

객체기반 모바일 콘텐츠 검색 기술 평가 지침

오원근
한국전자통신연구원
owg@etri.re.kr

Guideline for Object-based Mobile Contents Retrieval Technology Certification

Oh, Weon-Geun

Electronics and Telecommunications Research Institute

요약

본 논문에서는 객체기반의 모바일 콘텐츠 검색 기술에 대한 평가 항목, 평가 기준, 평가 방법을 설정하여 관련 기술의 성능을 객관적으로 검증하는 방법에 대해 설명한다. 객체기반의 모바일 콘텐츠 검색 기술은, 다양한 분야에서 활용이 가능한데, 대표적인 서비스 분야는 e-Learning 분야, 문화재 정보검색 분야, 모바일 AR 분야, 스마트 TV 분야, 모바일 쇼핑분야, 모바일 영상 감시분야 등 매우 다양하다. 본 논문에서 제시한 객체기반 모바일 콘텐츠 검색 기술 평가 지침은 국내 디지털 콘텐츠 업계에서 개발, 생산하고 있는 객체기반의 모바일 콘텐츠 검색 기술의 품질을 객관적으로 평가할 수 있는 체계를 제공함으로써, 사용자는 표준화된 객체기반의 모바일 콘텐츠 검색 기술의 인증서의 내용을 토대로 자신의 목적에 따라 선택하여 사용할 수 있도록 하고, 기술 개발자들에게는 보다 우수한 객체기반의 모바일 콘텐츠 검색 기술의 개발을, 평가자에게는 공정한 평가를 가능하게 할 수 있다.

1. 서론

시각은 인간의 생존과 지적 활동에 매우 중요하며, 뇌 정보처리 기능 중 가장 중요한 기능(60%이상)을 담당하고 있으며, 특히 지능 시스템의 가장 중요한 사물인식 및 정보입력 수단으로 활용되어 왔다. 이러한 인간의 시각기능을 검색분야에 적용한 응용기술이 내용(혹은 콘텐츠) 기반 영상 검색 기술이고 스마트폰을 이용하여 실세계에 적용한 기술이 모바일 시각 검색 기술(MVS : Mobile Visual Search) 혹은 객체기반 모바일 콘텐츠 검색 기술(이하 객체기반 모바일 콘텐츠 검색 기술)이다.

일반적인 모바일 콘텐츠 검색 기술은 객체를 근접촬영(그림 1.1 a) 하거나 객체가 포함된 영상의 전체를 입력수단으로 하여 가장 유사한 영상을 검색하는 방법(그림 1.1 b)을 사용하고 있다. 그러나 영상의 전체를 입력수단으로 하는 방법은 객체 이외의 영역이 검색 정보로 활용이 되어 검색 정확도를 낮추는 요인이 되며, 다양한 복수의 객체가 같이 존재할 경우(그림 1.1 c)에는 관심이 있는 객체를 독립적으로 검색하는 것이 불가능하다.



(a) 포도 (b) 귤 (c) 백화점 내부

그림 1.1 콘텐츠 기반 영상의 예

이러한 문제를 해결하기 위하여 제안된 기술이 객체기반 모바일 콘텐츠 검색 기술이며 객체기반 모바일 콘텐츠 검색 기술에서는 일반적인 콘텐츠 기반 영상 검색 기술에 비해 추가적인 요소기술이 필요하고 객체의 다양한 변형(크기, 시점, 조명, 가려짐 등)에도 강인한 기술이 요구된다(그림 1.2). 한편, 이러한 객체기반 모바일 콘텐츠 검색 기술은 검색의 정확도를 획기적으로 향상시킬 수 있으며 동시에 여러 객체를 검색할 수 있는 큰 장점이 있기 때문에 향후 이 기술에 대한 요구와 활용분야가 점차 증대될 것으로 사료된다.

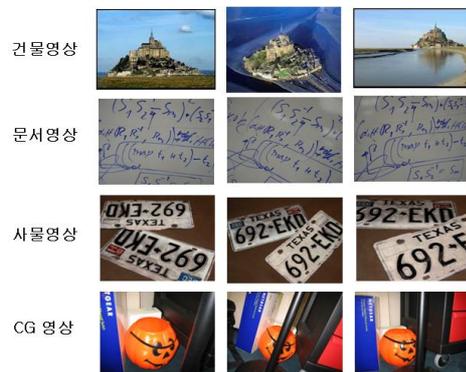


그림 1.2 객체영상의 예

2. 객체기반 모바일 콘텐츠 검색의 구조

객체기반 모바일 콘텐츠 검색기술의 흐름은 크게 그림 4.3과 같다.

