반도체디스플레이기술학회지 제13권 제2호(2014년 6월)
Journal of the Semiconductor & Display Technology, Vol. 13, No. 2. June 2014.

스마트 센서와 시각적 기술자를 결합한 사진 검색 시스템

이용환 · 김흥준*†
극동대학교 스마트모바일학과, *†경남과학기술대학교 컴퓨터공학과

Photo Retrieval System using Combination of Smart Sensor and Visual Descriptor

Yong-Hwan Lee and Heung-Jun Kim*†
Department of Smart Mobile, Far East University
*†Department of Computer Engineering, Gyeongnam National University of Science and Technology

Abstract

This paper proposes an efficient photo retrieval system that automatically indexes for searching of relevant images, using a combination of geo-coded information, direction/location of image capture device and content-based visual features. A photo image is labeled with its GPS (Global Positioning System) coordinates and direction of the camera view at the moment of capture, and the label leads to generate a geo-spatial index with three core elements of latitude, longitude and viewing direction. Then, content-based visual features are extracted and combined with the geo-spatial information, for indexing and retrieving the photo images. For user's querying process, the proposed method adopts two steps as a progressive approach, filtering the relevant subset prior to use a content-based ranking function. To evaluate the performance of the proposed scheme, we assess the simulation performance in terms of average precision and F-score, using a natural photo collection. Comparing the proposed approach to retrieve using only visual features, an improvement of 20.8% was observed. The experimental results show that the proposed method exhibited a significant enhancement of around 7.2% in retrieval effectiveness, compared to previous work. These results reveal that a combination of context and content analysis is markedly more efficient and meaningful that using only visual feature for image search.

Key Words : Photo Retrieval, Mobile Image Search, Geo-spatial Index, GPS Coordination and Image Direction, Content-based Visual Descriptor

1. 서 론

일상 생활에서 사용되는 모바일 장치가 꾸준하게 증가하고 있으며, 대부분의 장치들은 성능 좋은 카메라를 장착하여 고해상도의 영상을 손쉽게 제공하고 있다. 모바일 장치 사용자들은 저렴한 가격으로 대용량, 다량의 데이터를 축적하고 있으며, 멀티미디어 콘텐츠 중에서 가장 많은 사용자층을 가지는 영상은 시간이 지남에 따라 기하급수적으로 데이터 양이 증가하고 있다. 이에 따라, 영상 검색(Image Retrieval)은 최근 멀티미디어 관련 기술에서 부각되고 있는 연구 분야중의 하나가 되었다[1]. 영상 검색은 대용량의 디지털 자원을 활용하기 위해 인덱스를 만들고 필요한 자원으로 접근하는 방법을 연구하는 분야이며[1], 영상 검색 시스템은 입력 영상을 통해 디지털 영상 데이터베이스에서 사용자가 원하는 적절한 영상들을 열람하고 검색하는 시스템으로 정의된다[2]. 이러한 검색 시스템은 4개의 주요 기능을 수행하며, 이는 영상 레포지토리 구성, 영상 특징 추출 및 저장, 질의 영상을 통한 데이터베이스 질의와 검색 결과의 반환 및 정렬이다[3]. 영상 검색 시스템은 검색 유형에 따라 2가지 형태로 나눠진다[4]. 첫 번째 형태는 주석 또는 키워드(Keyword) 중심으로 검색을 지원하는 텍스트-기반 영상 검색(Text-based Image Retrieval)이며, 기존의 웹 기반 영상 검색 기능

†E-mail : thinkthe@gntech.ac.kr

을 제공하는 구글, 네이버 등이 여기에 포함된다. 텍스트-기반의 웹 영상 검색은 영상 키워드 부여에 있어, 주관적 해석이 반영되기 때문에 검색 만족도가 떨어지는 문제점이 있다. 또한 대용량, 다량의 영상 데이터베이스에서 키워드 생성에 따른 작업시간과 작업 부하가 높다는 단점이 있다. 이러한 단점을 보완하기 위한 방법으로, 내용-기반 영상 검색(Content-based Image Retrieval, CBIR)이 연구되고 있다. 일반적으로 CBIR에서는 색상, 질감, 모양과 같은 저급의 시각적 특징(Low-level Visual Feature)을 기반으로 자동 검색 인덱스를 구축한다[4-6]. CBIR의 핵심 알고리즘인 영상 기술자(Image Descriptor)는 색상, 질감, 모양, 공간 정보 등과 같은 영상의 내용과 관련된 특징을 추출한 다음, 영상 데이터베이스로부터 미리 추출하여 저장하고 있는 각각의 대상 영상들의 특징과 유사도를 평가하여, 사용자 질의 영상과 유사한 영상을 검색하여 반환하는 기능을 담당한다[6]. Fig. 1은 일반적인 영상 검색의 접근 방법을 보여준다[7]. Fig. 1 (a)에서는 사용자에 의해 입력된 하나 이상의 키워드를 기반으로 적절한 영상을 검색하는 텍스트-기반 검색을 나타내며, Fig. 1(b)는 질의 영상을 입력 받아, 영상 특징을 추출하고 저장된 특징 데이터베이스에서 유사한 영상을 검색하는 내용-기반 검색의 흐름도를 보여준다.

