

자동차 전방 감시를 위한 영상 구문 검출 평가 기술

*오원근 *고종국
 *한국전자통신연구원
 *owg, gko@etri.re.kr

Visual Phrase Detection and Evaluation Technology for Car Front Monitoring

*OH, Weon Geun and *KO, Jong-Gook
 *Electronics and Telecommunications Research Institute

요약

영상이 포함하고 있는 풍부한 정보를 검출하고 이해하기 위해서는, 영상속의 일관된 상호관계를 갖는 영상 객체 그룹을 이해하고 표현하는 영상 구문 검출 기술(Visual Phrase Detecting Technology)이 필수적이다. 영상 구문 검출 기술은, 영상이 포함하고 있는 다양하고 풍부한 정보를 추출하고 활용하기 위한 핵심 기술로 이를 이용한 자동차 주행중 전방 감시, 영상의 자동 주석 달기, 동영상의 검색, 쇼핑 공간에서의 부정행위 검출(fraud detection) 등 다양한 분야에 적용할 수 있다.

본 논문에서는 자동차 전방감시를 위한 영상 구문 검출 기술을 소개하고, 이를 평가하기 위한 표준 영상 DB의 선정과 평가 기준 및 절차에 대해서 설명한다.

1. 서론

영상 구문(Visual Phrase)은, 영상속에서 일관된 상호 관계(유형, 공간, 크기 및 시점 관계)를 갖는 두 개 이상의 영상 객체의 그룹을 의미한다. 이러한 영상 구문은 영상을 이해하기 위한 중요한 계층적 의미를 부여하며, 이를 이용하면 영상을 기술(description)하거나 관심이 있는 영상 객체를 검출하고, 해석하는데 있어서 보다 정확하고 안정된 결과를 얻을 수 있다.

이러한 영상 구문 검출 기술은, 반세기 이상에 걸쳐 수많은 대학과 연구소, 기업 연구소에서 진행이 되어 왔으며, 그 방법도 매우 다양하다. 그러나 개발된 기술이 실제 실세계에 적용이 된 경우는 극히 희박하다. 이러한 배경으로는 처리하는 데이터가 워낙 방대하고 이를 처리할 수 있는 컴퓨터 H/W(연산 및 저장 장치)의 부재와 이들을 처리, 분석하는 S/W(알고리즘)의 기술적 한계 등을 들 수 있다.

최근에 등장한 딥 러닝(Deep Learning) 기술은, 고성능 GPU 컴퓨팅 장치의 등장과 과적용(over-fitting) 등의 기술적인 문제의 해결, 기술을 적용할 수 있는 대용량의 데이터(big data)의 발생으로 영상과 같은 비정형데이터를 효율적으로 처리, 해석할 수 있게 되었다.

영상 구문 검출 기술은, 영상이 포함하고 있는 다양하고 풍부한 정보를 추출하고 활용하기 위한 핵심 기술로 이를 이용한 자동차 주행중 전방 감시, 영상의 자동 주석 달기, 동영상의 검색, 쇼핑 공간에서의 부정행위 검출(fraud detection) 등 다양한 분야에 적용할 수 있다.

본 논문에서는, 자동차 주행중 사고 방지를 위한 전방 감시를 목

적으로 전방 카메라를 통해서 입력된 영상을 해석하여 영상속 객체(사람, 자전거, 바이크, 킥 보드 등)를 인식하고 이들의 관계(Visual Phrase)를 검출하는 기술과 이 기술을 평가하기 위한 표준 영상 DB의 선정과 평가 기준 및 절차 설명한다.

2. 자동차 전방 감시를 위한 영상 구문 검출 기술의 활용

주행중인 자동차의 사고방지를 위해서는 자동차의 전방카메라에서 입력된 영상을 처리하여 영상속의 객체(사람이나 자전거 등)를 검출하고, 인식할 필요가 있다.

영상 객체 검출 및 인식 기술은 크게 영상 속성을 추출하고 추출된 영상 속성을 이용하여 영상 자체를 인식하는 방법과 영상속의 객체를 검출(detection)하고 객체를 해석하여 분류(classification)하고 인식(recognition)하는 방법으로 나눌 수 있다(그림 1 참조).

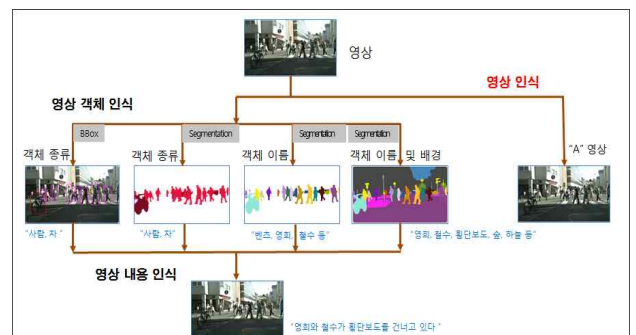


그림 1. 영상 객체 검출 및 인식 기술의 구분

객체 검출 기술은 다시 객체를 class, semantic object, instance object 및 image contents로 나누어 검출 하는 기술로 세분화 할 수 있다.

