
인공위성 영상의 객체인식을 위한 영상 특징 분석

Feature-based Image Analysis for Object Recognition on Satellite Photograph

이석준, Seokjun Lee¹⁾, 정순기, Soon Ki Jung²⁾

요약 본 논문은 특징검출(feature detection)과 특징해석(feature description) 기법을 이용하여, 영상 매칭(matching)과 인식(recognition)에 필요한 다양한 파라미터의 변화에 따른 인식률의 차이를 분석하기 위한 실험 내용을 다룬다. 본 논문에서는 영상의 특징분석과 매칭프로세스를 위해, Lowe의 SIFT(Scale-Invariant Transform Feature)를 이용하며, 영상에서 나타나는 특징을 검출하고 해석하여 특징 데이터베이스로 구축한다. 특징 데이터베이스는 구글 어스를 통해 획득한 위성영상으로부터 50여개 건물에 대해 구축되는데, 이는 각 건물 영상으로부터 추출된 특징 점들의 좌표와 128차원의 벡터의 값으로 이루어진 특징 해석데이터로 저장된다. 구축된 데이터베이스는 각 건물에 대한 정보가 태그의 형식으로 함께 저장되는데, 이는 카메라로부터 획득한 입력영상과의 비교를 통해 입력영상이 가리키는 지역 내에 존재하는 건물에 대한 정보를 제공하는 역할을 한다. 실험은 영상 매칭과 인식과정에서 작용하는 내-외부적 요소들을 제시하고, 각 요소의 상태변화에 따라 인식률의 차이를 비교하는 방법으로 진행되었으며, 본 연구의 최종적인 시스템은 모바일기기의 카메라를 이용하여 카메라가 촬영하고 있는 지도상의 객체를 인식하고, 해당 객체에 대한 기본적인 정보를 제공할 수 있다.

Abstract This paper presents a system for image matching and recognition based on image feature detection and description techniques from artificial satellite photographs. We propose some kind of parameters from the varied environmental elements happen by image handling process. The essential point of this experiment is analyzes that affects match rate and recognition accuracy when to change of state of each parameter. The proposed system is basically inspired by Lowe's SIFT(Scale-Invariant Transform Feature) algorithm. The descriptors extracted from local affine invariant regions are saved into database, which are defined by k-means performed on the 128-dimensional descriptor vectors on an artificial satellite photographs from Google earth. And then, a label is attached to each cluster of the feature database and acts as guidance for an appeared building's information in the scene from camera. This experiment shows the various parameters and compares the affected results by changing parameters for the process of image matching and recognition. Finally, the implementation and the experimental results for several requests are shown.

핵심어: *image matching, image retrieval, image recognition, SIFT, artificial satellite image*

1) 주저자 : 경북대학교 컴퓨터공학과 박사과정; e-mail: sukjuni@vr.knu.ac.kr

2) 교신저자 : 경북대학교 컴퓨터공학과 교수; e-mail: skjung@knu.ac.kr

1. 서론

개인용 모바일장치 기술이 발달되고 휴대용 기기의 컴퓨팅 파워가 증가하면서, 휴대용 기기에서도 다양한 멀티미디어 데이터 처리능력이 도입되고 있다. 이로 하여금 휴대용 기기를 이용하는 수많은 사용자들은 언제 어디서나 편리하게 개인용 멀티미디어 콘텐츠를 이용할 수 있게 되었고, 서비스 제공자들은 휴대기기 사용자들이 생활 속에서 필요로 하는 서비스 개발과 시장개척에 힘쓰고 있는 추세이다. 한 예로, 휴대기기에 도입된 여러 가지 기술 중에서도 카메라를 이용하여 영상 데이터를 획득할 수 있게 한 기능은 매우 편리하고 많은 사용자들이 선호하는 기능이다. 사용자들은 휴대기기에 장착된 카메라를 이용하여 개인 포토앨범이나 바코드 인식을 통한 전자신문 서비스를 이용하고 있으며, 최근 동영상 기능을 이용한 다양한 콘텐츠가 도입됨에 따라 휴대기기 카메라를 이용한 콘텐츠 서비스 시장은 무한한 가능성을 지니고 있다. 최근 삼성전자에서는 세계적인 인터넷업체 Google의 검색엔진을 이용한 새로운 형태의 모바일 기기 개발을 진행하고 있고, 모바일기기를 통한 정보 서비스 개발에 박차를 가하고 있다. 이처럼 무선 네트워크와 인터넷 서비스의 연동을 통한 휴대용 기기의 서비스가 확대되고 있고, 해당 서비스를 제공하기 위한 하드웨어와 모바일기기도 개발되고 있는 추세이다.