CBIR시스템은 자동 특징 추출과 유사도 평가를 통해 텍스트-기반 검색의 문제점을 해결할 수 있지만, 검색 성능 측면에서 사용자의 만족도가 높지 않다. 이는 CBIR에서 처리하는 저급의 영상 특징이 사용자의 고급의 의미론적 개념(High-level Semantic Concept)을 충족시키지 못하기 때문이다. 이러한 불일치 문제를 시맨틱 차이(Semantic Gap)이라 하며[8], 영상의 유사성 측정에 대한 주관적 사용자 해석과 검색 시스템에 적용되는 객관적 유사도 평가 모델간의 차이에서 발생하는 것이다. 이는 자동 추출된 특징들이 사용자의 의미적 개념을 분석하기에 충분하지 않기 때문에 발생한다.

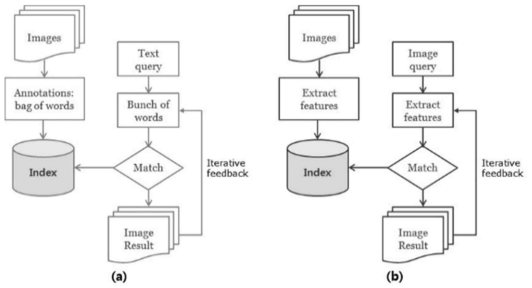

Fig. 1. Naïve system view of image search and content-based image retrieval: (a) Image search using annotations (e.g., keywords) and (b) content-based image search using an image as query (e.g., image retrieval based on visual feature) [7].

이러한 문제를 해결하기 위한 방법으로, (1) 검색과정에서 관련된 합리적 지식(Rational Knowledge)을 유도하거나 (2) 학습 메커니즘을 이용하거나 (3) 시각적 특징과 텍스트 기반 메타데이터를 결합하는 여러 연구가 진행되고 있다[9]. 지식-기반 및 학습-기반 검색의 경우, 관련된 많은 정보를 통해 광범위한 학습 훈련이 필요하기 때문에, 세 번째 방법인 메타데이터 결합 방법이 영상 검색에서 보다 간단하고 일반적으로 사용될 수 있다. 앞의 두 방법은 특정 응용분야에서 사용되기도 하며, 의료 영상검색이 대표적인 사례이다[10].

최근에는 GPS정보를 적용하는 스마트 디바이스 사용이 활성화되면서, 위치정보를 제공하는 다양한 서비스가 개발되고 있다. 대표적인 사례로, 위치 기반 서비스가 있으며[11], 이러한 응용분야 기술을 영상 검색에 적용하여 사진 촬영 위치를 활용한 사진 영상 검색의 효율성을 향상시키고자 한다. 본 논문에서는 사진 영상의 메타데이터로 위도/경도의 위치 좌표와 사진 촬영의 방향성 정보를 포함하고, 내용-기반 시각적 특징을 결합한 새로운 사진 영상 검색 기술자를 제안한다.

본 논문의 구성은 다음과 같다. 2장에서 내용 기반 영상 검색과 GPS 정보를 활용한 기존 연구를 살펴보고, 3장에서 제안하는 검색 기술자 알고리즘에 대해 설명한다. 4장에서 실험 데이터를 통해 제안 알고리즘의 성능을 분석하고, 5장에서 결론과 향후 연구 방향으로 마무리 한다.

2. 관련 연구

영상 검색과 관련하여, 다수의 범용적인 영상 검색 엔진이 연구 목적으로 또는 상용화된 제품으로 개발되고 있다. 대표적인 내용-기반 영상 검색 방법으로 IBM의 QBIC[12]을 시작으로, 콜롬비아 대학교의 VisualSeek와 WebSeek[13], 버클리대학교의 Blobworld, MIT의 PhotoBook 등이 연구 개발되었으며[4], 표준화와 관련하여 MPEG-7 XM이 있다[14]. 이러한 검색 툴들은 내용 기반으로 칼라, 질감, 모양과 공간 정보를 이용하여 영상의 특징을 추출하였다. 이외에 최근 영상 표준인 JPEG에서 JPSearch 분야를 구성하여 검색 방법과 검색 기술자, 메타데이터 포맷 등에 대한 표준화 연구가 활발히 진행되고 있다[7]. 시각적 특징 기반의 유사 영상 검색이 사용자 만족도를 충족시키지 못하여, 최근 많은 연구에서는 특징점을 기반으로 패턴인식 기

술을 적용하고 있으며, 대표적으로 SURF (Speed-Up Robust Features) 또는 SIFT (Scale Invariant Feature Transform) 알고리즘을 적용하는 사례가 있다[15]. 시맨틱 차이를 해결하는 대표적인 검색 엔진으로 ImageScape가 있으며[16], 영상 내에 객체를 검출하여 객체 기반의 부분 영상 검색을 연구하였다[28]. 모바일 환경에서 효율적인 검색을 지원하는 검색 기술자인 MPEG-7 CDVS (Compact Descriptor for Visual Search)가 연구되고 있다[17].