(a) 영상 인식 기술 : 영상 전체의 속성(global feature) 혹은 영상의 지역적인 속성(local feature)을 체계화하여 영상을 검색(retrieval)하거나 영상 속 객체의 존재여부를 확인하는 것이 주목적이다.

(b) 영상 객체 검출 및 인식 기술 : 영상을 해석하여 영상속의 객체를 검출하고 검출된 객체의 종류(class)나 이름(instance)을 인식하는 것이 목적이다. 여기서 객체를 추출하는 방법은 영상을 포함하는 외각 사각형(bounding box)을 검출하는 방법과 객체의 경계선(boundary)을 검출하는 방법으로 다시 나눌 수 있다.

(c) 영상 내용 인식 기술 : 검출된 객체와 배경을 해석하여 영상속 객체의 상황(situation)을 인식하는 것이 목적이다. 영상 내용 인식 기술은 다시, 영상 속 객체들의 상호 관계를 인식하는 영상 구분 인식과, 영상의 전반적인 내용을 기술하는 영상 기술(description)하는 기술로 나눌 수 있다(그림 2 참조).



그림 2. 영상 이해 (scene understanding)의 단계

이중에서도 영상 구분 검출 및 인식 기술은, 자동차 전방의 상황을 빠르고, 정확하게 인식하고, 급박한 상황에 대처할 수 있는 기술로 활용될 수 있어 많은 연구가 진행이 되고 있다^{(1), (2), (3)}.

특히 최근의 교통사고 추세를 보면, 어린이나 노인들의 사고가 급증하고 있고, 또 새로운 이동 수단(노인용 보행기, 전동 배달차 혹은 킥보드 등)에 대해서도 대응을 할 필요가 급증하고 있다⁽⁴⁾.

3. 자동차 전방 감시를 위한 영상 구분 검출 기술 평가를 위한 영상 DB의 구축

영상 구분 검출 기술의 평가를 위해서는 알고리즘의 학습, 테스트 및 평가를 위한 일정 규격의(혹은 표준) 영상 데이터베이스가 필요하며, 표준화 된 영상 데이터베이스는 새로운 알고리즘 개발을 위한 모델의 구현과 객관적인 평가 방법(평가 기준, 절차 등)을 제공하는데 필수적인 조건이다. 이것은 개발 기술의 성능을 객관적으로 평가, 인증할 수 있는 표준화된 실험 영상을 선정하고 제공함으로써, 기술 개발자들에게는 보다 우수한 영상 속성 구조화 기술의 개발을, 평가자(혹은 평가기관)에게는 공정한 절차를 통한 평가를 가능하게 할 수 있기 때문이다.

3.1 자동차 전방 감시를 위한 영상 구분 검출 기술 성능 평가를 위한 실험 영상의 유형

자동차 주행중 사고의 유형은 매우 다양하나, 그 중에서도 자전

거, 모터사이클과 노인들의 보행기 및 최근 급증하고 있는 배달차나 킥보드(전동 혹은 무동력 포함) 등의 이동 수단에 대해서도 새로운 영상 구분 검출 및 인식 기술과 평가 기술이 필요하다. 이 경우 실험 영상의 유형은 다음과 같이 5종류(복수 객체를 포함할 경우 10종류)로 나눌 수 있다.

- (a) 자전거를 타고 있는 사람(혹은 사람들)
- (b) 모터사이클(오토바이)을 타고 있는 사람(혹은 사람들)
- (c) 노인용 전동 스쿠터를 타고 있는 사람(혹은 사람들)
- (d) 노인용 보행기를 밀고 가는 사람(혹은 사람들)
- (e) 기타(전동 배달차, 전동휠, 킥보드(전동, 무동력 등)를 타고 있는 사람(혹은 사람들))



그림 3. 실험영상의 유형(예)

3.2 자동차 전방 감시를 위한 영상 구분 검출 기술 성능 평가를 위한 영상의 데이터베이스의 구성

영상 구분 검출 및 인식 기술의 성능을 평가하기 위한 영상 데이터베이스는 크게 알고리즘 개발을 위한 학습 데이터 셋(training dataset), 모델 검증을 위한 데이터 셋(validation dataset) 과 성능 평가를 위한 테스트 셋(test set)으로 나눈다(그림 4 참조).

