

비주얼 검색 알고리즘 성능분석

*제성관 **나상일 ***오원근

한국전자통신연구원

*skj@etri.re.kr

The Test of existing Visual Search Algorithms with the ETRI dataset

*Sung-Kwan Je **Sang-Il Na ***Weon-Geun Oh

Electronics and Telecommunication Research Institute

요약

최근 모바일환경에서 영상검색을 위한 연구가 활발히 진행됨에 따라 MPEG에서도 국제표준을 위한 연구가 진행되고 있다. 그러나 모바일의 제한된 성능환경과 외부적인 실외환경변화에 강인한 알고리즘개발이 시급하다. 따라서 본 논문에서는 다양한 환경변화에 강인한 영상검색 알고리즘을 개발하기 위하여 기존의 영상검색으로 많이 사용되고 있는 SIFT와 SURF 알고리즘의 성능을 분석하고자 한다. 실험에 사용된 데이터베이스는 MPEG에서 표준영상으로 채택된 ETRI dataset을 이용하였다. 실험결과, 기존의 알고리즘은 특징점 검출과 그 서술자가 서로 의존적임을 알 수 있었다.

1. 서론

영상검색(visual search)에 대한 관심이 증대하면서 최근에는 모바일환경에서의 영상검색을 하기 위해 간략(compact)하면서도 기기간 호환을 고려한 CDVS(Compact Descriptor Visual Search)에 대한 연구가 활발하게 진행되고 있다. 그 중 국제표준단체인 MPEG에서는 CDVS 표준에 대한 요구조건이 담긴 Call for Proposal과 경쟁을 위한 실험을 위해 필요한 Evaluation Framework가 정식으로 발표되었다. CDVS 표준을 위해서는 다음과 같은 사항이 보장되어야 한다[1].

- 영상 검색 응용 프로그램과 데이터베이스의 상호운용성
- 높은 수준의 구현 성능
- 영상 검색 응용 프로그램을 위한 추출과 정합의 간결한 설계
- 모바일 장치에서의 서술자의 추출과 정합의 하드웨어 구현 용이성
- 영상 검색을 위한 무선 네트워크 사용 부하 축소

이러한 요구사항을 만족하기 위한 알고리즘 개발을 위해 본 논문에서는 기존의 특징 기반 물체인식에 대한 연구 중 조명에 강인하고 이미지의 크기와 회전에도 일정한 특징점을 추출하는 SIFT(Scale Invariant Feature Transform) 알고리즘[2]과 SURF(Speed Up Robust Feature)[3] 알고리즘의 성능을 MPEG에서 채택된 ETRI dataset을 이용하여 분석하고자 한다.

SIFT와 SURF 알고리즘은 다음과 같다. SIFT 알고리즘은 특징점을 추출하기 위해 크게 세단계로 나뉜다. 첫 번째로 스케일 공간상의 후보 특징점을 추출하고 후보 특징점의 안정성을 검사하여 안정된 특징점의 위치를 세부 위치로 보정한다. 두 번째로 특징점에 기준 방향과 크기를 할당한다. 세 번째로 기준 방향을 중심으로 특징점 주변 영역의 국부영상에서 특징점 서술자를 생성한다. 이후 가우시안 가중치가 부여된 이미지 그래디언트는 각각의 배열이 8방향을 가지고 있는 4X4

배열로 재구성된다. 이러한 방향성 히스토그램은 128(8X4X4)차원의 벡터로 구성되며 조명에 강인한 서술자를 만들기 위해 정규화 과정을 거치며 이렇게 만들어진 128차원 벡터는 특징점을 대표한 서술자가 된다. SURF 알고리즘은 특징점 추출과 서술자 생성의 두단계로 구성된다. 특징점을 추출하기 위해 영상을 적분영상으로 변환하여 고속 헤시안 검출기(Fast Hessian Detector)를 이용한다. 이때 크기 변화에 불변하는 특징을 얻기 위해 가우시안 2차 미분이 근사화된 사각필터의 크기를 변환시켜 특징점을 추출한다. 그리고 회전에 강인하기 위해 특징점에 대한 방향 정규화를 한다. 특징점을 중심으로 4X4 크기의 세부 영역을 구성하고 각 세부영역에서 하르 웨이블릿(Haar Wavelet)을 이용하여 32, 64, 128차원의 서술자를 구성한다.