영상 검색에서의 GPS 정보 처리와 관련하여, 최근 연구가 진행되고 있으며, Toyama는 WWMX (World Wide Media eXchange) 데이터베이스 구축을 연구하였다[18]. 해당 논문에서는 영상의 위치정보(Location)와 시간정보를 검색 인덱스에 적용하였다. Geo-life 시스템에서는 GPS 로그를 추적, 저장하여 사용자의 위치 변화 패턴을 분석하고, 위치에 따른 이벤트를 메타데이터로 생성하여 검색에 적용하였다[19]. 해당 논문에서는 영상 캡처 장치와 영상 관리 시스템간의 지리적 위치 정보 동기화 문제를 다루었다. GeoTracker 시스템에서는 사용자가 디지털 지도상에서 영상 열람과 검색을 지원한다[20]. 해당 논문에서는 위도/경도의 위치 좌표에 따라 영상들을 표현하고, 지역에 따른 그룹핑 기능을 제공하였다. Qamra는 사진 위치정보를 기반으로 주석, 위치 문맥정보와 위치적 랜드마크(Landmark)를 사용자가 직접 입력하여 주석 기반의 영상 검색 시스템을 제안하였다[21]. Kennedy는 Flickr 데이터셋을 이용하여 사진 영상에 위치정보 태그를 추가하는 관리 시스템을 제안하였으며, 위치 태그를 통해 영상 표현, 열람, 검색 기능을 제공하였다[22]. Tollmar는 모바일 사용자를 대상으로 카메라의 위치정보와 알려진 랜드마크 데이터셋을 결합시키는 시스템을 제안하였으며, 해당 논문에서는 랜드마크를 기반으로 영상을 식별하는 방법을 제시하였다[23]. Khan은 SMS로 제공되는 텍스트와 영상을 결합하여 영상 데이터베이스를 구축하고 텍스트-기반의 주석으로 검색하는 시스템을 제안하였다[24].

이러한 여러 연구에서 위치정보와 결합된 영상 검색 알고리즘이 제안되고 있다. 이에 반해, 본 논문에서 다른 연구와의 가장 큰 차이는 사진 영상의 위치정보를 활용하는 것이다. 본 연구에서는 위도와 경도의 위치 좌표를 포함하여 촬영되는 사진 방향정보를 함께 다룬다. 일반적으로 사진 영상에서 촬영하는 카메라의 방향성에 따라, 서로 다른 사물이 사진에 나타날 확률이 높으며, 이는 사용자에게 서로 다른 영상으로 인식하는 개념적 차이를 제시할 수 있다. 이러한 접근 방법을 실험적으로 평가하여 위치 정보와 방향성 정보의 검색 적용 여부를 판별하고 검색 효율성의 향상 정도를 평가하는 것이 본 연구의 주된 목적이다.

3. 제안 검색 알고리즘

본 논문의 목적은 지리적 공간정보와 카메라의 방향정보를 시각적 기술자와 결합하여 영상 검색의 성능을 향상시킬 수 있는지를 검증하는 것이다. 이를 위해, 제안된 방법은 사진을 찍는 시점에서 GPS 센서 정보를 활용한다. 사진 영상은 촬영하는 실세계에서 특정 위치와 높은 관련성을 가지기 때문에, 사진 검색 기능을 제공하는데 있어, 시각적 특징 이외에 GPS를 기반으로 한 위치 정보가 중요한 역할을 담당하며, 본 논문에서는 GPS 위치 정보와 함께 카메라의 방향성을 고려하여 검색 효율성을 높이는 방법을 제안한다.

Fig. 2는 제안하는 사진 검색 기술자를 이용한 전체 시스템 구성도이며, 제안 시스템은 3개의 주요 기능을 수행한다.

첫 번째 기능은 촬영된 사진 영상이 미디어 서버에 전송될 때, 검색을 위한 자동 인덱싱을 수행하며, 시공간적 정보인 위치, 촬영시간과 카메라 촬영 방향 정보를 포함한다. 두 번째 기능으로, 내용-기반 검색 시스템에서 많이 활용되는 저급의 시각적 특징인 색상과 질감 정보를 추출한다. 세 번째 기능은 사용자 질의 입력 사진에 대해 적절한 영상을 검색하는 유사도 계산 기능을 수행하여 결과값을 반환한다.

3.1. 사진 영상 클러스터링을 위한 지리적 공간 인덱싱

지오-코드 영상(Geo-coded Image)은 지리적 위치가 결합된 영상을 의미하며[2], 일반적으로 영상의 위치

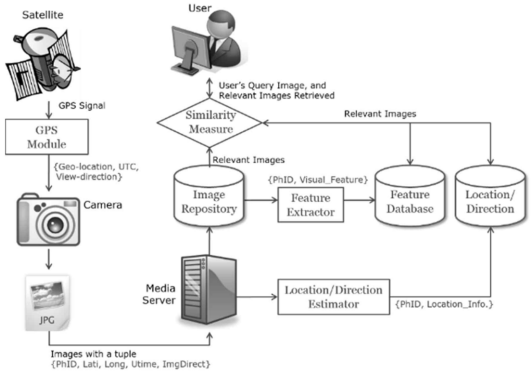

Fig. 2. Block diagram of the proposed photo retrieval system.

