

모바일 비주얼 검색: 기술과 표준화 동향

Technology and Standardization Trend of Mobile Visual Search

| | |
|------------------|-------------------|
| 이승재 (S. J. Lee) | 콘텐츠저작권보호연구실 선임연구원 |
| 이근동 (K. D. Lee) | 콘텐츠저작권보호연구실 연구원 |
| 나상일 (S. I. Na) | 콘텐츠저작권보호연구실 연구원 |
| 제성관 (S. K. Je) | 콘텐츠저작권보호연구실 선임연구원 |
| 정다운 (D. U. Jung) | 콘텐츠저작권보호연구실 연구원 |
| 오원근 (W. G. Oh) | 콘텐츠저작권보호연구실 책임연구원 |
| 서영호 (Y. H. Seo) | 콘텐츠저작권보호연구실 실장 |
| 손욱호 (W. H. Son) | 콘텐츠플랫폼연구부 부장 |

* 본 연구는 미래창조과학부 산업융합원천기술개발사업(정보통신)의 일환으로 수행 하였음
(R2012030111, UVD(Unified Visual Descriptor) 기반 Smart Mobile Search 기술개발).

모바일 비주얼 검색(MVS: Mobile Visual Search) 기술은 스마트폰 환경에서 질의 영상의 특징을 검출하고 그 주변 영역의 정보를 이용하여 추출한 서술자를 미리 구축된 영상정보 DB와 비교하여 검색하는 기술로 스마트폰의 보급과 네트워크 환경발달에 따라 응용서비스의 범위와 가능성이 확대되고 있다. 본 동향에서는 기존의 텍스트, 음성, QR코드 등의 검색을 넘어 인간이 보는 그대로를 검색하고자 하는 모바일 비주얼 검색의 기술 및 표준화 동향을 분석하고 향후 전망에 대해서 살펴보고자 한다.

차세대 콘텐츠기술 특집

- I. 서론
- II. 모바일 비주얼 검색 기술
- III. 관련 기술동향
- IV. 표준화 동향
- V. 결론

I. 서론

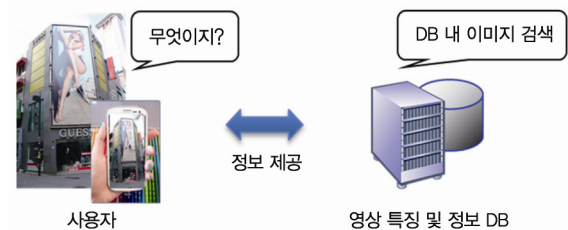
스마트폰 보급 확대와 네트워크 환경발달은 영상 데이터의 급속한 증가로 이어지고 있으며, 모바일기기를 통한 영상정보의 검색도 증가하고 있다. 페이스북에는 매일 300백만장의 사진이 업로드 되고 있으며, 유튜브에는 한 사람이 1,700년동안 시청해야 다 볼 수 있는 비디오가 업로드되어 있다[1]. 모바일 데이터의 트래픽 증가율은 연평균 66%에 달하며, 이중 스마트폰은 67.5%를 차지하고, 2017년에는 전체 세계 모바일 트래픽에 93%를 차지할 것으로 전망된다[2]. 검색의 어려움을 측정하는 평균 검색어 수가 지속적으로 증가하고 있으며, 좀 더 편리하고 정확한 검색 방법에 대한 수요가 증가하고 있다[3]. 특히 영상 기반의 정보검색은 사용자가 명칭이나 키워드를 모를 경우 사용자의 편의성을 만족시킬 수 있어 그 효과가 크다. 이에 따라, 세계 각국의 IT 기업들은 영상 기반 정보검색기술을 차세대 성장산업으로 보고, 기술 및 서비스개발을 하고 있으며, 특히, 모바일 환경에 적합한 모바일 비주얼 검색기술에 집중하고 있다[4]. 구글의 'Goggles'[5], 아마존의 'Flow'[6], 마이크로소프트의 'Bing Vision'[7], 쉐컴의 'Vuforia'[8] 등의 서비스가 출시되었다. 이와 함께 관련 기술에 대한 표준화도 MPEG을 통해 진행되고 있다.

본 동향에서는 이러한 MVS(Mobile Visual Search) 기술에 대한 기술동향과 서비스 및 표준화 동향을 살펴보고, MVS 기술의 발전방향 및 향후 전망에 대해서 살펴보고자 한다.

II. 모바일 비주얼 검색 기술

1. 개요

본 절에서는 모바일 비주얼 검색기술의 개요, 요구사항과 함께 기술 구성요소 및 서비스 시나리오에 대해서



(그림 1) 모바일 비주얼 검색기술 개념도: 사용자는 검색 대상을 촬영하여 검색을 요청하면 DB 내에 이미지를 검색하고, 해당 이미지에 대한 정보를 제공

