

실시간 비디오 검색 기술 평가 및 인증

오 원근^o, 이 승재, 이 근동, 정 다운, 손 형관
한국전자통신연구원 차세대콘텐츠연구본부

{owg, seungilee, zacurr, crisia, kwanson}@etri.re.kr

Guideline for Real-Time Video Search Technology Certification

Oh Weon-Geun, Lee Seung-Jae, Lee Keun-Dong, Jung Da-Un and Son Hyung-Kwan

Electronics and Telecommunications Research Institute

요약

최근 디지털 콘텐츠 및 멀티미디어는 거의 모든 IT 산업에서 중요한 역할을 하고 있으며, 이중에서도 비디오 콘텐츠는 그 동안의 검색, 유통, 관리 등의 수동적인 범위를 넘어 실세계와의 실시간 상호작용을 통한 능동적이며 혁신적인 역할을 수행하고 있다. 실시간 비디오 검색 기술은, 다양한 분야에서 활용이 가능한데, 대표적인 서비스 분야는 ALV(Autonomous Land Vehicle : 무인자동차), SNS 서비스, 오락/스포츠/광고 서비스, 모바일 쇼핑, AR, Surveillance 분야 등 매우 다양하다.

본 논문에서는, 실시간 비디오 검색 기술의 개요와 적용분야 및 사례를 설명하고 실시간 비디오 검색기술을 객관적으로 평가할 수 있는 방법, 절차에 대한 인증서에 대한 규격을 제정하여, 사용자가 표준화된 실시간 비디오 검색 기술의 인증서의 내용을 토대로 자신의 목적에 따라 기술을 선택하여 사용할 수 있도록 하였다.

1. 서론

최근 디지털 콘텐츠 및 멀티미디어는 거의 모든 IT 산업에서 중요한 역할을 하고 있으며, 이중에서도 비디오 콘텐츠는 그 동안의 검색, 유통, 관리 등의 수동적인 범위를 넘어 실세계와의 실시간 상호작용을 통한 능동적이며 혁신적인 역할을 수행하고 있다. 특히 최근의 영상처리 기술의 급속한 발전은 콘텐츠의 의미적 검색의 실용화까지 가능하게 하였으며 이것은 시각정보를 이용한 실세계의 실세계 비주얼 검색 기술의 발전에 힘입은 바가 크다. 이중에서도 영상의 공통적인 속성(generic attribute)을 기반으로 한 비주얼 검색은 현재 가장 어려운 과제로 간주되고 있다. 이것은 일반적으로 영상이 표시하는 의미(혹은 기능)의 범위가 넓어 동일한 범주에 속하는 대상의 외형의 변화가 매우 크기 때문이다. 그러나 최근의 (1) 새로운 영상 표현방법의 발전, (2) 기계 학습 기술의 발전, (3) Web의 보급에 의한 데이터베이스 구축의 용이성, (4) 계산기의 고속화·대용량화에 의해 연구가 크게 발전하여 일반 물체 인식의 실용화가 현실이 되어가고 있으며 그중에서도 실시간 비디오 검색 기술은, 다양한 분야에서 활용이 가능한데, 대표적인 서비스 분야는 ALV(Autonomous Land Vehicle : 무인자동차), SNS 서비스, 오락/스포츠/광고 서비스, 모바일 쇼핑, 미디어 검색/관리, AR, Surveillance 분야 등 매우 다양하다.

최근 실시간 상황인식(자율주행 로봇, 자동차 등) 및 실세계와 관련된 정보(이동환경, 동식물/음식/의상 등) 검색 및 서비스의

필요성이 급격하게 증가됨에 따라 많은 기업과 표준 기구 및 단체에서 이와 관련된 기술개발과 표준 및 상용화를 진행하고 있다.

ILSVRC(ImageNet Large Scale Visual Recognition Challenge)에서는 PASCAL VOL 보다 확장된 DB에서 2010년부터 현재까지 대규모 비주얼 검색 기술 개발을 목적으로 다양한 실 객체의 검출(detection) 및 분류(classification)에 대한 국제적 기술경쟁을 진행하고 있으며(그림 1.1), MPEG CDVA(Compact Descriptor for Video Analysis)에서는, 비디오 해석을 통한 물체 탐지 및 분류, 미디어 검색에 대한 국제표준의 제정을 목적으로 현재 요구사항(requirements)을 정리하고 있다.