하지만 모바일장치와 인터넷 서비스연동 등의 모바일 서비스 환경이 개선되고 있음에도 불구하고 사용자가 정보를 획득하기 위해서는 여전히 텍스트 기반의 검색기법을 사용할 수밖에 없다. 컴퓨팅 기술의 초기부터 컴퓨터 로컬검색과 인터넷 웹 검색에서 사용되어왔던 텍스트 기반의 검색기법은, 최근 다양한 멀티미디어 데이터의 양적증가로 인해 멀티미디어 데이터 검색 결과의 질과 합리성에 대해 여러 가지 어려움과 문제점이 제시되고 있고, 이에 대한 대안으로써 검색 질의어가 텍스트가 아닌 영상을 이용한 컨텍스트 검색기법에 대한 새로운 접근법이 제시되고 있다. 특히 모바일 장치에서의 텍스트 기반 검색기법은 휴대기기의 텍스트 입력방식이 지니고 있는 몇 가지 한계점을 지닐 수밖에 없고, 그중에서도 특히 휴대기기에 장착된 키패드나 입력장치에 익숙하지 않은 사용자층에서는 정보 서비스를 이용하는데 있어 어려움이 따를 수밖에 없다. 따라서 검색 질의어가 텍스트가 아닌 카메라로 획득한 영상을 이용한다면 이러한 한계점을 극복할 수 있을 것이다. 휴대용 기기의 카메라를 이용하여 사용자가 정보를 필요로 하는 영상을 획득하고 획득영상을 서버에서 검색하여 입력된 영상에 대한 정보를 제공해 주는 기술은, 이미 대중화 된 카메라 장착 휴대기기 사용자들에게 좀 더 편리한 정보 제공 서비스를 지원해 줄 수 있을 것이다. 이러한 기술은 다양한 응용분야에 적용될 수 있고 폭넓게 활용될 수 있을 것이다.

본 논문에서는 기존의 정보검색 메커니즘인 텍스트 기반의 검색으로 해결할 수 없었던, 영상데이터 자체를 이용한 영상 정보 검색시스템을 구축한다. 이는 영상데이터로부터 추출한 특정 특징을 미리 저장하고, 해당하는 특징과 유사도 검사를 통해 가장 높은 카운트를 가지는 영상을 결과로 도출하는 시스템으로써, 사용하는 영상 데이터는 인공위성 영상 데이터를 사용한다. 기본적으로는 영상 매칭에 대한 실험을 기반으로, 영상 매칭에 작용하는 여러 가지 파라미터의 변화에 따른 매치 카운트를 비교해 보고, 인공위성 영상을 매칭하는데 필요한 특징들을 알아보는 데 본 연구의 의의가 있다. 검색된 결과 영상은 서비스 제공자가 원하는 형태의 정보 데이터베이스로 가공되어, 해당 정보는 원하는 사용자에게 제공 될 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 영상매칭에 관한 배경이론과 관련연구를 서술하고, 3장에서는 본 논문에서 제안하는 영상매칭기반 정보검색 시스템의 처리과정과 원리를 설명한다. 4장에서는 본 논문에서 제안한 영상기반 검색시스템의 구현내용과 실험결과에 대해 기술하고, 5장에서는 본 논문의 내용을 정리하면서 향후 연구에 관하여 언급한다.

2. 배경이론과 관련연구

컴퓨터 비전 분야에서 영상 매칭(image matching)은 다양한 접근법과 과정으로 연구되어왔고, 영상 매칭을 이용하는 다양한 응용분야의 발전을 가져왔다. 영상 매칭은 객체인식(object recognition), 장면인식(scene recognition), 스테레오정합(sterео correspondence), 다중영상으로부터 3차원 구조물 복원(3D structure reconstruction from multiple images), 모션트래킹(motion tracking)등의 분야에 활용되어 왔다.

본 장에서는 영상을 해석하는데 필요한 특징(feature)에 대한 기존의 정의와 이론을 설명하고, 특징을 이용한 영상매칭 응용분야의 다양한 연구들을 소개한다.