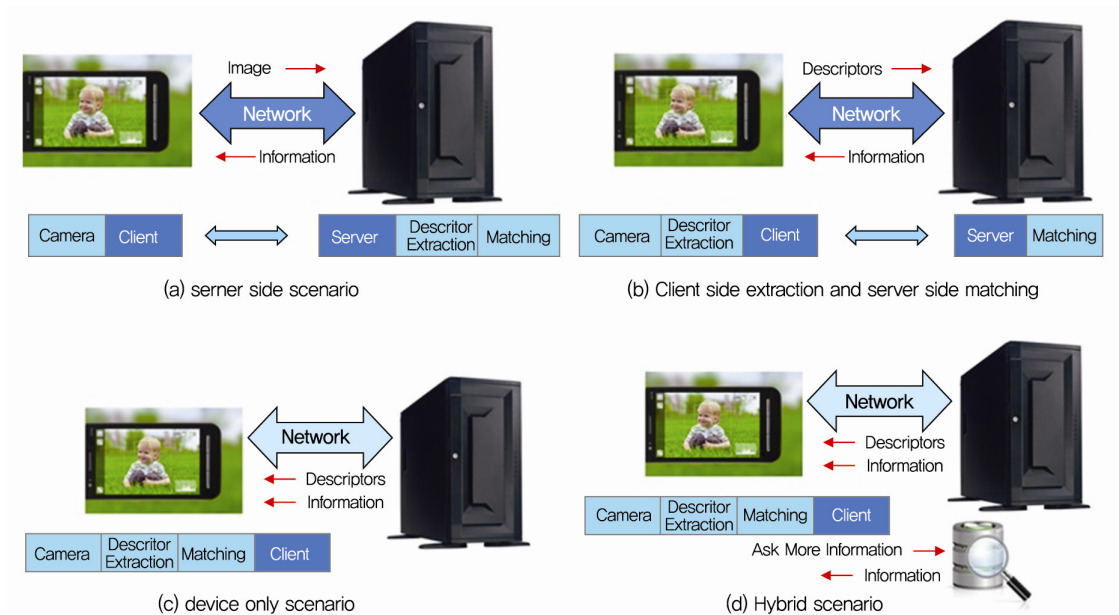
전반적으로 살펴본다.

모바일 비주얼 검색기술의 개념도는 (그림 1)과 같다. 사용자는 검색하고자 하는 대상을 촬영하거나 촬영된 이미지를 활용하여 미리 구축된 영상 DB 검색을 통해 촬영 대상과 연계된 정보를 얻게 된다. 예를 들어 상품의 경우, 검색하려는 상품을 촬영하고 이를 DB에서 검색하여 상품명, 가격, 기능 및 판매처 등 부가정보를 얻을 수 있다. 혹은 거리의 건물을 촬영하고, 건물을 인식하여 건물 내의 정보나 주변정보를 제공받을 수 있다.

이러한 MVS 기술은 크게 특징 추출과 검색의 단계로 이루어지며, MVS 응용서비스에 있어 특징 추출과 검색의 주제에 따라 (그림 2)와 같이 4가지의 시나리오가 가능하다. 응용서비스에 따라 특징 추출과 검색이 서버에서 모두 이루어질 수도 있고, 특징 추출은 사용자의 단말에서 검색은 서버에서 이루어 질 수 있다.

이러한 다양한 서비스 시나리오를 만족시키기 위해 MVS 기술은 아래와 같은 요구사항을 만족시켜야 한다[9].

- 강인성(robustness): 조명 및 환경변화에 강인해야 하며, 촬영시점의 변화에도 강인해야 함. MVS 기술은 실내외를 포함한 왜곡에 대한 강인성을 지원할 수 있어야 함.
- 구별성(pairwise independence): 서로 다른 객체를 포함한 영상 특징은 구별되어야 함.
- 일반성(generality): 하나의 특징으로 여러 응용분야에 활용이 가능해야 함. 예를 들어 CD 및 책 표지를 위한 기술은 건물이나 다른 상품검색에 활용



(그림 2) 특징 추출과 검색에 따른 MVS 서비스 시나리오

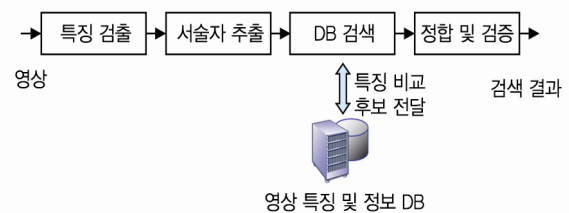
될 수 있어야 함.

- 계층성(scalability): 네트워크환경 및 다양한 단말을 지원할 수 있도록 특징 크기를 계층적으로 구성하고 상호호환성(interoperability)을 확보하여야 함.
- 검색 효율성(DB efficiency): 대용량 영상데이터를 효율적으로 검색할 수 있는 구조를 지원해야 함.
- 하드웨어 구현지원(support to hardware implementation): 고속 특징 추출과 검색을 위한 하드웨어 구현을 지원해야 함.

2. 요소기술 및 서비스 시나리오

모바일 비주얼 검색 기술은 일반적으로 특징 검출에 기반한 서술자 추출방법을 사용하고 있으며, 이러한 기술은 (그림 3)과 같은 요소기술로 구성된다. 각 요소기술의 역할은 아래와 같다.

- 특징 검출(feature detection): 영상에서 시점, 환경 변화 등의 왜곡에 불변하는 영역 혹은 점을 검출하는 과정



(그림 3) MVS 기술의 구성요소

- 서술자 추출(descriptor extraction): 검출된 특징 주변영역을 이용하여 강인하며, 구별 가능한 특징 벡터를 추출하는 과정
- DB 검색 (database search): 추출된 특징들로 구성된 DB에 대해서 정합 및 검증을 위한 후보군을 선정하는 과정
- 정합 및 검증(matching and verification): 후보에 대해서 정합 유무를 판별하고, 정합이 이루어진 영상에 대해서 기하학적 검증을 수행하는 과정

모바일 비주얼 검색기술의 각 요소기술에 대한 다양한 방법들이 제안되고 연구되었으며 <표 1>과 같이 다양한 응용서비스에 활용될 수 있다.

