모바일 환경에서의 영상 검색 표준 MPEG-7 CDVS

김 해 광 세종대 컴퓨터공학과 교수

1. 머리말

하루에 전 세계에서 만들어지는 영상은 몇 장이나 될까? 지구 인구가 70억쯤 된다고 하고 세계 인구의 75%가 휴대폰(60%가 스마트폰)을 보유하고 있다고 하니, 한 사람이 하루에 휴대폰 카메라로 한 장씩만 찍으면 하루에 약 52억 장의 영상이 만들어진다고 계산된다. 영상은 휴대폰뿐만 아니라 다른 방법을 통해서도 생성되고, 또한 이런 영상은 복제되고 전송되므로 정확히 계산하기 힘들지만 곱하기 10만 해도 520억 장의 영상 콘텐츠가 매일 생성되는 셈이다. 금년 2월 미국 인터넷 기사에 따르면 페이스북에 하루에 업로드 되는 사진의 수가 평균 2억 5천 장이라고 하니, 참 어마어마한 양이다.

영상 정보의 양이 천문학적인 숫자로 늘어가고 있고 이에 따라 원하는 영상 정보를 빨리 정확하게 찾는 기술, 즉 영상 검색 기술이 매우 필요한 상황이다. 현재 가장 많이 사용하고 있는 영상 검색은 영상에 키워드로 태그를 붙여서, 사용자가 인터넷 검색 엔진에 문자 질의어를 전송해서 검색하는 것이다. 하지만 키워드에 의한 방법은 문자로 표현할 수 없는 영상의 특징(예를 들면 질감이나 색의 분포 등)으로 영상을 찾을 수 없다. 코끼리를 묘사한 100쪽 짜리 책이 코끼리 사진 1장을 대신할 수는 없다는 것이다.

영상의 특징으로 영상을 검색하는 기술을 '콘텐츠 기반 영상 검색 기술'이라고 하는데, 사실 이 기술은 1980년대 말부터 연구가 시작되어 1997년에는 MPEG에서 MPEG-7 표준을 다루기 시작했고, 2001년에 MPEG-7 버전 1의 표준을 제정하였고, 그 후에도 계속 새로운 영상 특징 기술들을 포함하면서 확장됐다. MPEG-7 표준은 정지 영상뿐만 아니라 동영상 검색에 대한 기술도 포함한다. 최근에는 모바일 환경에서의 영상 검색을 위한 새로운 영상 검색 표준 기술 CDVS(Compact Descriptor for Visual Search)의 표준 작업을 진행하고 있다. CDVS 이전의 MPEG-7 영상 검색 기술과 CDVS 영상 검색 기술은 기술적인 측면에서 크게 두 가지가 다르다. CDVS 이전의 표준 기술들은 영상 전체에 대한 특징을 기술하는 전역적 영상 특징을 다루는 반면, CDVS 표준 기술은 영상의 특징점들과 그 특징점들 주변의 특징을 기술하는 지역적 영상 특징 기술이다. 또 다른 차이점은 CDVS 표준에서는 하나의 영상에서 발생할 수 있는 수많은 특징

점과 특징점들에 대한 주변 특징을 기술해야 하기 때문에 압축 기술이 특별히 표준화의 중요한 부분이다. 본 고 2장에서 CDVS 이전의 MPEG-7 영상 검색 표준 기술에 대해 전반적으로 소개하고, 특히 현재 이슈가 되고 있는 CDVS 기술에 대해 3장에서 상세히 설명한다.

2. MPEG-7 영상 검색 기술(CDVS 이전까지)

2.1 MPEG-7 버전 1 영상 특징(2001년 제정)

영상에 대한 색, 질감, 모양 등에 대한 특징들이 제정되었다. 색에 기반한 특징들로서는, 색 히스토그램(Color Histogram), 지배 색(Dominant Color), 색 공간(Color Space) 등이 표준화되었으며, 모양에 관련해서는 모양의 연결된 외곽 정보를 기술하는 외곽선 모양 기술자(Contour Shape Descriptor)와 상표와 같은 모양을 기술하는 영역 모양 기술자(Region Shape Descriptor)가 제정되었다. 질감 특징으로는, 어떤 일정한 패턴을 갖는 질감에 적합한 주파수영역 기반과 일정한 패턴을 갖지 않는 질감에 적합한 에지 기반의 질감 특징이 포함되었다. 또한, 얼굴 인식을 위한 얼굴 특징이 제정되었다.

2.2 MPEG-7 영상 특징의 진화

MPEG-7 영상 특징은 버전 1 이후 지속해서 새로운 영상 특징을 포함시키는 방향으로 확장되었다. 버전 1에 있었던 얼굴 특징과 완전히 다른 새로운 얼굴 특징 기술이 개발되었으며, 영상에 대한 따뜻함과 차가움의 정도를 표현할 수 있는 컬러온도 특징, 원 영상을 변화하여 복제한 영상을 검출하기 위한 영상 서명 특징 등이 표준으로 제정되었다.