2.1 디지털 영상의 특징요소

디지털 영상을 사용하는 컴퓨터 비전의 다양한 분야에서는 주어진 영상을 컴퓨터가 인식하고 해석하게 하기 위해서 특징이라고 일컫는, 영상내의 특징을 추출하는 과정을 반드시 거치게 된다. 따라서 특징[1]는 디지털 영상처리에서 주요한 입력요소로써 작용하며, 마커리스(marker-less)한 환경에서 영상의 오리엔테이션(orientation)이나 포지션(position)을 찾는 데 필수적인 요소이다.

Haralick과 Shapiro[2]의 저서에서는 영상 매칭을 위해 필요한 차별적인(distinctive) 특징에 대해 다음과 같은 다섯

개의 특징으로 정의하였다: 배경과 명확하게 구별되어야 하고(distinctness), 복사적(radiometric)-기하학(geometric)적인 변형에 영향을 받지 않아야 하며(invariance), 해석가능(interpretability)해야 하고, 영상의 잡음에 강하여 안정적(stability)이어야 하며, 다른 여러 포인트(point)들과 차별성(uniqueness)이 있어야 한다. 추가적으로, 영상매칭과 인식에서 특징적인 요소를 사용하기 위해서는 조명이나 3D 카메라 시점(viewpoint)의 변화에도 어느 정도는 불변적(invariant)일 필요가 있다. 영상을 촬영하는 매체가 사람이 손에 들고 사용하는 카메라(handheld camera)이고, 카메라로 촬영하는 영상은 정확한 기계적인 조율 없이는 촬영하는 방향이나 위치가 변경될 가능성이 높기 때문이다.

또 다른 연구로써, 위의 Haralick과 Shapiro가 정의한 다섯 가지의 정의를 만족하면서, 3차원 카메라 시점변화에 의한 영상의 왜곡(distortion)을 극복한 연구가 있다. Lowe[3]의 연구에 의하면 특징을 추출하는데 필요한 비용을 줄이는 방법으로써, 단계적인 필터링 과정을 거칠 것을 제안했다. 이것은 프로세싱 비용이 많이 드는 작업을 뒤로 배치하여, 이전 프로세스나 초기 프로세스를 통과한 영상의 국부적인 지역에 한정하여 높은 프로세싱 비용의 작업을 수행하는 것이다. Lowe는 주어진 영상에서 추출한 특징들의 세트를 구성하여 프로세싱 하는 주요 과정을 다음의 네 단계로 정립하였고, 이것을 SIFT(scale invariant feature transform)로 명명하였다:

1. **Scale-space extrema detection:** 첫 번째 단계에서는 영상의 모든 스케일(scale)과 위치(location)들을 찾아서 정의해 둔다. DoG(difference-of-Gaussian)함수를 이용하여 카메라의 시점 변경에 의한 스케일과 오리엔테이션(orientation)변화에 영향을 받지 않는 잠재적인 관심 점(interest point)들을 식별해낸다.
2. **Keypoint localization:** 위 논문에서는 관심 점들을 기반으로 키포인트를 생성하여 특징으로 정의하여 사용한다. 키포인트(key-point)들은 Haralick이 정의한 요소들을 만족하는 범위 내에서 안정적인 키포인트들을 선택하여 사용한다. 추출된 키포인트들은 각각의 후보위치를 갖게 되며, 첫 번째 단계에서 계산해 둔 스케일 값과 위치에 대한 정보를 이용하여 영상과 일치시키고, 정합-정확한 키포인트의 위치를 찾아 배치하는 과정-하는 과정을 거친다.
3. **Orientation assignment:** 오리엔테이션은 각각의 키포인트가 위치하는 국부적인 위치에서의 영상 내부 픽셀 값이 변화하는 방향(local image gradient directions)들을 이용하여 지정할 수 있다.
4. **Keypoint descriptor:** 국부적인 위치에서의 영상 내부 픽셀 값의 변화량은 각각의 키포인트가 위치하고 있는 영

역의 주변영역을 이용하여 계산할 수 있다. 앞서 세 번째 단계에서 수행한 것처럼 영상의 그라디언트 방향의 벡터(vector)들을 가중치가 부여된 가우시안 윈도우(gaussian window)내의 모든 벡터들을 2×2 또는 4×4 , 8×8 등의 샘플 배열(sample array)로 나타낼 수 있으며, 각 영역에 속한 벡터들의 가중치 값들을 통하여 키포인트 특징(feature)의 경향성을 나타낼 수 있다.