〈표 1〉 MVS 응용서비스의 예

| 분야 | 응용서비스 |
|-------|-------------------------------------|
| 정보검색 | 검색대상 촬영을 통한 정보 검색서비스 |
| 증강현실 | MVS 기술을 적용한 영상과 3D 객체 간의 연동을 통한 서비스 |
| 로봇 비전 | MVS 기술을 로봇 비전에 적용하여 사물 인식에 활용 |
| TV | TV 콘텐츠 내의 상품이나 건물 검색서비스 |
| 자동차 | 자동차 주행 중 안전을 위한 위험물 감지 및 중요 표지판 식별 |
| 영상 감시 | 객체 감시 및 이벤트 감지서비스 |
| 콘텐츠관리 | 콘텐츠 내 객체 검색 및 분류서비스 |

III. 관련 기술동향

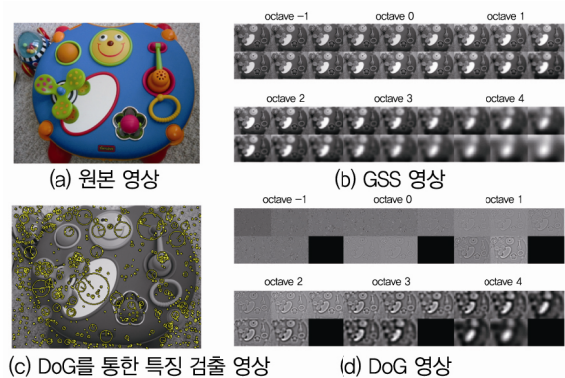
본 장에서는 모바일 비주얼 검색기술의 주요 기술 및 관련 서비스에 대한 국내외 동향을 살펴본다.

1. 주요 기술동향

가. 특징 검출(feature detection)

모바일 비주얼 검색기술에서 가장 중요한 요소기술 중 하나인 특징 검출 기술은 영상 내에서 시점이나 조명 변화에 불변한 영역을 찾는 것으로 MVS 기술에 핵심 부분이라 할 수 있다. 이러한 특징 검출에 대해서는 강인성뿐만 아니라 계산의 효율성, 구별성 등이 요구된다[10].

일반적인 특징 검출을 위해서는 영상 변형에 불변성을 가지는 위치를 선택하도록 설계한다. 이러한 특징 검출의 예로는 Harris 검출기[11], Hessian 검출기[12], Harris-Laplace 검출기[13], SIFT(Scale Invariant Feature Transform) 검출기[14], MSER(Maximally Stable Extremal Regions) 검출기[15], SURF(Speeded Up Robust Features) 검출기[16], FAST 검출기[17]~[19] 등이 있다. 본 동향에서는 성능이 우수하다고 알려진 DoG(Difference of Gaussian) 기반의 SIFT 검출기와 빠른 속도로 인해서 많이 활용되고 있는 SURF 검출기, FAST 검출기에 대해서 살펴본다.



(그림 4) DoG 기반 SIFT 검출기의 예

1) SIFT 검출기

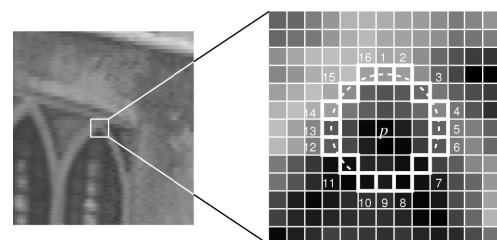
SIFT 검출기[14]는 (그림 4)와 같이 Gaussian Scale Space에서 Laplacian 필터의 응답에 대한 지역 극점을 찾는 방법으로 LoG(Laplacian of Gaussian)의 계산량을 DoG를 통해 줄이는 방법이다. 하지만, 특징 검출에 있어 DoG를 사용하는 부분이 특허로 등록되어, 청구 1항의 'difference of images' 부분, 즉 DoG를 특징 검출 시 회피해야 한다.

2) SURF 검출기

SURF 검출기[16]는 적분 이미지(Integral Image)와 박스 필터(Box filters)를 활용하여 Hessian의 근사값으로 극점을 찾고, 영상의 크기를 줄이는 대신 필터의 크기를 키워 Scale Space를 계산하는 속도를 개선하였다. SURF 검출기 역시 특허로 인해 사용제한이 있다.

3) FAST 검출기

FAST 검출기[17]~[19]는 (그림 5)와 같이 Bresenham의 알고리즘 기반으로 그린 원 위의 픽셀값과 중심의 픽셀값 사이의 관계를 이용해서 빠른 특징 추출을 하



(그림 5) FAST 검출기에서 검출 영역 판별을 위한 계산 영역[18]

는 방법이다. ID3(Iterative Dichotomiser 3)를 이용한 결정 트리를 통해 성능을 향상시켰으며[18], 더 많은 픽셀과의 비교를 통한 성능을 개선한 버전[19]도 있다. 관련 소스코드가 공개되어 있고, BSD(Berkeley Software Distribution) 라이선스로 많은 응용에 활용되고 있다.