그림 1.1 ILSVRC 2014의 다양한 Object Class

한편 구글 등은 이러한 실세계 비디오 검색 기술을 활용한 응용제품의 상용화를 오래전부터 추진하고 있는데 그 목적을 인명 사고의 대폭 감축, 에너지 절감, 출근시간의 효과적인 활용 등 3가지를 제시했다. 이 외에도, 세계 자동차 10대 생산기업(BMW, Ford, GM, Mazda, Mercedes-Benz, Tesla, Toyota, Volkswagen, Audi and Volvo)는 최근, 자동차의 에어백시스템과 같이 자동 브레이크 시스템(Automatic Braking System)을 신규 자동차의 표준 사양으로 도입할 것에 동의하였으며, 국내에서는 네이버가 로봇·자율주행차 등 생활과 밀접한 3대 기술 분야에 향후 5년간 1000억 원을 투자하는 ‘프로젝트 블루’도 발표했다.

본 논문에서는 실시간 비디오 검색 기술에 대한 평가 항목, 평가 기준, 평가 방법을 설정하여 관련 기술의 성능을 객관적으로 검증하는 방법에 대해 설명한다. 실시간 비디오 검색 기술 평가 지침은 국내 디지털 콘텐츠 업계에서 개발, 생산하고 있는 실시간 비디오 검색 기술의 품질을 객관적으로 평가할 수 있는 체계를 제공함으로써, 사용자는 표준화된 실시간 비디오 검색 기술의 인증서의 내용을 토대로 자신의 목적에 따라 선택하여 사용할 수 있도록 하고, 기술 개발자들에게는 보다 우수한 실시간 비디오 검색 기술의 개발을, 평가자들에게는 공정한 평가를 가능하게 할 수 있다.

2. 실시간 검색 기술의 개요

2.1 실시간 비디오 검색 기술의 개념 및 구조

실시간 비디오 검색 기술은 다양한 영상처리 응용 분야에서 입력된 영상을 적용할 응용분야(예; 보행자, 장애물 감지 및 경고 혹은 자동차의 자동 멈춤, 침입자, 비상상황 감지 및 경고 등)에 적응적으로 실시간 처리하여 사고를 미연에 방지하는 것이 그 목적으로, 고속 처리시간과 정확성이 기본적으로 요구되는 기술적 사항이다. 실시간 비디오 검색 기술의 구조는 (그림 2.1)과 같으며 크게 비디오 입력 모듈, 실시간 영상 해석 모듈, 응용 시스템의 세 부분으로 나눈다.

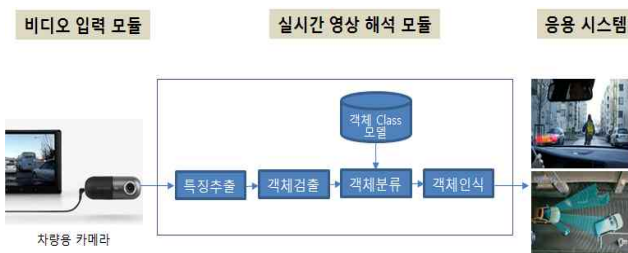


그림 2.1 실시간 비디오 검색 기술의 구조

먼저 비디오 입력 모듈인 차량용 카메라는 운전중 혹은 주차중 상황을 녹화하여 사고발생시 사고의 원인을 밝히는 목적으로 사용된다. 차량용 카메라는 크게 1) 화면 해상도, 2) 화각, 3) 프레임율 등으로 구분할 수 있는데, 최근에는 후방녹화 기능과, 선명한 번호판 식별을 원하는 소비자들을 위한 Full HD제품과 차선 이탈 방지 등의 운전자 지원기능도 개발, 도입중이다.

실시간 영상 해석 모듈은, 입력된 영상을 처리하여 다양한 환경(조명, 시점, 위치 변화)에도 강하고 재연성이 높은 특징을 추출, 처리하여 목적에 맞는 객체가 존재하는 영역을 검출하고 객체를 다양한 클래스(예; 사람, 차 등)로 분류한 다음 적절한 객체(성인, 어린이, 승용차, 트럭, 오토바이, 자전거 등)로 인식하는 기능을 수행한다.