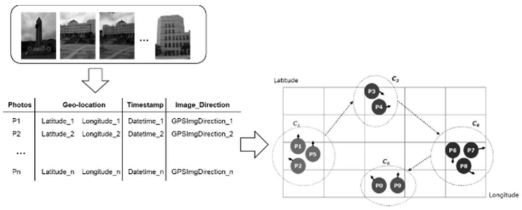

Fig. 3. Example of photo clustering with time-sequences and the corresponding geo-location track. Dotted arrows represent a user's movements, and dotted circle denotes a location cluster. Photo p_1 has an own short line arrow, which is indicates image view direction obtained from EXIF GPSImg Direction.

정보를 위도(Latitude)와 경도(Longitude)로 표시한다[2]. 이러한 위치정보를 지리적 식별자(Geo-graphical Identifier)로 활용하여 적절한 영역 집합을 적용하면 사진 영상의 위치 클러스터링(Clustering)이 가능하다. 그 Fig. 3은 사진 영상들에 대한 GPS 기반 위치와 영상 촬영 방향을 통한 그룹핑 사례를 설명한다. 집합 P = $\{p_1, p_2, \ldots p_n\}$는 T_1 에서 T_n까지의 특정 시간에서 시간의 순차적 순서와 상관없이 생성된 사진 영상 집합을 나타내며, 집합 $C_i = \{p_i, p_{i+1}, \ldots p_j\}$는 위치 클러스터에 대응하는 각 집합 C에 대해, 사진 촬영 위치가 근접한 사진들의 집합을 나타낸다. 사진 영상들은 사용자-정의 거리(Distance threshold, ε)에 따라, 하나의 위치 클러스터에 포함되며, 각 사진 영상은 영상 데이터베이스에 존재하는 사진 인덱스와 관계성을 가진 위치 상대 정보를 자동으로 생성하게 된다.

지리적 위치 정보에 대한 인덱스는 GPS 위치 좌표와 사진 촬영 방향성에 따라, 다음과 같은 6개의 속성 정보를 갖는다. 각 속성은 사진 식별자(PhotoID), 위도(Latitude), 경도(Longitude), 시간, 방향성(Image Direction), 공간적 관계(Spatial Relation)이다. 2차원 위치정보를 나타내는 위도와 경도를 통해 사진의 촬영 위치를 해석하며, 저장공간과 처리시간을 고려하여 고도(Altitude)는 표현하지 않는다. 영상의 방향성에 대응하는 방향각은 영상 메타데이터 산업표준인 EXIF의 GPSImgDirection에 저장되어 관리되며, 북쪽(North)을 기준으로 시계 방향으로 각도값을 계산한다. 영상의 방향각에 따라 중첩되지 않는 4개의 공간적 방향-관계성을 다음의 기준에 따라 산출한다.

3.2. 시각적 특징 추출

본 논문에서, 내용-기반 시각적 특징으로 추출하는 요소는 잘 알려진 2개의 영상 기술자를 적용한다. 그

공간적 방향-관계성 결정 알고리즘

```
For each image,
switch(GPSImgDirection by clockwise) {
  case(between 45.1 and 135.0) :
SpatialRelation = E; break;
  case(between 135.1 and 215.0) :
SpatialRelation = S; break;
  case(between 215.1 and 315.0) :
SpatialRelation = W; break;
  default: SpatialRelation = N; break;
}
```

중 하나는 HSV(Hue, Saturation, Value) 칼라 모델에서 64-빈(Bin)을 갖는 칼라 히스토그램(Color Histogram)이며, 다른 하나는 그레이-레벨에서 MPEG-7 에지 히스토그램 기술자(EHD, Edge Histogram Descriptor)이다. 칼라 히스토그램은 영상의 색상 분포를 표현하는 방법으로, RGB 칼라 채널을 HSV 칼라 채널로 변환한 후, 각 채널에 대해 64-비트 양자화(Quantization)을 수행하고 각 색상에 대한 발생 빈도 계산을 통해 산출한다. HSV로 변환하는 이유는 영상 처리 및 검색 분야에서 가장 좋은 성능을 보이는 칼라 모델이기 때문이며[25], 64-비트 양자화를 통해 처리량 및 계산량을 줄인다. 에지 히스토그램은 영상에서 지역 에지(Local Edge) 분포를 나타내며, 그레이-스케일(Gray-scale)로 변환된 영상을 16 (4 × 4)개의 중첩되지 않는 블록으로 구분하고, 지역 영역에서 5개 유형(0°, 45°, 90°, 135°와 분류되지 않는 비-방향성)의 블록 에지로 분류하여 발생 빈도를 계산한다.