(a) 특징점 추출(Feature Detector) : 영상을 구성하는 각각의 특정 객체 영역들에서 고유한 특징점을 추출한다. 특징점은 영상의 크기나 영상 변환에 강인하며, 노이즈 및 조명의 변화에도 강인한 특성이 있어야 한다. Harris Detector, Harris Laplace detector, Hessian Laplace detector, SIFT & SURF detector 등이 대표적이다.

(b) 서술자 추출(Feature Description) : 입력 패턴으로부터 대상을 표현하는 공통적인 성질을 추출하는 것. 대표적인 전역 서술자(global descriptor)로는 MPEG-7(색상, 질감, 모양, 움직임 등)이 있고, 지역 서술자(local descriptor)로는 SIFT, SURF, CHOG 등이 있다.

(c) 영역 레이블링(Feature Labeling) : 대부분의 영상은 하늘, 구름, 건물 등과 같이 다양한 의미를 지니고 있는 영역의 조합으로 구성되어 있는데, 이때 의미가 있는 영역을 객체 영역이란 하고 이들을 구성하고 있는 객체 영역을 구체적이며 독립적인 특징들로 표현하는 것을 영역 레이블링이라고 한다.

(d) 특징 정합(Feature Matching) : 두 개의 영상간에서 대응되는 특징을 찾는 과정을 일컬으며, 서로 다른 환경에서 추출된 특징들간에 안정된 다수의 특징을 찾는 것이 주목적임. KD-tree, index 구조에 의한 방법이 대표적이다.

(e) 기하학적 검증(Geometrical Verification) : 서로 다른 기하학적 환경에서 정합이 된 두 개의 영상간의 Parameter를 일치시키는 과정을 말한다. 일반적으로 대응되는 특징점들의 3차원 공간에서의 상대좌표 및 이들간의 통일된 기하학적 관계를 계산하는 방법을 사용한다. RANSAC, Advanced RANSAC 등이 있다.



그림 2.1 객체기반 모바일 콘텐츠 검색기술의 흐름

3. 객체기반 모바일 콘텐츠 검색 기술 성능 평가 기준 및 절차

3.1 성능 평가를 위한 실험 영상 DB의 구성 및 주석달기

객체기반 모바일 콘텐츠 검색 기술의 성능을 객관적이고 체계적으로 평가하기 위해서는 기술을 평가할 영상을 선정하는 구체적인 조건이 필요하다. 본 절에서는 객체기반 모바일 콘텐츠 검색 기술의 다양한 응용 분야와 성능 평가 기준을 충족시키기 위한 표준 영상의 구성과 성능평가 실험을 위해 영상 DB에 첨가된 주석(annotation) 대해 설명한다.

객체기반 모바일 콘텐츠 검색 기술의 성능을 평가를 위해서 선정된 영상의 리스트는 <표3-1>과 같이 5개 범주(category)의 정합영상 16,796매와 비정합영상 167,960매로 구성되어 있다. 선정된 영상은 성능평가 실험을 위한 영상 DB(정합 및 비정합 영상)에 다음과 같이 네 종류의 주석을 정의하였다.

(a) 영상 이름 : 한 장의 영상에 대해서 한 줄의 아스키 텍스트 파일

로 구성된 영상의 이름

(b) 정합 실험을 위한 주석 : 정합 실험을 위한 주석은 각각의 데이터 셋에 정합 영상 및 비정합 영상의 쌍으로 구성된 두개의 파일로 정의된다. 여기서 비정합 영상은 동일한 객체의 다른 각도에서의 촬영된 영상을 포함해서는 안 된다.

(c) 기하학적 검증을 위한 주석 : 실험 영상에는 기하학적 검증을 위한 영상 객체의 좌표정보가 제공이 되는데, 질의 영상의 객체와 대응되는 참조 영상의 객체 파일에 .bbox라는 확장자로 표기하고 영상 객체의 외접 4각형에서 추출한 4개의 모퉁이(corner)의 X,Y 좌표를 포함한 ASCII text file로 나타낸다.

표 3.1 성능평가를 위해서 선정된 영상의 종류 및 주석

실험 번호	영상의 범주(category)	실험영상의 구성	주석
1	Text와 그림 혼합영상	정합영상 : 3,000매 비정합영상 : 30,000매	(a)영상 이름 (b)정합 실험을 위한 주석
2	회화(painting) 영상	정합영상 : 400매 비정합영상 : 4,000매	
3	비디오 캡처 영상	정합영상 : 400매 비정합영상 : 4,000매	(c)기하학적 검증을 위한 주석
4	랜드마크/빌딩 영상	정합영상 : 5,346매 비정합영상 : 53,460매	
5	사물/정경(scene)영상	정합영상 : 7,650매 비정합영상 : 76,500매	

3.2 성능평가 방법 및 절차

3.2.1 강인성 평가

객체기반 모바일 콘텐츠 검색 기술의 강인성을 평가하기 위해서 본 표준에서는 한 장의 영상을 표현하는 서술자의 크기를 다음과 같이 세분화하였다.