이러한 접근법은 영상 내에서 추출한 포인트(corner-like points)나, 영역들의 그룹을 정립된 몇 개의 특징으로 표현함으로써, 영상의 스케일이나 오리엔테이션의 변화가 있을 경우에도 해당하는 특징을 찾을 수 있다는 장점이 있다.

2.2 디지털 영상 매칭을 이용한 다양한 연구들

컴퓨터 비전분야에서 영상을 이용하기 위한 노력은 너무나 다양하고 폭넓게 진행되고 있다. 특히 영상으로부터 특징 점을 찾고, 찾아진 특징 점을 이용하여 특정한 의미 있는 일을 할 수 있도록 하는 응용연구와 소프트웨어 개발도 활발히 진행되고 있다.

앞서 SIFT를 제안한 Lowe와 함께한 Matthew[4]의 최근 연구를 참조하면, 실영상(real scene)으로 촬영된 이미지 시퀀스로부터 추출된 매칭 점을 기반으로, 오리엔테이션과 스케일, 그리고 전 후 이미지의 매칭관계와 이미지간의 위치적 배치관계까지 자동적으로 정렬(image mosaic)해 주는 파노라믹 영상(panoramic image) 생성 기법을 보여주었다. Autostitch[5]라는 프로젝트로 진행되는 파노라믹 알고리즘은, Autopano Pro, Serif PanoramaPlus, Calico라는 이름의 상업용 제품으로 이용되고 있다.

Nister[6]의 연구에서는 40000여장의 음악CD커버로 사용되는 대량의 이미지 파일들로부터 MSERs(Maximally Stable Extremal Regions)[7] 영역 추출기(region detector)를 이용하여 특징을 추출하고, 그들이 제안한 vocabulary tree로 데이터베이스화 한다. 시스템이 구성되면, 사용자의 카메라로부터 입력된 영상 프레임에서 같은 방법으로 특징을 추출하여 구축된 데이터베이스 검색으로 영상 매칭과정을 수행한다. 영상 데이터베이스로부터 구축된 특징 데이터를 매칭하는 방법은 구축과정에서 색인(indexing)된 로컬 특징 인덱스들을 TF-IDF(Term Frequency Inverse Document Frequency) 스코어링을 이용하여 매치 카운트의 엔트로피가 가장 높은 노드를 쿼리 영상과 같다고 판단하는 방법이다.

Sivic과 Zisserman[8]의 연구를 보면 영화와 같은 동영상으로부터 발췌한 대량의 이미지 프레임으로부터 특징을 추출하고, 특징의 위치와 오리엔테이션, 스케일 등에 영향을 받지 않고 검색, 데이터베이스화 하여 특정 오브젝트를 찾으려 하는 연구 결과가 있다. 이는 'Video Google'이라는 프로젝트

트 이름으로, 인터넷 포털업체 Google의 이미지 검색기능에도 적용가능성을 놓고 테스트 중에 있다.

영상매칭기법은 최근에 각광받고 있는 AR(augmented reality)시스템을 위한 연구에도 적용되고 있다. AR분야에서는 마커 없이(marker-less) 객체(object)나 장면(scene)에 대한 자세정보를 파악하기 위해 이미지기반(image based) 접근법이 사용되고 있다. 'bazAR'[9]은 카메라로 입력된 영상에 대한 특징을 추출하고, 가상의 트레이닝이나 기하학적 변환관계(affine transformation)를 계산하여 장면이나 객체의 자세를 추정하여 AR시스템을 구축한다. 또한 앞서 나열한 연구들 외에도, 영상 데이터를 이용한 이미지 왜곡(warping)[10]이나, 3차원 트래킹을 위한 상용소프트웨어[11,12,13,14,15,16,17]에도 많이 사용되고 있다.

3. 본론

이번 장에서는 본 논문에서 제안하는 인공위성 영상의 객체 인식을 위한 시스템을 설명하고, 본 시스템에서 사용되는 영상매칭방법과 특징색인, 검색에 사용되는 요소들과 그 처리 과정을 설명한다. 광역의 지역을 촬영한 인공위성영상에서 특정 객체, 즉 건물과 같은 객체를 인식하는데 있어서 어떤 요소가 특징이 되기에 적당한 지를 관찰하는 과정이 본 연구에서 가장 주안점이 된다.

3.1 시스템 개요

본 논문에서 제안하는 위성영상 내부의 특정 시설물을 인식하는 시스템의 기본적인 처리흐름은 그림 1과 같다.