나. 서술자 추출(descriptor extraction)

특징 검출에 의해서 시점 및 조명 변화 등에 불변한 영역을 찾은 후에는 영상의 지역정보를 이용하여 구별이 가능하도록 서술자를 추출한다. 서술자 추출방법으로는 대표적인 SIFT 특징[14]과 SURF[16]과 이진특징을 이용한 LBP(Local Binary Pattern)[20][21], BRIEF(Binary Robust Independent Elementary Features)[22], BRISK(Binary Robust Invariant Saleable Key points)[23], LIOP(Local Intensity Order Pattern for Feature Description)[24] 등이 있다. 이 중에서 SIFT와

SURF, LBP, LIOP를 살펴본다.

1) SIFT

SIFT[14] 서술자 추출은 검출된 특징정보를 기반으로 해당 스케일에서의 영상의 패치를 얻어 4 by 4 영역을 나누고 각 영역에 대해서 8단계로 구성된 각도 히스토그램을 얻는다. 이 값을 모아 128차원의 특징 벡터로 구성하고, 밝기 변화에 강인하도록 정규화한다.

2) SURF

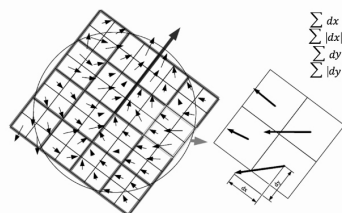
SURF[16] 서술자 추출은 (그림 6)과 같이 특징 검출 영역의 패치에 대해서 4 by 4영역을 할당하고, 각 영역을 4영역으로 나누어서 그림과 같이 특징 벡터로 구성한다. 그러므로 서술자 차원은 64차원이 된다.

3) LBP

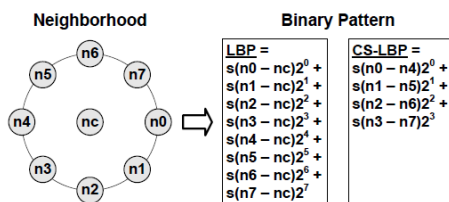
LBP 영상의 텍스처를 인식하기 위해서 처음 제안된 것[20]으로 영상의 중심 픽셀과 주변 픽셀 간의 차 정보를 이용해서 생성한 값을 서술자로 활용한다. (그림 7)처럼, 기본적인 LBP와 CS-LBP(center-symmetric LBP)를 활용 가능하다.

4) LIOP

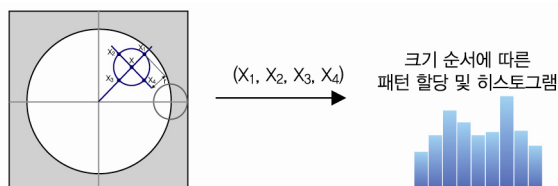
LIOP[24]는 (그림 8)처럼 서술자 추출 영역에서 서술자 추출 시 중심영역에서 떨어진 주변의 밝기정보의 순서를 이용한다. 예를 들어 4개의 주변 영역을 선택하는 경우 총 24개의 경우에 수에 대해서 히스토그램을 계산하고 이 값을 서술자로 사용하는 방법이다.



(그림 6) SURF 서술자 추출 과정[16]



(그림 7) LBP 및 CS-LBP 서술자 추출과정[21]



(그림 8) LIOP의 서술자 추출과정[24]

다. DB 검색(database search)

모바일 비주얼 검색기술이 실질적으로 활용되기 위해서는 강인한 특징 검출과 서술자 추출 및 구현[25] 외에도 대용량의 서술자를 효율적으로 검색하는 방법이 반드시 필요하다. 대표적인 검색 구조로는 VT(Vocabulary Tree)[26]와 REVV(Residual Enhanced Visual Vector)[27], VLAD(Vector of aggregated local descriptors)[28] 및 FV(Fisher Vector)[29]를 기반으로 한 방법들이 사용되

고 있다. 본 동향에서는 BoW(Bag of Words) 기반의 VT와 VLAD에 대해서 살펴본다.

1) Vocabulary Tree

영상 서술자 값들을 BoW에서의 단어의 개념으로 생각하고 검색의 대상이 되는 영상으로부터 Vocabulary Tree를 구성한다. 이 트리를 기반으로 검색 시에는 각 비주얼 워드(visual word), 즉 특징의 빈도에 따라 후보 군을 생성하고, 정합 및 검증방법에 의해서 결과를 결정한다.

2) VLAD

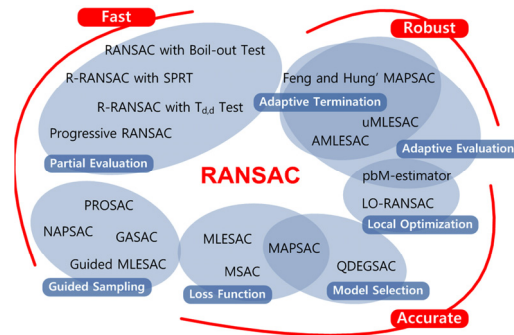
VT 기반의 방법은 비주얼 워드의 빈도수만을 이용하고 있어, 정보를 표현하는데 있어 제약이 있다. 이러한 점을 개선하기 위해서 해당 비주얼 워드를 선택한 특징들의 의 통계적 특성을 반영하는 방법이 제안되었다. VLAD에서는 평균 정보를 활용하고 있다. 이 방법은 Fisher Kernel의 일종으로 생각할 수 있으며, REVV도 VLAD의 변형으로 생각할 수 있다.