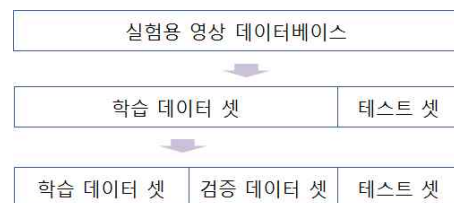


그림 4. 실험용 영상 데이터베이스 구성

- 실험용 영상 데이터베이스 : 실험용 영상은 다음의 10가지 유형(복수 객체 유형 포함)의 1,350매의 영상으로 구성되고, 각 유형에 대한 영상의 수는 다음과 같다.

영상의 유형	유형 1 (복수)	유형 2 (복수)	유형 3 (복수)	유형 4 (복수)	유형 5 (복수)	합계
영상의 개수(매)	180 90	180 90	180 90	180 90	180 90	1,350

- 학습용 영상 데이터 셋 : 학습용 영상은, 각 유형의 영상으로부터 다음과 같이 임의로 50%씩 선정한다.

영상의 유형	유형 1 (복수)	유형 2 (복수)	유형 3 (복수)	유형 4 (복수)	유형 5 (복수)	합계
영상의 개수(매)	90 45	90 45	90 45	90 45	90 45	675

- 검증용 영상 및 테스트용 영상 데이터 셋 : 검증용 영상과 테스트용 영상은, 각 유형의 영상으로부터 다음과 같이 임의로 각각 1/3씩 선정한다.

영상의 유형	유형 1 (복수)	유형 2 (복수)	유형 3 (복수)	유형 4 (복수)	유형 5 (복수)	합계
영상의 개수(매)	60 30	60 30	60 30	60 30	60 30	450

3.3 실험용 영상의 주석 달기(annotation)

영상 구문 검출 및 인식 실험을 위해서는 각각의 영상에 대해 유형정보와 객체의 위치 및 크기에 대한 정보가 색인이 되어 있어야 한다. 주석 달기는 크게 세 종류로 나눌 수 있다.

- (1) 영상 자체에 대한 색인 : 영상의 유형정보
- (2) 영상 객체의 위치 및 크기에 대한 색인(외접 사각형) : 영상의 유형 및 영상속 객체(단일 혹은 복수의 객체)의 외접 사각형(bounding box)과 외접 사각형의 중심점 좌표(position)
- (3) 영상 객체의 위치 및 크기에 대한 색인(경계선) : 영상의 유형과 영상속 객체(단일 혹은 복수의 객체)의 경계선(boundary)과 경계선영역의 중심 좌표(position)
- (4) 영상 객체(혹은 복수 객체)간 상호 관계(위/아래, 좌/우 등) 그림 5 참조.

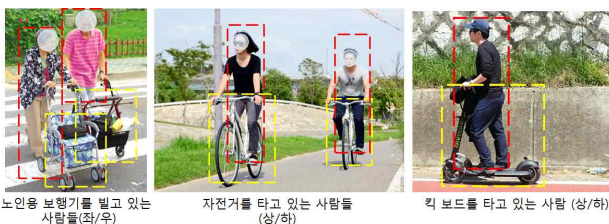


그림 5. 영상 구문 및 상호 관계에 대한 주석 달기

4. 자동차 전방 감시를 위한 영상 구문 검출 기술 평가 방법

영상 객체 인식 기술의 성능 평가 방법은 크게 세 가지로 나눌 수 있다. 첫 번째는 영상의 유형만을 확인하는 방법이며, 두 번째는 영상속의 객체를 검출하고 검출된 객체를 인식하고 객체의 위치(중심 좌표) 및 크기(외접 사각형의 면적)를 확인하는 방법, 세 번째는 영상속의 객체를 검출하고 검출된 객체를 인식하고 객체의

위치(중심 좌표) 및 크기(경계선 내부의 화소 수)를 확인하는 방법, 네 번째는 검출된 객체간(혹은 복수 객체간) 상호관계를 인식하는 방법이다.

각 기술에 대한 성능 평가 방법은 다음과 같다.

4.1 영상 유형의 확인

10종류의 영상 유형을 확인하는 것이 목적. 테스트 영상 450매에 대하여 실험을 하여 영상의 유형이 일치하면 0, 일치하지 않으면 1로 계산을 하여, 누적된 값을 error율(%)로 계산.

4.2 영상 객체 추출(외접 사각형) 및 인식

영상속 객체의 위치와 크기를 검출하고 유형을 확인하는 것이 목적. 영상의 객체(복수 객체 포함)를 검출하고 유형이 일치하고, 외접 사각형이 50%이상 겹치면 객체당 0, 겹치지 않으면 1로, 겹치는 비율을 정확도로 계산을 하여, 유형별로 정밀도(정확도, error율)를 계산하고 이의 평균값(mAP)으로 평가.