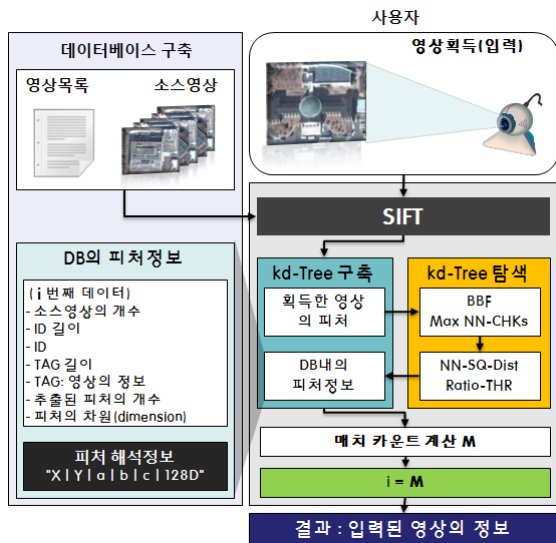


그림 1. 시스템 처리 흐름도.

본 시스템은 크게 두 가지 처리 영역으로 나뉜다. 영상이 주어지면 영상을 기반으로 특징을 추출 하고, 다차원 데이터에

적합한 데이터구조를 이용하여 데이터를 구조화 한다. 본 연구에서는 가장 손쉽게 구현할 수 있고, 그 효과 또한 검증이 되어 많이 사용하고 있는 kd-tree[18]를 이용한다. 데이터 구조화 과정이 마치면 새로이 입력되는 특징데이터를 갱신하고, 해당 특징 데이터와 기존 데이터베이스 내의 특징 정보와 일치하는지 비교하는 특징매칭 과정이 수행된다. 이 과정에서 일치되는 특징의 수가 가장 많게 나온 데이터베이스의 영상이 새로이 입력되는 영상과 같은 영상이라고 판단하게 된다.

3.2 특징추출



그림 2. 특징추출의 예 - 흰색 점으로 표시 (1280×999×24bit, 총 24,245 특징점, 경북대학교, 위성영상 출처: 구글 어스).

본 연구에서는 SIFT를 사용하여 특징을 추출하였다. SIFT는 크게 추출기(detector)와 해석기(descriptor)로 나뉘는데, 추출기는 MSERs와 같은 원리로 원형의 영역내부의 그라디언트 값을 이용하여 특징 점(corner point)을 찾고, 특징 점을 둘러싼 원형의 스케일과 회전정도를 파악하여 정규화 시킨다[19,20]. 찾아진 특징 점들은 DoG극값과의 조화정도를 이용하여 스케일 공간 극값(scale-space extrema)을 찾는다. 찾아진 값은 특징 점(feature point)으로 정의되며, 위치를 측정하고 필터링을 하여 사용하기에 좋지 않은 특징 점들을 개선하는 과정을 거친다. 그림 2는 구글 어스(Google earth, <http://earth.google.com>)로부터 획득한 경북대학교 위성사진의 특징추출 예이다.

3.3 데이터베이스 구축

특징 데이터베이스는 특징에 대한 기하학적(geometric)정보와 해당 특징들의 그룹을 이루는 객체에 대한 정의를 담고 있다. 본 논문에서 실험한 영상은 경북대학교 캠퍼스의 전체 위성사진을 기반으로 캠퍼스 내에 존재하는 50여개의 시설물에 대해 객체를 정의하고 정보를 할당하였다. 따라서 50여개 시설물에 대한 위성영상 목록을 작성하고 각 시설물에 ID와 TAG를 할당하였다. ID는 객체를 정의하기 위해서 할

당하고, TAG는 해당 객체에 대한 정보를 삽입하기 위해서인데, 여기서는 시설물의 이름으로 정의하였다.



그림 3. 데이터베이스 구축과정.

그림 3에서는 데이터베이스 구축과정을 알 수 있고, 그림 4는 실제로 구축된 데이터베이스를 텍스트 파일로 출력해 본 것이다. 출력된 데이터인 그림 4에서 보면, 전체 데이터에서 1번 데이터의 일부가 나타나 있는데, 총 50개 시설물에 대해, ID가 1번인 '본관'건물은 758개의 특징이 검출되었으며, 758의 특징데이터가 ID와 TAG정보 밑으로 연속되어 기록된다. 그림 4에서 보듯이, 하나의 시설물마다 해당 시설물이 포함하는 특징정보들을 저장하고 있고(X, Y, a, b, c, description), 각각이 ID와 TAG가 할당되어있다.