2.3 MPEG-7 영상 특징의 산업적 성과

MPEG 국제 표준은 산업계를 수요자로 하는 표준으로서 오디오 압축과 동영상 압축 표준에서 매우 커다란 제품·콘텐츠·서비스 시장을 창출하였다. 하지만 전역적 특징을 기반으로 하는 MPEG-7 영상 표준은 아직 대규모의 시장을 형성하지 못하고 있다. 그 이유 중 하나는 아직 전역적 기반의 MPEG-7 특징을 이용한 영상 검색의 효과가 산업계의 기대에 못미치고 있기 때문일 것이다. 따라서 보다 효과적인 영상 검색을 위한 특징 연구 및 개발 표준화가 요구되고 있다.

3. CDVS 영상검색 기술

3.1 지역기반 영상 특징 기술

지역기반 영상 특징 기술은 영상에서 특징이 될 만한 점들을 검출해서 검출된 특징점들 주변의 지역적인 영상 정보에 대한 특징을 기술하는 기술이다. 입력 영상이 들어오면, 출력되는 정보는 검출된 각 특징점마다 좌표값과 지역 특징이다. 지역 영상 특징 기술은 기본적으로 전역 영상 특징 기술이 제공할 수 없는 기능을 제공한다. 두 영상 '가'와 '나'가 주어졌을 때, 지역 영상 기술은 '가' 영상의 일부분과 일치하는 '나' 영상의 일부분을 지역 특징을 비교하여 찾을 수 있다. 이러한 장점은 영상 검색의 효과를 획기적으로 개선하여 영상 검색 제품과 서비스 시장을 창출할 높은 가능성을 가진다.

이전의 MPEG-7 영상 특징에서 지역기반 영상 특징 기술이 개발되지 않았던 이유는 지역 특징 기술이 전역 특징 기술에 비해 매우 많은 계산량을 요구하며, 또 영상의 크기 변화에 따라 특징점이 변화하는 기술적 문제점 등의 원인이 있었기 때문이다.

최근 반도체의 발달로 계산 복잡성이 더 이상 심각한 문제가 아닌 상황이며, 지역 특징 기술들의 성능이 놀랍게 향상되고 있다. 현재까지 알려진 가장 성능이 우수한 기술은 SIFT(Scale Invariant Feature Transform)이다. SIFT는 영상의 크기 변화에 따라 특징점이 변화하는 문제에 매우 강인하며, 현재 CDVS 표준 과정에서 지역 특징 기술로서 채택되었다. SIFT가 영상 크기 변화를 해결하는 방법을 개념적으로 설명하자면, 입력된 영상을 다

양한 크기(scale) 변화에 의해 얻어진 여러 영상으로 이루어진 영상 공간에서 특징점을 추출하고 특징점의 주변 정보를 기술한다. 따라서 '가' 영상과 '가'를 축소한 '나' 영상이 있을 때, 영상 '가'로부터 구성한 영상 공간과 영상 '나'로부터 구성한 영상 공간에는 공통된 영상 공간이 있게 되고, 이 공통된 영상 공간에서 추출한 특징점은 동일한 특징점이 된다. 이러한 기술을 통칭해서 크기 공간(scale space) 기술이라고 한다.

3.2 CDVS 기반 영상 검색 기술의 응용

CDVS 표준은 영상으로부터 추출한 특징점들과 주변 특징 정보의 부호화에 대한 표준이다. CDVS는 기본적으로 다양한 영상 검색에 사용될 수 있으나, 특히 모바일 환경에서의 응용을 목표로 하고 있다. 아주 핵심적인 CDVS가 휴대폰에 제공하는 기능은 사용자에게 "이게 뭐지?"라는 사용자 인터페이스를 제공하는 것이다. 여기서 "이게"와 "뭐지"는 상황에 따라 다양한 의미를 가질 수 있다. 여러분이 백화점에 옷을 사러 갔다고 하자. 여러분은 관심 있는 옷을 찾은 다음, 가격 등 필요한 정보를 얻기 위해 점원에게 선택한 옷을 손으로 가리키며, "이 옷 얼마예요?"라고 물어볼 것이다. 하지만, CDVS가 장착된 카메라 휴대폰을 소지했다면 백화점의 점원에게 물어보는 대신, 휴대폰으로 그 옷의 사진을 촬영하고(이 부분이 "이게"의 역할이다) 와이파이 무선 인터넷으로 연결된 서버(가상 점원)에 사진으로부터 추출한 CDVS 특징을 전송한다(이 부분이 "뭐지"의 역할이다). 서버는 전송받은 CDVS 특징을 미리 구축한 옷들의 사진으로부터 추출한 CDVS 데이터베이스에서 검색해 어떤 옷인지를 확인하여 관련된 정보를 알려준다. 그러면 그건 바코드나 QR 코드를 사용하면 되는 것 아니냐고 반문할 수 있다. 맞다. 하지만 바코드나 QR 코드가 없다면? 예를 들어 여러분은 지금 매우 번화한 중국 상하이 푸둥의 남경동로에서 어떤 건물에 대한 정보를 알고 싶다. 건물에 바코드나 QR 코드를 달 수는 없지 않은가? 여러분은 휴대폰으로 그 건물의 사진을 찍어서 "이게 뭐지"하고 CDVS를 서버로 보내면 되는 것이다. 아주 특수한 예를 하나 들었지만, CDVS는 기본적으로 언제 어떤 상황에서든 "이게 뭐지"라는 기능을 제공한다.