512, 1K, 2K, 4K, 8K, 16K bytes

정합 실험의 성능을 평가하기 위해서 여기서는 각각의 서술자에 대해서 다음의 세 가지를 평가 기준으로 설정하였다.

(a) 정합 성공률(success rate) : 주어진 false alarm rate내에서의 정합 성공률

(b) 비정합 영상 실험 : 비정합 영상을 대상으로 실험

$$\text{False Positive Rate} = \text{FP}/(\text{FP}+\text{TN})$$

(c) 정합 영상 실험 : 정합 영상을 대상으로 실험

$$\text{True Positive Rate} = \text{TP}/(\text{TP}+\text{FN})$$

여기서,

TP : the number of true positive

FP : the number of false positive

TN : the number of true negative

FN : the number of false negative를 각각 나타낸다.

3.2.2 기하학적 정합의 정확도 평가

객체기반의 콘텐츠 기술의 성능을 평가하기 위해서는 질의 영상과 참조 영상에 있는 객체간의 기하학적 정합성을 측정하는 것이 매우 중요하다. 여기서는 다음과 같이 두 객체영역의 영역의 합과 겹치는 영역

의 비율로 정하였다.

<두 객체의 외접 사각형 영역의 겹치는 영역>
<두 객체의 외접 사각형 영역의 합>

기하학적 정합의 정확도의 평균값은 다음의 두 가지로 계산이 된다.

- (a) 모든 정합쌍(정합 및 비정합)에 대한 기하학적 정합의 정확도의 평균값
- (b) 정합쌍에 대한 기하학적 정합의 정확도의 평균값

3.2.3 계산 복잡성 평가

객체기반 모바일 콘텐츠 검색 기술의 계산의 복잡성 평가는, 제한 성능의 PC(CPU≤3.4GHz)상에서 식별자 추출 복잡성 평가와 식별자 정합 복잡성 평가로 나눌 수 있으며 계산 방법은 다음과 같다.

표 3.2 식별자 추출 복잡성

식별자 추출 복잡도	식별자 추출 시간/식별자 크기
------------	------------------

표 3.3 식별자 정합 복잡도

식별자 정합 복잡도	식별자 정합 시간/식별자 크기
------------	------------------

표 3.4 식별자 정합 복잡성(기하학적 검증시간 포함)

식별자 정합 복잡도	식별자 정합 시간/식별자 크기
------------	------------------

3.2.4 성능 평가표

객체기반 모바일 콘텐츠 검색 기술의 성능을 나타내는 전체적인 성능평가표는 표 3.5와 같다.

표 3.5 객체기반 모바일 콘텐츠 검색 기술의 성능평가표

실험 번호	서술자 크기	성능		계산 복잡성	
		정합 성공율	기하학적 정합 정확도(a,b)	기하학적 정합과정 (X)	기하학적 정합 과정 (O)
1~5	512				
	1K				
	2k				
	4k				
	8k				
	16k				

4. 결론

본 논문은 국내 디지털 콘텐츠 업계에서 개발, 생산하고 있는 객체기반의 모바일 콘텐츠 검색 기술의 품질을 객관적으로

평가할 수 있는 체계를 제공함으로써, 사용자는 표준화된 객체기반의 모바일 콘텐츠 검색 기술의 인증서의 내용을 토대로 자신의 목적에 따라 선택하여 사용할 수 있도록 하고, 기술 개발자들에게는 보다 우수한 객체기반의 모바일 콘텐츠 검색 기술의 개발을, 평가자에게는 공정한 평가를 가능하게 할 수 있다.

한편, 본 기술은 최근 국내외 모바일 검색 및 영상 서비스 산업에서 기술 개발 및 시장 진입 경쟁에서 국내 산업이 선도적인 역할을 하는데 기여를 할 것이다.

참고문헌

- 1) The MPEG Home Page(<http://mpeg.chiariglione.org/>)

본 연구는 지식경제부 정보통신표준화 및 인증지원사업의 일환으로 수행하였음. [2012-PM10-04, 디지털 콘텐츠 핑거프린팅 표준개발]