라. 정합 및 검증(matching and verification)

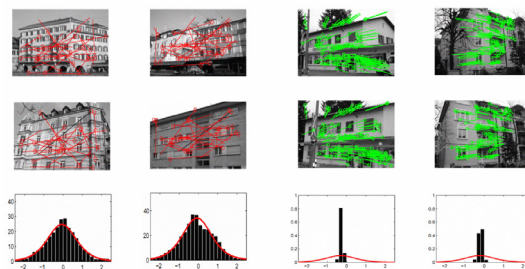
정합 및 검증은 특징으로 구성된 DB에서 후보 영상들에 대해서 실제 일치하는 정보가 있는지 확인하는 과정으로 서술자 값과 위치 정보를 이용하여 기하학적인 관계를 추정한다. 일반적으로 RANSAC(RANdom SAmple Consensus)[30]이란 방법이 활용된다. 최근에는 확률 모델을 기반으로 한 방법도 제안되어 효율성을 높이고 있다.

1) RANSAC

RANSAC은 모델을 가정하고, 그 모델이 맞는지 반복적으로 검증하는 방법으로 모델을 따르는 확률(p)과 모델을 풀기 위한 점의 개수(k) 및 반복 횟수(n)에 의해 모델 검증 확률을 계산할 수 있다. RANSAC 관련 기술은 (그림 9)와 같이 다양하며, 영상 정합에서는 임의의 4쌍의 점을 선택하고, 영상 변환을 모델링해서 정합과 및 검증을 하게 된다. 이 과정에서 RANSAC의 경우 모



(그림 9) RANSAC과 관련 기술현황[31]



(그림 10) 비정합쌍과 정합쌍 간의 분포 차[33]

델을 따르는 확률(p)이 낮을 경우 많은 시간이 소요되므로 SIFT에서 제안한 것처럼, 특징 간의 유사도를 기반으로 점의 개수를 줄인다.

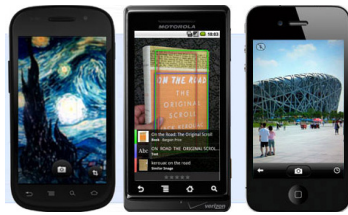
2) 통계 모델 기반 정합 방법

정합 여부를 판단하기 위해서 RANSAC를 통한 방법 외에 (그림 10)과 같이 정합쌍과 비정합쌍 간의 통계 모델을 통한 방법도 제안되었다[32][33]. 정합 후보쌍의 로그 거리비율에 따라 정합 유무를 판단한다. 그림과 같이 정합쌍과 비정합쌍 간에는 통계적으로 차이가 발생한다. 이러한 정보를 활용하여 동일한 개수의 점에 대해서 RANSAC보다 빠르게 정합 여부를 확인할 수 있다.

2. 관련 서비스동향

가. 구글

구글은 'goggles'라는 (그림 11)과 같은 이미지 인식 응용프로그램을 2010년 10월 출시하여 서비스하고 있다. 현재 동물, 식물, 자동차, 가구, 의류 등은 검색이



(그림 11) 구글의 'goggles' 서비스의 예[5]

어려우며, 최대 3개까지의 물체를 인식 가능하다.

나. 아마존

아마존은 자사의 온라인 쇼핑 서비스와 MVS 기술의 상승효과를 기대하고 2009년 영상 검색 기업인 'Snap Tell'을 인수하였다. 이후 (그림 12)의 상품 검색 앱, 'Flow', 신발 전용 검색 앱, 'Fabulous'등을 출시 서비스 하고 있다. 특히 A9.com이라는 별도의 회사를 두고 MVS 기술을 미래 수종사업으로 선정 다양한 기술을 개발 중이다.

다. 쉐컴

쉐컴은 MVS 기술을 활용한 'Vuforia'[8]라는 응용서비스 개발 플랫폼을 제공하고 있다. 현재, 영어 문자 인

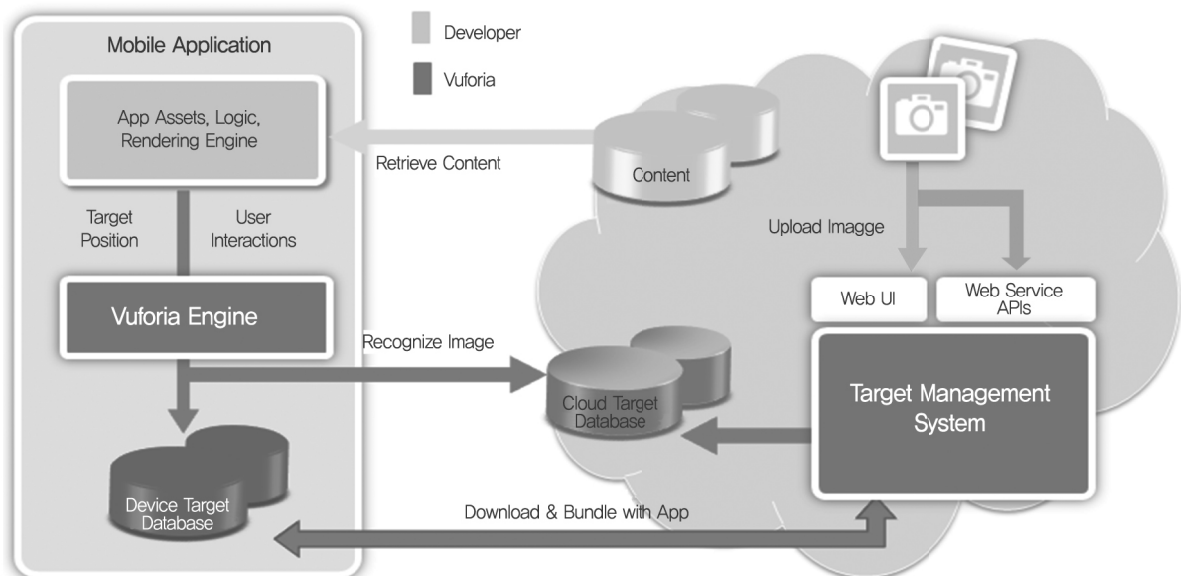


(그림 12) 아마존의 상품 검색 서비스 'Flow'[6]