2개의 특징 벡터는 서로 다른 차원과 특징 요소의 변화량을 가지기 때문에, 이러한 영향을 최소화하기 위해 특징 벡터의 정규화 과정이 필요하다. 색상과 질감 특징을 결합하는 시각적 특징 F_v는 수식(1)을 통해 정규화한다.

$$F_v = \left[\omega_c \times \frac{f_c}{N_c \cdot \delta_c \mu_c}, \omega_t \times \frac{f_t}{N_t \cdot \delta_c \mu_t} \right] \quad (1)$$

수식에서 N_c와 N_t는 색상과 질감 특징 벡터의 차원을 나타내며, δ_c, δ_t와 μ_c, μ_t는 각각 색상과 질감에 대한 표준편차와 평균을 의미한다. ω_c, ω_t는 색상과 질감 특징에 대한 가중치를 의미하며, 가중치는 $0 \le \omega_c$, $\omega_t \le 1$와 $\omega_c + \omega_t = 1$의 조건을 만족하는 범위에서 지정된다. 성능평가에서는 여러 번의 실험을 통해, ω_c = 0.65과 ω_t = 0.35로 지정하였다.

칼라 히스토그램은 색상을 표현하는 S채널과 V채널을 각각 연산하여 128(64 × 2)차원의 특징 벡터를, 에지 히스토그램은 80(16 × 5)차원 특징벡터를 갖는다.

3.3. 유사도 평가

본 연구에서 검색 유사도(Similarity)는 사용자 질의 영상 입력에 따라 점진적 평가 방법을 적용한다. 먼저, 지리적 위치와 영상의 방향성 정보를 통해 적절한 필터링을 수행하고, 내용-기반의 시각적 특징으로 유사도를 평가하여 사용자 질의 영상과의 순위 검색 결과를 반환한다.

위치와 방향성 정보를 적용하여 적절한 영상을 필터링하는 수식은 다음과 같다.

$$f_i \in \{\forall I: (dis(A,B) \le \varepsilon) \wedge (dis(A,B) \le \delta)\} \quad (2)$$

수식에서 I는 필터링된 영상들의 부분집합을 의미하며, ε 는 촬영 위치를 중심으로 필터링하는 사용자 지정 범위를 나타낸다. $dis(A,B)$는 질의 영상 A와 검색 대상 영상 B간의 거리를 나타내며, 두 영상의 거리는 수식(3)을 통해 계산한다.

$$dis(A, B) = C \times arccos[(sin(lat_A) \times sin(lat_B)) + (cos(lat_A) \times cos(lat_B) \times cos(long_A - long_B))]$$

수식에서 lat와 $long$은 위도/경도 좌표에 대한 십진법으로 표시되는 도(DD, Decimal Degree) 단위이며, 인덱스로 사용된 A와 B는 질의 영상 A와 검색 대상 영상B를 나타낸다. C는 라디언(Radian)에서 십진법 도 표시(DD)로 각도를 변환하기 위해 사용된 상수이다.

$dir(A,B)$는 질의 영상A와 대상 영상B간의 영상 뷰 방향 차이를 나타낸다. 이는 2개의 영상에 대한 촬영 방향의 유사성을 의미하며, 다음의 수식으로 산출한다.

$$\begin{aligned} \mathrm{dir(A, B)} = & \ |(\theta_A - \theta_B)|, \\ & \text{if } (\{(\theta_A - \theta_{mov}) \le \delta \vee (\theta_A - \theta_B) \\ & \le (\delta/2)\} \wedge \{(\theta_A - \theta_{mov}) \\ & \le 0 \vee (SR_A - SR_B)\}) \end{aligned} \quad (4)$$

수식에서 θ_A와 θ_B는 EXIF GPSImgDirection에저장된 영상 A와 B에 대한 GPS 이미지 방향각을 나타내며, θ_{mov}는 두 영상에 대한 위치 방향각의 차이를 의미한다. δ는 영상 촬영 방향에 대해 사용자가 지정하는 임계치 각도이며, SR_A는 영상A에 대하여 북쪽을 기준으로 공간-방향성에 대한 관계를 나타낸다. Fig. 4는 각 변수들에 대한 상관성을 보여준다.

GPS 위치와 사진 방향성을 이용하여 1차 필터링을 한 후, 시각적 특징에 대한 유사도 평가를 통해 검색 순위를 결정한다. 시각적 특징의 유사성은 정규화된 히스토그램 인터섹션(Normalized Histogram Intersection)을 사용하여 계산하며, 수식은 다음과 같다.