2. 실험결과

2.1 실험환경

알고리즘의 성능을 평가하기 위해 사용된 ETRI dataset은 밝기변화(주간, 야간) 조명변화(illumination change), 모션 블러(motion blur) 등과 같은 다양한 변화에 강인성을 테스트하기 위한 실외영상들로 구성되어 있다. 실험영상에 대해서 2가지 방법을 이용하였다. 첫째, 영상에 대해 각 알고리즘별로 재현율(recall)과 정확도(precision)을 계산하여 성능을 분석하였다. 둘째, 각 알고리즘에서 사용된 특징점 검출을 서로 변경하여 재현율과 정확도를 계산하여 성능을 분석하였다. 특징점 검출을 변경하여 테스트함으로써 각 알고리즘의 서술자가 특징점에 얼마나 독립적인지 알 수 있다. 재현율과 정확도는 정보인출(information retrieval)에서 중요한 성능평가 측정 기준으로 재현율은 질의영상과 관계되는 영상을 전체 영상에서 몇 개를 검색하였는지 평가하며, 정확도는 검색결과 중 실제 관계되는 영상이 몇 개인지를 평가하는 것이다.

2.2 실험결과

각 알고리즘의 성능분석을 위하여 ETRI dataset 1500장을 테스트하였으며 그 결과는 표 1과 표 2에 나타내었다.

표. 1 각 알고리즘의 NN-Ratio 결과

Detector	SIFT detector	
	SIFT descriptor NN-Ratio : 0.49	SURF descriptor NN-Ratio : 0.65
TRUE	52969	25395
FALSE	159829	20336
MISS	83120	110695
RECALL	0.389223229	0.186604453
PRECISION	0.248916813	0.555312589

Detector	SURF detector	
	SURF descriptor NN-Ratio : 0.49	SIFT descriptor NN-Ratio : 0.65
TRUE	21400	44878
FALSE	114779	36044
MISS	105005	91211
RECALL	0.169297101	0.329769489
PRECISION	0.157146109	0.554583426

표. 1은 각 알고리즘의 NN-ratio 기본값을 0.49, 0.65로 두어 테스트한 결과이다. 재현율과 정확도를 계산하기 위해서 "True", "FALSE", "MISS"를 수동으로 산출하였다. "TRUE"는 검색된 결과가 일치하는 값이고, "FALSE"는 검색된 결과가 일치하지 않았을 때 값이며, "MISS"는 검색된 결과에 관계되는 영상이 부재한 값이다. 표. 1에서 살펴본 바와 같이 배경이 복잡한 영상일 때 그 결과가 너무 낮게 나타남을 알 수 있었다. 그리고 각 알고리즘의 특징점 검출알고리즘을 변경하여 그 성능을 평가한 결과 특징점검출에 대한 NN-distance 차이에 의해 그 값이 차이가 남을 알 수 있었다.

표. 2는 NN-ratio 값을 변경하여 실험한 결과이다. 표. 2에서 살펴본 바와 같이 NN-ratio값을 변경했을 경우 그 성능도 변경됨을 알 수 있었다. 따라서 기존의 알고리즘인 SIFT나 SURF의 경우, 알고리즘에서 제안된 특징점 검출기법과 그 서술자가 서로 의존적임을 알 수 있었다.

표. 2 각 알고리즘의 변경된 NN-Ratio 결과

Detector	SIFT detector	
	SIFT descriptor NN-Ratio : 0.65	SURF descriptor NN-Ratio : 0.78
TRUE	59209	46094
FALSE	154700	182621
MISS	76880	89995
RECALL	0.435075576	0.338704818
PRECISION	0.276795273	0.201534661

Detector	SURF detector	
	SURF descriptor NN-Ratio : 0.65	SIFT descriptor NN-Ratio : 0.78
TRUE	35675	62663
FALSE	165950	143544
MISS	10414	73426
RECALL	0.262144626	0.460456025
PRECISION	0.176937384	0.303883961

Acknowledgement

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2012년도 콘텐츠 산업기술지원사업의 연구결과로 수행되었음

참고문헌

- [1] N11674, Compact Descriptors for Visual Search: Draft Call for Proposals, Requirements Subgroup, MPEG output document, Oct. 2010, Guangzhou, CN.
- [2] D. G. Lowe, Object recognition from local scale-invariant features, in Proceedings of International Conference on Computer Vision, pp.1150-1157, Corfu, September, 1999.
- [3] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008