식, 3차원 객체, 영상 인식 등을 지원하고 있다. 안드로이드, iOS를 모두 지원하는 개발자 환경 및 클라우드 서비스를 제공하고, 서비스 개요도는 (그림 13)과 같다. 현재 등록된 개발자만 130여개국 8만명을 넘으며, 플랫폼을 활용한 앱이 6,500개를 넘고 있다.

라. HP

HP는 클라우드 기반 'Multimedia Analytic Platform' 서비스[34]를 개발하였으며, 이 서비스를 통해서 얼굴인식, 특징 추출 및 영상 관련 API(Application Programming Interface)를 손쉽게 활용하여 관련 서비스를 개발할 수 있다.



(그림 13) 쉐컴의 'Vuforia' 서비스 개요도[8]

마. 제록스

제록스는 영상 검색에 대한 다양한 기술을 개발하고 있으며, Open Xerox[35]를 통해서 다양한 서비스를 공개하고 있으며, 대용량 영상 검색에 대해서 기술을 선도하고 있다.

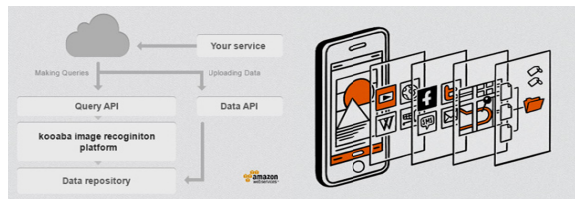
- Similar Image Search: 영상 및 키워드 기반의 유사 영상 검색
- Image Categorizer: 706개의 다른 분류에 따라 영상 분류
- Aesthetic Image Search: (그림 14)와 같이 영상에 대한 심미적 기준에 따른 이미지 검색

바. 마이크로소프트

마이크로소프트는 노키아 인수를 통해서 MVS 관련 'Point& Find'서비스 기술을 확보하였으며, 윈도우폰에 QR코드, 바코드, CD, 책표지 등을 검색할 수 있는 'Bing Vision'앱을 제공 중이다.

사. Kooaba

Kooaba[36]는 ETH에서 시작한 모바일 영상 검색 회

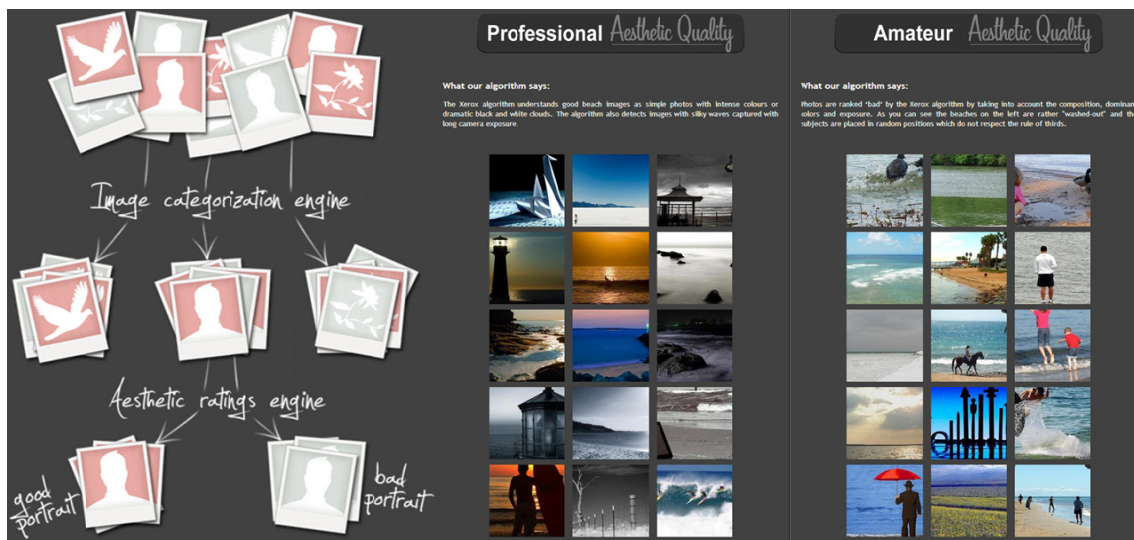


(그림 15) Kooaba의 플랫폼 및 서비스[36]

사로 6천만장이 넘는 영상에 대한 검색 서비스를 제공한다. MVS 기반 응용 서비스를 위한 API 기능을 (그림 15)와 같이 제공하며, 다양한 관련 앱을 개발하였다.