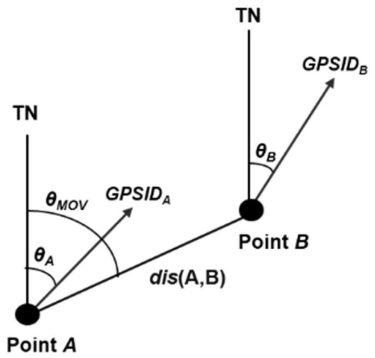

Fig. 4. Relation evaluation for field of view. GPSID is the image heading direction from EXIF GPSImg Direction, and TN indicates True North.

$$simA,B) = \frac{\Sigma_{k=0}^{n-1} min(F_k^A, F_k^B)}{\Sigma_{k=0}^{n-1} F_k^T} \quad (5)$$

수식에서 F_k^A 와 F_k^B 는 각각 질의 영상A와 대상 영상 B의 시각적 특징 벡터이며, n은 각 영상 히스토그램에 대한 빈(Bin) 수이다.

4. 실 험

여러 영상 검색 연구에서 일반적으로 Corel 또는 MPEG-7 데이터와 같은 공용의 영상 데이터베이스를 사용하지만, 본 연구에서는 자체적인 사진 영상 데이터들을 활용한다. 이러한 이유는 기존의 공용 데이터베이스는 영상의 GPS 정보를 포함하고 있지 않기 때문이다. Flickr에서 제공하는 영상 데이터들은 GPS 정보를 포함하지만, 영상의 촬영 방향 정보를 포함하지 않는다. 본 연구에서 활용된 영상 데이터는 아이폰과 10-핀을 통해 GPS 정보 수신 기능을 제공하는 니콘(Nikon) D200에 Easytag GPS 모듈을 장착하여 2,100장의 데이터 셋을 생성하였다. 모바일 환경에서 효과적인 영상 처리를 지원하기 위해, 640 × 480 JPEG 포맷을 사용하였다. 데이터셋은 20개의 카테고리로 분류되었으며, 카테고리는 평가를 위한 기초자료(Ground True Set)로 활용된다. 다른 영상 검색 연구와 마찬가지로, 실험에서 반환된 결과 영상이 질의 영상과 동일한 카테고리에 해당하는 경우, 적절한 영상으로 판단한다.

Fig. 5는 구글 맵(Google Map) 상에서 위도/경도를

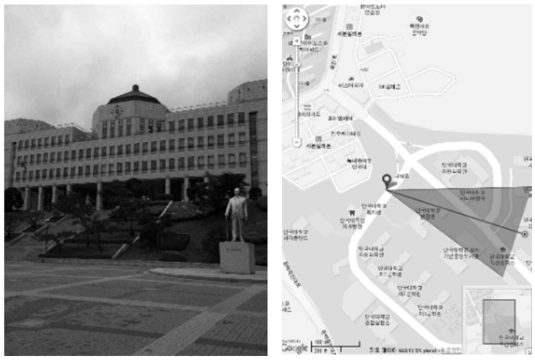

Fig. 5. Example of image and screenshot for viewing GPS coordinate and heading direction of image on Google map.

활용한 지리적 좌표(사진 촬영 위치)와 사진 영상의 방향을 표시한 예제이다.

많은 검색 연구에서 성능 평가 요소로 정확도(Precision)와 재현도(Recall)를 활용한다[27]. 정확도는 적절한 영상을 검색한 비율을, 재현도는 검색 결과에서 적절한 영상의 비율을 의미한다[2]. 집합A를 적절한 영상들의 집합으로, 집합B를 검색된 결과 영상들의 집합으로 가정하면, 정확도와 재현도는 다음과 같이 계산된다.

$$precision = P(A/B) = \frac{P(\mathrm{A} \cup \mathrm{B})}{P(B)}$$

$$= \frac{number\ of\ relevant\ images\ retrieved}{number\ of\ image\ retrieved}$$

$$= \frac{a}{a+c}$$

수식에서 a는 검색된 적절한 영상의 수를, b는 검색된 영상들 중에서 적절하지 않은 영상의 수를, c는 적절한 영상들 중에서 검색 결과로 반환되지 않은 영상의 수를 의미한다.

정확도와 재현도가 검색의 성능평가 요소로 많이 활용되기는 하지만, 두 요소간의 상관성 때문에 최근에는 *F-score*를 적용하기도 한다. 이는 수식(7)을 통해 산출된다.

Fig. 6. Example images of top ranking lists for each retrieval method. The query image is at the top-left. (a) Query with the proposed method using combination of GPS coordinates, image heading direction and visual features, (b) Query using combination of GPS coordinate and visual features, and (c) Query using visual feature alone.

$$F\text{-}score = \frac{2 \times precision \times recall}{precision + recall} \tag{7}$$

실험에서는 사진 위치와 방향 정보를 활용한 제안 알고리즘을 통한 검색 결과를 측정하고, 측정 결과를 시각적 특징만으로 검색 결과와 위치 정보 및 시각적 특징을 결합한 검색 결과와 비교 검증한다. 실험 검색에서는 높은 신뢰성을 보장하기 위해 LOOCV(Leave-one-out cross validation) 방식을 적용하여 모든 사진 영상을 대상으로 검색 실험을 수행하였다.