3. 실시간 검색 기술의 성능 평가 절차 및 기준

3.1 실험 영상의 선정 조건

실험영상은 비디오 입력 카메라의 사양과 입력 대상 및 조건에 따라 다음과 같이 나눌 수 있다

가) 비디오 입력 카메라의 사양

- 화면 해상도(Display Resolution) : 화면 해상도란 한 화면에 화면을 구성하는 점, 즉 화소(畫素, pixel)의 개수를 의미하며, 크게 VGA(Video Graphics Array, 640×480), HD(High Definition, 1280×720), FHD(Full HD, 1920×1080)로 구분된다.
- 초당 프레임율(Frame rate per second) : 초당 프레임율은 1초 동안 보여주는 화면의 수를 가리키며, 로마자 약어 및 단위로는 fps(또는 프레임/초)를 쓴다. 프레임이 높으면 높을수록 찍히는 사진이 많아 영상 화면이 부드럽지만 반대로 프레임이 낮으면 화면이 단속적인 느낌이 든다. 여기서는 15fps와 30fps 두 종류로 나눈다.

나) 입력 대상 : 입력 대상은 사람과 차량으로 나누며, 사람은 성인과 아동으로, 차량은 승용차, 트럭, 오토바이 및 자전거로 구분한다.

다) 입력 조건 : 입력 조건은 크게 조명 조건, 촬영 각도 및 촬영 위치로 나눌 수 있는데 여기서는 주간과 야간으로만 구분한다.

3.2 실험 영상의 구성

가) 실험 영상 : 실험 영상은 앞 절의 조건에 따라 각각의 조건을 만족하는 표준 영상 72종(각 20매) 1,440매와 디스트랙터 영상 720매 등 총 2,160매로 구성된다.

나) 참조 영상(reference image) : 총 1,440매의 표준 영상에서 임의의 720매의 영상으로 구성되며 이중 360매는 UGC(User Generated Content)로서 선정되어야 한다.

다) 질의 영상(query image) : 표준 영상에서 참조 영상을 제외한 720매의 비디오 영상으로 구성되며, 검색 대상 객체가 존재하는 프레임의 앞뒤의 3초를 포함하는 영상이어야 한다.

라) 디스트랙터(distractor image) 영상 : 검색 대상 객체가 존재하지 않은 비디오 영상 720매로 구성되며, 실시간 비디오 검색 SW의 성능을 상대적으로 점검할 수 있는 비디오 영상이어야 한다.

3.3 실험 영상 DB의 주석달기(annotation)

참조 영상과 질의 영상은 사전에 객체가 존재하는 프레임의 프레임 이름과 객체 영역(사각 영역) 및 이름이 색인되어 있어야 한다.

4. 실시간 검색 기술의 성능 평가

4.1 성능 평가 방법

실시간 비디오 검색 기술의 평가는 크게 검색 평가(retrieval evaluation)와 쌍대 비교평가(pair-wise comparison evaluation)가 있다. 검색 평가는 질의 영상에 대해 전체 실험 영상 DB로부터 적절한(relevant) 영상을 찾아내는 실험 평가이고(그림 4-1), 쌍대 비교평가는 질의 영상과 질의 영상에 대응하는 참조 영상을 일대일로 비교하여 참조 영상내의 객체를 제대로 찾는 가를 평가하는 실험이다(그림 4-2).



그림 4.1 검색 평가의 구조

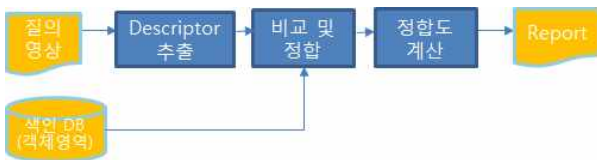


그림 4.2 쌍대 비교 평가의 구조

4.2 성능 평가의 종류

성능평가는 크게 검색 성능 평가, 쌍대 비교 성능 평가, 디스크립터 크기, 검색 및 정합 복잡도로 나눌 수 있다.

가) 검색 성능 평가

- 평균 검색 정확도(MAP : Mean Average Precision, %)
- 제 1순위 정확도(Success rate for top match, %)
- 재현율(Recall, %)

나) 쌍대 비교 성능 평가

- 주어진 오검출률(false alarm rate)에서의 검색 성공률(%)
- 질의 영상과 참조 영상 객체간의 정합 정밀도(%)

다) 디스크립터 경량성(KB)

라) 검색 및 정합 복잡도(Complexity)

- 디스크립터 추출 시간(sec)
- 평균 검색 시간(sec)
- 평균 쌍대 비교 시간(sec)

4.2 성능 평가표

<표 4-1>은 전체 실험 영상 DB에 대한 시스템의 성능표이다.