IV. 표준화 동향

모바일 비주얼 검색기술에 대한 표준화는 현재 MPEG-7 CDVS(Compact Descriptors for Visual Search)로 진행되고 있다. CDVS의 목적은 다양한 목적과 응용환경에 맞도록 강인하며, 작고 효율적인 서술자 추출과정을 표준화하는 것으로 제 92회부터 제 96회 회의까지 응용 서비스 시나리오, 표준화의 범위 및 요구사항과 평가 방법이 논의 되었고[37]~[40] 제 97회 회의에서 'Call for Proposal'이 공지되었으며[41] 제 106회까지 제안 기술에 대한 평가가 지속적으로 이루어지고 있다.



(그림 14) 제록스 'Aesthetic Image Search' 데모[35]

1. 목적과 범위

CDVS는 모바일 검색 응용서비스를 위한 설계가 가능하도록 최대한의 성능과 응용 환경에 대한 호환성을 보장하며, 최소한의 크기 및 효율적인 구현과 검색을 목표로 한다. 표준화의 범위는 추출된 서술자의 비트 스트림과 상호호환성이 필요한 서술자 추출 과정을 포함한다 [39].

2. 주요 요구조건 및 실험방법

CDVS는 강인성을 비롯하여 다양한 조건들[9]을 만족시켜야 한다. 이러한 요구 조건을 평가하기 위해서 ZuBud, UKBench, Stanford, ETRI, PKU, Telecom Italia, Telecom SudParis, Huawei의 데이터셋을 모아 총 30,256장의 사진, mixed text and graphics, painting, video frame, landmark, common object로 5개의 평가 DB를 <표 2>와 같이 구성했다[37].

제안된 기술의 성능은 정합 및 검색 성능실험을 통해 검증된다. 정합 성능을 위해서는 187,846장의 영상쌍을 이용하여, 최소 서술자 크기 512byte에서 최대 크기

16kbyte까지 정합율(TPR: True Positive Rate)과 오정합율(FPR: False Positive Rate)을 계산한다. 오정합율은 1% 이내를 목표로 한다. 검색 성능검증을 위해서는 검색 영상과 별도의 100만장 영상을 구성하여, 11,313장의 질의영상(검색 영상의 변형)을 통해 평균 정확도(mAP: mean Average Precision)과 Top Match값을 비교한다. 또한, 정합 성능과 함께, 위치 추정 정확도(Localization Accuracy)를 측정한다. 모든 실험결과들은 상대 기관과의 교차 실험을 통해서 검증되어야 한다.

3. 주요 기술현황

현재 106회까지 진행된 CDVS는 기존의 영상특징을 압축하고, 다양한 응용환경을 지원하기 위해 검출된 특징의 선별, 좌표 압축, 서술자의 양자화 과정을 포함한다. 이를 통해 CDVS에서는 512byte~16kbyte(512 byte, 1k, 2k, 4k, 8k, 16k)의 다양한 크기의 서술자 추출을 지원한다. 또한, 대용량 검색을 위해 효율적인 검색 구조가 제안되어 테스트 중이다. 특징 검출에 있어 기존의 특징 검출기, 예를 들어 DoG 기반의 SIFT 검출기의 특허를 회피하기 위해서 노력 중이며, 제 106회에 기존의 특허를 회피하는 검출 방법[42]으로 채택되었다.

4. 향후 일정

향후 MPEG-7 CDVS에서는 구현의 효율성을 위한 메모리 및 계산량의 최적화가 다루어질 예정이다. 또한, 검색 성능의 안정화 및 상호호환성에 대한 검증을 통해 필수적 요소만을 표준화할 예정이다.

VI. 결론

본 동향에서는 모바일 비주얼 검색의 기술 및 표준화 동향에 대해서 살펴보았다. 모바일 비주얼 검색기술은 영상특징을 이용하여 직관적이고 편리한 검색을 제공할

<표 2> MPEG-7 CDVS 표준 평가 DB

| 카테고리 | 영상 개수 | 영상의 예 |
|-------------------------|--------|---|
| Mixed text and graphics | 2,500 |  |
| Painting | 500 |  |
| Video frame | 500 |  |
| Landmark/buildings | 16,556 |  |
| Common object | 10,200 |  |

수 있다. 현재 기술의 가능성과 가치를 보고, 세계 다수의 기관이 기술 및 서비스 개발에 노력하고 있으며, 모바일 비주얼 검색을 이용하여 보이는 모든 것을 검색할 수 있는 시대가 도래할 예정이다.

약어 정리

| | |
|--------|---|
| API | Application Programming Interface |
| BoW | Bag of Words |
| BRIEF | Binary Robust Independent Elementary Features |
| BRISK | Binary Robust Invariant Scalable Key-points |
| BSD | Berkeley Software Distribution |
| CDVS | Compact Descriptors for Visual Search |
| CS-LBP | center-symmetric LBP |
| DoG | Difference of Gaussian |
| FPR | False Positive Rate |
| FV | Fisher Vector |
| ID3 | Iterative Dichotomiser 3 |
| LBP | Local Binary Pattern |
| LIOP | Local Intensity Order Pattern |
| LoG | Laplacian of Gaussian |
| mAP | mean Average Precision |
| MSER | Maximally Stable Extremal Regions |
| MVS | Mobile Visual Search |
| RANSAC | RANdomSAMple Consensus |
| REVV | Residual Enhanced Visual Vector |
| SIFT | Scale Invariant Feature Transform |
| SURF | Speeded Up Robust Features |
| TPR | True Positive Rate |
| VLAD | Vector of Locally Aggregated Descriptors |
| VT | Vocabulary Tree |

참고문헌

- [1] R. Sumit, "Facebook Photo Statistics and Insights," Online Marketing Trends. Feb. 28th, 2011.
- [2] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017, 2013. 2. 6.
- [3] J. Bughin et al., "The impact of Internet technologies:

Search," McKinsey & Company, July 2011.