제안 알고리즘의 성능 평가를 위해, 프로토타입 시스템을 개발하였으며, 개발 환경은 애플 iOS 플랫폼으로, MacBook Pro 노트북, OS X 10.7(Lion) 운영체계, Xcode 4.4.1개발툴에서 Objective-C와 C++을 사용하

Table 1. Comparisons of retrieval effective using F-score over all queries, and summary of computational characteristics of the methods. The proposed method (LOC-DIR-VIS) exhibited an improvement in retrieval effectiveness. Feature extraction time, retrieving time and length for feature vector are located on the right column.

	Retrieval Effectiveness			Computing Time		
	Precision	Recall	F-score	Extraction	Retrieving	Features
VIS	35.9	50.7	40.8	0.441	0.041	208
LOC-VIS	47.5	68.2	54.4	0.452	0.117	232
LOCDIR-VIS	53.9	77.0	61.6	0.513	0.152	256

여 구현하였다.

Table 1은 질의 유형에 따른 검색 효율성을 산출한 결과이다. 해당 수치는 데이터셋에 포함된 모든 영상을 대상으로 질의 검색을 수행한 평균값이다. 위치 정보 활용을 위한 사용자 지정 거리(ε)는 실험을 통해 얻은 가장 적절한 값으로 150m로 설정하였으며, 영상 촬영 방향각(δ)은 54.4도를 사용하였다. 해당 수치는 iPhone4S에서 생성한 사진 영상의 기본 화각(FOV, Field Of View) 값이다[26].

좌측에 표시된 정확도와 재현도는 실험 검색의 상위 150개 결과 영상들에 대한 평균값이다. GPS 위치 정보와 사진 방향성 정보를 결합하여 검색한 제안 알고리즘은 LOC-DIR-VIS로 표시하였다. 시각적 특징만으로 검색한 결과(VIS) 및 시각적 특징에 위치 정보만을 결합한 검색 결과(LOC-VIS)에 비해, 제안 알고리즘이 F-score에서 보다 좋은 성능을 보였다. LOC-DIR-VIS와 VIS를 비교하면, VIS는 낮은 정확도와 높은 재현도를 보였으며, 이는 관련없는(혹은 적절하지 않은 영상)의 검색 비율이 높은 것을 의미한다. 모든 질의에 대한 평균 F-score를 살펴보면, 영상 데이터베이스의 모든 적절한 영상에 대해, LOC-VIS와 LOC-DIR-VIS는 각각 54.4%와 61.6%의 평균 F-score를 얻었다. 이에 반해 VIS는 40.8%의 결과를 얻었다. 이는 위치 정보를 결합한 두 방법(LOV-DIR-VIS, LOV-VIS)이 시각적 특징만을 적용한 방법(VIS)보다 높은 검색율을 얻을 수 있다는 것을 나타내며, GPS 위치 좌표 정보가 사진 검색에서 중요한 특징으로 적용할 수 있다는 것을 의미한다. LOC-DIR-VIS와 LOC-VIS를 비교하면, 제안 알고리즘인 LOC-DIR-VIS가 7.2% 높은 검색 효율성을 보였으며, 영상의 촬영 방향성을 다루기에 사진 영상의 검색에서 보다 좋은 성능을 보임을 알 수 있다. Table 1의 우측 열에는 전체 데이터베이스를 대상으로 영상 특징점 추출 평균시간과 검색 수행 평균시간을 보여준다. 다른 알고리즘에 비해, 제안 알고리즘이 계산시간과 저장공간상에서 약간의 오버헤드가 발생하였다.

5. 결 론

본 논문에서는 사진 촬영시점에 GPS정보와 시각적 특징을 결합한 사진 영상 검색방법을 제안하였다. 제안 알고리즘은 지리적 위치 정보와 사진 영상의 방향성 정보를 활용하여 지오-코드 인덱스를 자동으로 생성하고, 이를 통해 점진적 필터링을 적용하여 사진 검색을 수행하였다. 실험적 평가 결과, 제안 알고리즘은 기존의 검색 알고리즘보다 향상된 검색 효율성을 보였다.

본 논문의 주요 목적은 GPS 기반 정보를 활용함에 있어, 위도/경도의 위치 좌표뿐만 아니라, 사진 영상의 방향성(Image View Direction) 정보를 추가적으로 다루는 검색 방법을 제안한 것이다. 이는 사진 영상 검색에서 내용(Content)-기반의 시각적 특징과 문맥(Context)-기반의 메타데이터를 결합하여 보다 높은 검색율을 제공하는 방법을 제시하였다.

감사의 글

본 논문은 2013년도 경남과학기술대학교 기성회연구비 지원으로 수행되었음.