<표 4-1> 전체 시스템 성능표

구분	세부 기능	성능
검색 성능	평균 검색 정확도(%)	
	제 1순위 정확도(%)	
	재현율(%)	
쌍대 비교 성능	검색 성공률(%) (k false alarm rate 기준)	
	정합 정밀도(%)	
디스크립터 경량성	디스크립터 크기(KB)	
	디스크립터 추출 시간(sec)	
검색 및 정합 복잡도	평균 검색 시간(sec)	
	평균 쌍대 비교 시간(sec)	

<표 4-2>과 <표 4-3>는 검색 대상(사람, 차량)에 대한 검색 성능을 나타내는 표이다. 검색 성능은 각각의 질의 영상에 대해 1순위로 정확하게 검색된 영상의 확률(%)을 나타낸다.

<표 4-2> 검색 성능표(사람)

구분	성인/어린이	주간/야간 (day/Night)	카메라 해상도 (Definition)	프레임율 (Frame Rate)	성공률(%) (Success rate)	
사람	성인	주간	VGA	15 fps		
				30 fps		
				15 fps		
			30 fps			
			15 fps			
			30 fps			
		야간	주간	VGA	15 fps	
				30 fps		
				15 fps		
			30 fps			
			15 fps			
			30 fps			
어린이	주간	주간	VGA	15 fps		
				30 fps		
				15 fps		
			30 fps			
			15 fps			
			30 fps			
		야간	주간	VGA	15 fps	
				30 fps		
				15 fps		
			30 fps			
			15 fps			
			30 fps			

<표 4-3> 검색 성능표(차량)

구분	성인/어린이	주간/야간 (day/Night)	카메라 해상도 (Definition)	프레임율 (Frame Rate)	성공률(%) (Success rate)		
차량	승용차	주간	VGA	15 fps			
				30 fps			
				15 fps			
				30 fps			
				15 fps			
				30 fps			
			야간	VGA	15 fps		
				30 fps			
				15 fps			
				30 fps			
				15 fps			
				30 fps			
	트럭	주간	주간	VGA	15 fps		
					30 fps		
					15 fps		
				30 fps			
				15 fps			
				30 fps			
			야간	주간	VGA	15 fps	
					30 fps		
					15 fps		
				30 fps			
				15 fps			
				30 fps			
오토바이	주간	주간	VGA	15 fps			
				30 fps			
				15 fps			
			30 fps				
			15 fps				
			30 fps				
		야간	주간	VGA	15 fps		
				30 fps			
				15 fps			
			30 fps				
			15 fps				
			30 fps				
자전거	주간	주간	VGA	15 fps			
				30 fps			
				15 fps			
			30 fps				
			15 fps				
			30 fps				
		야간	주간	VGA	15 fps		
				30 fps			
				15 fps			
			30 fps				
			15 fps				
			30 fps				

5. 인증

5.1 인증 신청 및 성능 평가

실시간 비디오 검색 기술(혹은 제품)의 성능평가 및 인증절차는 (그림 5-1)과 같다.

먼저 기술 소유자는 실시간 비디오 검색 기술과 인증서 양식(A)과 같이 인증기관 정보, 인증신청자 정보, 실시간 비디오 검색 기술의 일반정보를 포함한 인증신청서를 이메일이나 웹페이지 등의 전자적 방법을 사용하여 인증기관에 제출한다. 인증기관은 제출된 실시간 비디오 검색 기술을 앞의 4절의 기술평가를 통하여 성능평가 과정을 수행한다.

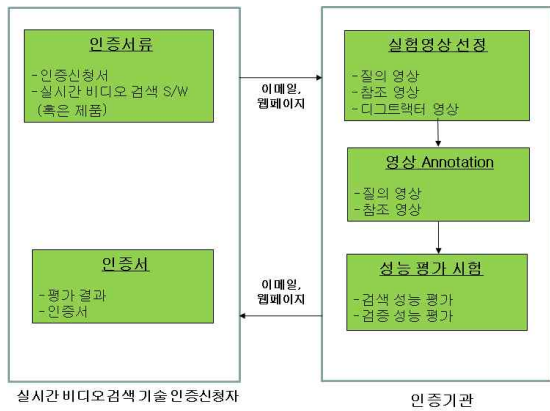


그림 5.1 실시간 비디오 검색 기술의 평가 및 인증 절차

5.2 인증서 교부

실시간 비디오 검색 기술의 인증기관은 실시간 비디오 검색 기술의 평가결과를 인증서에 포함하여 인증 신청기관에 보낸다. 이때, 인증기관은 인증서 데이터베이스를 구축하여 발급된 인증서들을 관리하여야 한다.