3.3 CDVS 영상 검색 기술의 요구사항

CDVS 영상 검색 기술의 가장 중요한 요구사항은 2개이다. 하나는 주어진 영상에 일치하는 영상을 검색하는 검색의 정확성이고, 다른 하나는 특징점과 특징점 주변 정보의 데이터양이 매우 많으므로 정확성을 유지하면서 데이터양을 줄이는 압축 성능이다.

3.4 CDVS 영상 검색 기술의 표준화 동향

CDVS 표준은 2011년 11월 제안된 기술들에 대한 평가로 표준화가 시작됐으며, 2013년 1월 위원회 초안, 2014년 1월에 최종 국제표준 완료를 목표로 하고 있다. 현재는 작업안 버전 2의 상태다. 한국에서는 한국전자통신연구소를 비롯해 인하대, 세종대가 참여하고 있으며, 이탈리아(이탈리아 텔레콤), 미국(스탠퍼드대, 퀄컴, 노키아, 삼성전자 댈러스 연구소), 영국(미쓰비시), 중국(북경대, 화웨이) 일본(NEC) 등이 표준 기술 경쟁을 하고 있다.

현재 CDVS 표준은 핵심적인 지역 특징 기술로서 SIFT를 그대로 사용하고 있으며, 이를 응용하고 압축하는 기술을 개발해 표준화하고 있다. 그런데 표준화에 있어 큰 문제가 하나 있는데, SIFT 기술은 캐나다 브리티시 컬럼비아 대학의 David G. Low 교수가 연구하여 특허를 등록해 이 대학이 특허권을 갖고 있다. 따라서 이 표준이 제정되기 위해서는 브리티시 컬럼비아 대학이 SIFT를 표준으로 사용할 수 있도록 허가해주거나 SIFT 특허에 저촉되지 않는 새로운 지역 특징 기술을 개발해야 한다. 현재 상태는 SIFT 특허 문제가 해결되지 않고 있으며, 이 특허의 청구 범위가 매우 포괄적이어서 이 특허에 저촉되지 않은 새로운 기술을 개발하기가 쉽진 않다는 것이다.

4. 맺음말

1980년 말부터 효과적인 영상 검색에 대한 요구에 의해 많은 연구·개발이 이뤄져 왔고, 초기 MPEG-7 표준 제정에 있어서는 세계의 유수한 연구기관들로부터 회의마다 백수십 명의 연구원이 참여하였다. 하지만 전역 특징을 기반으로 하는 이전의 MPEG-7 영상 특징 표준은 시장에서 성공하지 못했고, 최근에 반도체의 발달과 함께 많은 계산량이 요구되는 지역 특징을 기반으로 한 영상 검색 기술이 향상된 정확성과 새로운 기능으로 연구되어 왔다. MPEG에서는 지역 특징 기반의 표준을 CDVS라는 이름으로 표준화를 진행 중이다.

CDVS 표준이 성공하면 막대한 시장을 창출할 것으로 예상된다. 세계적인 연구기관과의 기술 경쟁, SIFT 원천 특허 문제를 해결 등 커다란 도전이 있는 분야로서 우리나라에서도 적극적인 참여가 요구된다.

정보통신 용어해설

일체형 터치 스크린 패널 일체형 TSP(one piece Touch Screen Panel) [단말기기]

※

디스플레이 혹은 커버에 센서를 일체화한 터치 스크린 패널. 화면 손가락 위치를 인식하기 위해 2개의 센서(ITO) 층이 필요하다. ITO 센서는 유리에 증착하거나 필름에 인쇄하여 구현되는데, 필름 부착 방식이 아니라 디스플레이 혹은 커버에 센서를 일체화한 제품을 일체형 TSP라고 한다. 커버 일체형 터치는 센서층 수에 따라 G2, G1으로 구분되는데, 단일층으로 구현하면 G1, 복층으로 구현하면 G2라 부른다.