- [4] B. Girodet et al., "Mobile Visual Search," *IEEE Signal Process. Mag.*, vol. 28, no. 4, 2011, pp.61-76.
- [5] Google Goggles(https://support.google.com/websearch/topic/25275?hl=en&ref_topic=1733205)
- [6] Flow (<http://a9.com/whatwedo/mobile-technology/>)
- [7] Bing Vision (<http://www.windowsphone.com/ko-kr/how-to/wp8/web/scan-codes-tags-and-text>)
- [8] Vuforia (<http://www.qualcomm.com/solutions/augmented-reality>)
- [9] MPEG, "White Paper on Compact Descriptors for Visual Search," ISO/IEC JTC1/SC29/WG11/N13951, 2013. 4.
- [10] K. Mikolajczyk et al., "A comparison of affine region-detectors," *IJCV*, vol. 65, no.1-2, 2005, pp. 43-72.
- [11] C. Harris and M. Stephens, "A combined corner and edge detector," *4th Alvey Vision Conf.*, Mahchester, UK, 1988, pp. 147-151.
- [12] Beaudet, "Rotationally invariant image operators," *International Joint Conf. Pattern Recognition*, 1978, pp. 579 - 583.
- [13] T. Lindeberg, "Feature detection with automatic scale selection," *International J. Comput. Vision*, vol. 30, no. 2, 1998, pp. 77-116.
- [14] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International J. Comput. Vision*, vol. 60, 2004, pp. 91-110.
- [15] J. Matas et al., "Robustwide-baseline stereo from maximally stable extremal regions," *Image Vision Comput.*, vol. 22, no. 10, 2004, pp. 761-767.
- [16] H. Bay et al., "Surf: Speeded up robust features," *European Conf. Comput. Vision*, vol. 1, 2006, pp. 404-417.
- [17] E. Rosten et al., "Real-Time Video Annotations for Augmented Reality," *Proc. Int'l Symp. Visual Comput.*, 2004.
- [18] E. Rosten and T. Drummond, "Machine Learning for High Speed Corner Detection," *Proc. Ninth European Conf. Comput. Vision*, vol. 1, 2006, pp. 430-443.
- [19] E. Rosten et al., "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 32, 2010, pp. 105-119.
- [20] T. Ojala et al., "Performance evaluation of texture measure with classification based on Kullback discrimination of distri-

- butions," *ICPR*, 2004, pp. 582–585.
- [21] M. Heikkilä, M. Pietikainen, and C. Schmid, "Description of Interest Regions with Center-Symmetric Local Binary Patterns," *ICVGIP, LNCS 4338*, 2006, pp. 58–69.
- [22] M. Calonder et al., "Brief: Binary robust independent elementary features," *European Conf. Comput. Vision*, 2010.
- [23] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," *Proc. IEEE ICCV*, 2011.
- [24] Z. Wang, F. Bin, and F. Wu, "Local intensity order pattern for featuredescription," *Proc. IEEE ICCV*, 2011, pp. 603–610.
- [25] VL Feat, <http://www.vlfeat.org>
- [26] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," *Proc. IEEE Conf. Comput. Vision Pattern Recognition(CVPR)*, New York, June 2006.
- [27] D. Chen et al., "Residual enhanced visual vector as a compact signature for mobile visual search," *Signal Processing*, 2012.
- [28] H. Jegou et al., "Aggregating local descriptors into a compact image representation," *Proc. IEEE Conf. Comput. Vision Pattern Recognition(CVPR)*, San Francisco, CA, June 2010.
- [29] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the Fisher kernel for large-scale image classification," *ECCV*, 2010.
- [30] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, June 1981, pp. 381–395.
- [31] S. Choi et al., "Performance evaluation of RANSAC family," *BMVC*, 2009.
- [32] S. Tsai et al., "Fast geometric re-ranking for image-based retrieval," *Proc. IEEE International Conf. Image Process.*, 2010.
- [33] S. Lepsøy et al., "Statistical modelling of outliers for fast visual search," *Proc. International Conf. Multimedia Expo(ICME)*, 2011.
- [34] HP Public Cloud, <http://www.hpcloud.com/>
- [35] Open Xerox, <http://open.xerox.com/>
- [36] Kooaba, <http://www.kooaba.com/>
- [37] MPEG, "Compact Descriptors for Visual Search: Evaluation Framework," ISO/IEC JTC1/SC29/WG11/N12202, 2011. 7.
- [38] MPEG, "Compact Descriptors for Visual Search: Requirements," ISO/IEC JTC1/SC29/WG11/N11531, 2010. 7.
- [39] MPEG, "Compact Descriptors for Visual Search: Context and Objectives," ISO/IEC JTC1/SC29/WG11/N11530, 2010. 7.
- [40] MPEG, "Compact Descriptors for Visual Search: Applications and UseScenarios," ISO/IEC JTC1/SC29/WG11/N11529, 2010. 7.
- [41] MPEG, "Call for Proposals for Compact Descriptors for Visual Search," ISO/IEC JTC1/SC29/WG11 N12201, 2011. 7.
- [42] MPEG, "CDVS: Telecom Italia's response to CE1 – interest point detection," ISO/IEC JTC1/SC29/WG11 M31369, 2013. 10.