5.3 인증기관 요건

인증기관은 인증 업무를 수행할 수 있는 인증시설을 구비하여야 하며, 인증업무에 대한 전담자를 두어야 한다. 여기서 인증시설은 실시간 비디오 검색 기술 소유자로부터 이 규격에서 정해진 대로 인증신청을 받아 실시간 비디오 검색 기술을 인증할 수 있는 시험 영상 데이터베이스, 실시간 비디오 검색 평가 S/W를 갖추고 성능평가를 실시하여 그 결과 성능에 대한 인증서를 발부할 수 있는 컴퓨터 하드웨어 및 소프트웨어를 말한다.

5.4 인증서 양식

<표 5-1>은 실시간 비디오 검색 기술 평가를 위한 인증서의 양식이다. 인증서 양식에 있는 각 항목의 의미는 다음과 같다.

<표 5-1> 인증서 양식

인증기관 정보	인증기관 명	
	인증기관 주소	
	인증기관 일련번호	
인증신청자 정보	인증신청자 명	
	인증신청자 주소	
	인증신청자 식별 번호	
실시간 비디오 검색 기술 일반 정보	실시간 비디오 검색 기술 식별 번호	
	특허 (선택)	
	알고리즘 설명	
	대상 응용	

가) 인증기관 정보

인증기관 정보는 인증기관 명, 인증기관 주소, 인증기관 일련번호로서 구성된다.

- 인증기관 명: 인증기관의 법인 명
- 인증기관 주소: 우편번호와 인증기관이 소속되어 있는 나라, 시/도, 동, 번지
- 인증기관 일련번호: 인증기관을 구분하기 위해 발부되는 유

일한 번호

나) 인증신청자 정보

인증신청자 정보는 인증신청자 법인/개인명, 인증신청자 주소, 인증신청자 일련번호로서 구성된다.

- 인증신청자 이름: 관심영역의 정지영상 식별 기술 소유자의 자연인 혹은 법인명
- 인증신청자 주소: 나라, 시/도, 동, 번지, 우편번호 등의 주소
- 인증 신청사 식별번호: 관심영역의 정지영상 식별 기술 소유자의 신원을 확인할 수 있는 주민등록 번호/사업자 등록 번호

다) 실시간 비디오 검색 기술 일반 정보

- 실시간 비디오 검색 기술 식별번호: 인증기관이 부여하는 각 실시간 비디오 검색 기술을 구별할 수 있는 유일한 번호
- 특허 (선택): 실시간 비디오 검색 기술에 관련된 특허 정보로 선택 사항임
- 알고리즘 (선택): 실시간 비디오 검색 기술에 관련된 알고리즘 정보로 선택 사항임
- 응용 분야: 자율 주행, 상황 감시, 미디어 및 엔터테인먼트

6. 결론

본 논문은 국내 디지털 콘텐츠 업계에서 개발, 생산하고 있는 실시간 비디오 검색 기술의 품질을 객관적으로 평가할 수 있는 체계를 제공함으로써, 사용자는 표준화된 실시간 비디오 검색 기술의 인증서의 내용을 토대로 자신의 목적에 따라 선택하여 사용할 수 있도록 하고, 기술 개발자들에게는 보다 우수한 실시간 비디오 검색 기술의 개발을, 평가자에게는 공정한 평가를 가능하게 할 수 있다.

한편, 본 기술은 최근 국내의 차세대 자동차, 실시간 감시 및 영상 서비스 산업에서 기술 개발 및 시장 진입 경쟁에서 국내 산업이 선도적인 역할을 하는데 기여를 할 것이다.

참고문헌

- 1) <http://image-net.org/challenges/LSVRC/2015/>
- 2) <http://wg11.sc29.org/>
- 3) 네이버, 로봇·자율주행차 만든다, 중앙일보 2015. 09.1 5
- 4) Automakers Will Make Automatic Braking Systems Standard in New Cars, BILL VLASIC: The New York Times(2015. 9. 11),
- 5) 자동운전 자동차(self-driving car), 어떻게 볼 것인가, 공영일: [방송통신정책 제25권 7호] 정보통신정책연구원

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.R0132-15-1005, 온.오프라인에서의 콘텐츠 비주얼 브라우징 기술개발)