참고문헌

1. M. Li, Z. Chen, H. J. Zhang, “Statistical Correlation Analysis in Image Retrieval”, Pattern Recognition, vol.35, issue.12, pp.2687-2693, 2002.
2. Website: http://en.wikipedia.org.
3. N. Singhai, S. K. Shandilya, “A Survey on: Content Based Image Retrieval Systems”, International Journal of Computer Applications, vol.4, no.2, pp.22-26, 2010.
4. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, “Content-based Image Retrieval at the End of the Early Years”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.12, pp.1349-1380, 2000.
5. B. S. Manjunath, H. R. Ohm, V. V. Vasudevan, A. Yamada. “Color and Texture Descriptors”, IEEE Transactions on Circuits and System for Video Technology, vol.11, no.6, pp.703-715, 2001.
6. R. S. Torres, A. X. Falcao. “Content-based Image Retrieval: Theory and Applications”, Brazilian Symposium on Computer Graphics and Image Processing, vol.13, no.2, pp.165-189, 2006.
7. International Organization for Standards, “ISO/IEC TR 24800-1:2012 Information Technology – JPSearch – Part 1: System Framework and Components”, http:// www.iso.org/iso/home/store/catalogue_ics/ catalogue_detail_ics.htm?csnumber=60232.
8. H. H. Wang, D. Mohamad, N.A. Ismail, “Semantic Gap in CBIR: Automatic Objects Spatial Relationships Semantic Extraction and Representation”, International Journal of Image Processing, vol.4, issue.3, pp.192-204, 2010.
9. A. Lakdashti, M. S. Moin, K. Badie, “A Novel Semantic-based Image Retrieval Method”, Interna-

tional Conference on Advanced Communication Technology, pp.969-974, 2008.
10. V. V. Kumar, N. G. Rao, A.L. N. Rao, "RTL: Reduced Texture Spectrum with Lag Value Based Image Retrieval for Medical Images", International Journal of Future Generation Communication and Networking, vol.2, no.4, pp.39-48, 2009.
11. R. Jain, "Photo Retrieval: Multimedia's Chance to Solve a Real Problem for Real People", IEEE Multimedia, vol.14, no.3, pp.111-112, 2007.
12. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, "Query by Image and Video Content: The QBIC System", IEEE Computer, pp.23-32, 1995.
13. J. R. Smith, S. Chang, "Visually Searching the Web for Content", IEEE Multimedia, pp.12-20, 1997.
14. T. Sikora, "The MPEG-7 Visual Standard for Content Description – An Overview", IEEE Transactions on Circuits and Systems for Video Technology, vol.11, no.6, pp.696-702, 2001.
15. K. Velmurugan, Lt.Dr.S. Santhosh Baboo, "Content-based Image Retrieval using SURF and Colour Moments", Global Journal of Computer Science and Technology, vol.11, issue.10, 2011.
16. M. S. Lew, "Next-Generation Web Searches for Visual Content", IEEE Computer, pp.46-53, 2000.
17. Hae-Kwang Kim, "Image Retrieval Standard MPEG-7 CDVS on Mobile Environment", TTA Journal, vol.144, pp.35~38, 2012.
18. K. Toyama, R. Logan, A. Roseway, P. Anandan, "Geographic Location Tags on Digital Images", ACM International Conference on Multimedia, pp.156-166, 2003.
19. Y. Zheng, X. Xie, W. Y. Ma, "GeoLife2.0: A Location-based Social Networking Service", International Conference on Mobile Data Management: Systems, Services and Middleware, pp.357-358, 2009.
20. Y. F. Chen, G. D. Fabbrizio, D. Gibbon, R. Jana, S. Jora, B. Renger, B. Wei, "GeoTracker: Geospatial and Temporal RSS Navigation", International Conference on World Wide Web, pp.41-50, 2007.
21. A. Qamra, E. Y. Chang, "Scalable Landmark Recognition using EXTENT", Multimedia Tools and Applications, Vol. 38, Issue. 2, pp.187-208, 2008.
22. L. Kennedy, M. Naaman, S. Ahern, R. Nair, T. Rattenbury, "How Flickr Helps us Make Sense of the World: Context and Content in Community-Contributed Media Collections", International Conference on Multimedia, pp.631-640, 2007.
23. K. Tollmar, T. Yeh, T. Darrell, "IDeixis - Searching the Web with Mobile Images for Location-based Information", Mobile Human-Computer Interaction, LNCS Volume 3160, pp.288-299, 2004.
24. M. F. Khan, S. Beg, F. Kashif, "Image Transference and Retrieval over SMS", International Journal of Computer Science Issues, Vol.8, Issue.5, No.2, pp.357-361, 2011.
25. W. Bailer, H. Furntratt, P. Schallauer, G. Thallinger, W. Haas, "A C++ Library for Handling MPEG-7 Descriptions", Proceedings of ACM Multimedia Open Source Software Competition, pp.731-734, 2011.
26. P. Geiger, R. Pryss, M. Schickler, M. Reichert, "Engineering an Advanced Location-based Augmented Reality Engine for Smart Mobile Devices", Technical Report, University of Ulm, pp.1-25, 2013.
27. V. Castelli, L.D. Bergman, "Image Databases: Search and Retrieval of Digital Imagery", Wiley Inter-Science, 2002.
28. Yong-Hwan Lee, Je-Ho Park, Youngseop Kim, "Object Identification and Localization for Image Recognition", Journal of the Semiconductor and Display Technology, vol.11, no.4, pp.49~55, 2012.

접수일: 2014년 6월 2일, 심사일: 2014년 6월 12일,
게재확정일: 2014년 6월 20일