4.3 영상 객체 추출(경계선) 및 인식

영상속 객체의 위치와 크기를 검출하고 유형을 확인하는 것이 목적. 영상의 객체(복수 객체 포함)를 검출하고 유형이 일치하고, 경계선 내부의 화소수가 50%이상 겹치면 객체당 0, 겹치지 않으면 1로, 겹치는 비율을 정확도로 계산을 하여, 유형별로 정밀도(정확도, error율)를 계산하고 이의 평균값(mAP)으로 평가.

4.4 영상 객체간(혹은 복수 객체) 상호 관계 인식

테스트 영상 450매에 대하여 실험을 하여, 영상속의 객체간(혹은 복수 객체간) 상호 관계(상/하, 좌/우)가 일치하면 0, 일치하지 않으면 1로 계산을 하여, 누적된 값은 error율(%)로 계산.

4.5 영상 구문 검출 및 인식 기술 성능표

<표 1>은 10개의 유형에 대한 영상 구문 검출 및 인식 기술의 성능을 나타내는 표이다.

<표 1> 영상 구문 검출 및 인식 기술의 성능표

영상의 유형	영상 유형 인식	영상 객체 추출 및 인식(외접 사각형)	영상 객체 추출 및 인식(경계선)	영상 객체(혹은 복수 객체) 상호 관계 인식
(1) 자전거를 타고 있는 사람	error율(%)	정밀도(정확도/%, error율/%) 및 평균 정밀도(mAP)	정밀도(정확도/%, error율/%) 및 평균 정밀도(mAP)	error율(%)
(2) 자전거를 타고 있는 사람들	//	//	//	//
(3) 모터사이클을 타고 있는 사람	//	//	//	//
(4) 모터사이클을 타고 있는 사람들	//	//	//	//
(5) 노인용 전동 스쿠터를 타고 있는 사람	//	//	//	//
(6) 노인용 전동 스쿠터를 타고 있는 사람들	//	//	//	//
(7) 노인용 보행기를 밀고 가는 사람	//	//	//	//
(8) 노인용 보행기를 밀고 가는 사람들	//	//	//	//
(9) 기타(전동 베달자, 전동휠, 리보드 등)를 타고 있는 사람	//	//	//	//
(10) 기타(전동 베달자, 전동휠, 리보드 등)를 타고 있는 사람들	//	//	//	//

<표 2>는 전체 실험 영상 DB에 대한 영상 구문 검출 및 인식 기술의 성능을 나타낸다.

<표 2> 전체 기술의 성능 평가표

구분	세부 기능	성능
구문 검출 및 인식 성능	평균 구문 검출 및 인식 정확도(%)	
영상 속성값의 경량성	영상 속성값의 크기(Byte)	
객체 인식 복잡도	영상 속성값 추출 시간(msec)	
	평균 구문 추출 및 인식 시간(sec)	

5. 결론

본 논문에서는 자동차 전방 감시를 위한 영상 구문(Visual Phrase) 검출 및 인식 기술의 소개와 기술의 성능 평가를 위한 표준 영상 DB의 선정과 평가 기준에 대해 설명하였다.

본 논문에서 설명한 기술은, 영상 구문 검출 및 인식의 관련 기술을 연구하고 있는 대학(원)의 학생, 연구 분야에서 관련 연구를 수행하고 있는 연구자들, 산업 현장에서 본 기술을 활용하여 응용 제품을 개발하고 있는 현장의 기술자들에게 객관적인 기술 평가와 제품 도입을 위한 중요한 수단으로 활용 될 수 있을 것으로 사료된다.

참고문헌

- [1] Bryan A. Plummer, Arun Mallya, Christopher M. Cervantes Svetlana Lazebnik, "Phrase Localization and Visual Relationship Detection with Comprehensive Image-Language Cues," Proceedings of the IEEE International Conference on Computer Vision, pp. 1946-1955, October(2017)
- [2] Mohammad Amin Sadeghi1, Ali Farhadi, "Recognition Using Visual Phrases," Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1745-1752. June (2011)
- [3] Yu Kong and Yun Fu, "Human Action Recognition and Prediction: A Survey," JOURNAL OF LATEX CLASS FILES, VOL. 13, NO. 9, SEPTEMBER(2018)
- [4] 2018년판 교통사고 통계 요약(<http://taas.koroad.or.kr>)

 This work was supported by Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIP) (No. 2018-0-00198), Object information extraction and real-to-virtual mapping based AR technology)