1	50	소스영상의 개수
2	1	id 길이 / id
3	1	Tag 길이 / Tag
4	4	추출된 피쳐개수
5	758	피쳐정보의 차원
6	128	128 차원 해석 데이터
7	219	336844 152 254127 0.000000 0.000000 0.000000 21.000000 17.000000 38.000000
8	219	336844 152 254127 0.000000 0.000000 0.000000 13.000000 0.000000 0.000000
9	21	44 152 254 127 0.000000 0.000000 0.000000 4.000000 17.000000 54.000000
10	175	109 296 235 0.000000 0.000000 0.000000 109.000000 46.000000 1.000000

그림 4. 특징 데이터베이스 샘플.

특징 해석데이터는 기본적으로 특징의 2차원 영상내의 좌표 X, Y와 영역에 대한 변환관계 a, b, c(affine region parameter), 그리고 각 특징마다 고유한 해석 데이터가 있다. SIFT를 이용한 본 실험에서는 한 지역에 대한 영상의 그라디언트의 백터가 각각이 8방향의 오리엔테이션으로 이루어진 4×4배열로 구성하였으므로, 각각의 특징마다 4×4×8=128개의 키폰트 벡터들로 이루어진다.

3.4 특징 색인과 검색

특징색인과 검색과정은 기본적으로 kd-tree를 이용하여 영상 특징의 근사 k개의 최 근사 이웃들(k-nearest neighbors)을 찾고, 각 이웃 특징들 간의 거리가 증가하는 순서대로 정렬한다. 본 논문에서는 2차원 평면 영상의 특징 점들을 kd-tree로 구성하였으므로, 유클리디안-공간(euclidean-space)상에서 최 근사거리 검색(approximate Nearest Neighbor search)[21,22]을 통하여 같은 특징들을 검색할 수 있고, 각 특징들의 거리관계를 파악할 수 있다. 위의 작업을 전체 지도에 대해 미리 수행하여 데이터베이스로 구축하고, 검색하고자 하는 영상의 입력이 들어왔을 때,

같은 방법으로 최 근사거리 검색을 수행한 특징 점들의 수를 스코어링 하여 카운트가 가장 높게 나타난 유닛의 ID를 가진 영상이 입력된 영상과 유사한 영상이라고 판단한다.

4. 구현 및 실험결과

본 논문은 SIFT[23]를 기반으로 하여, 인공위성으로부터 획득한 위성영상에서 영상의 전체, 또는 부분을 인식하는 시스템을 개발하는데 필요한 다양한 변수들의 변화에 의한 인식률의 차이를 관찰하기 위해 진행되었다. 영상을 인식하고 해석하는 데는 여러 가지 변수가 적용되는데, 본 논문에서는 영상해석과 검색에 사용되는 변수들의 변화를 설정하고 그 변화에 따라 영상해석, 또는 영상 검색의 정확도에 얼마나 영향을 미치는지 살펴보았다.

일반적으로 시스템을 구현할 때, 시스템의 응답률을 높이기 위해 트레이닝과정을 거치거나 빌드하기 위한 데이터의 양을 풍성하게 하는 방법을 사용한다. 하지만 본 연구에서는 응답률을 높여 시스템의 응답률을 보장하는 것보다, 단일 영상의 데이터를 통해 컴퓨터 비전 기술로 얼마나 응답률을 높일 수 있는가를 관찰하기 위한 것이 본 실험의 목적이고, 또 특정 파라미터의 조정이 동일 환경내의 영상 응답률에 얼마나 영향을 미치는 가를 관찰하는 것이 연구의 의도 이므로 여타의 트레이닝 과정은 거치지 않도록 실험환경을 설정하였다.

4.1 구현시스템

본 논문에서 제안한 시스템은 그림 5와 같이 구성되었으며, 영상을 획득하는 장면과 처리된 결과를 출력하는 장면을 확인할 수 있다. 카메라가 장착된 UMPC(ultra mobile PC)에 본 시스템을 탑재하여 카메라로 입력된 영상 데이터로부터 특징을 추출하고 데이터를 구조화 한 후, 기존에 구축된 데이터베이스를 참조하여 매치되는 특징을 계산하게 된다. 데이터베이스 구축하는 과정은 대량의 영상데이터로부터 특징을 추출하고 구조화 하는 작업을 해야 하므로, 서버나 워크스테이션에서 미리 구축해 두는 것이 효율적이다.



그림 5. UMPC를 활용한 모바일기기 영상기반 검색.

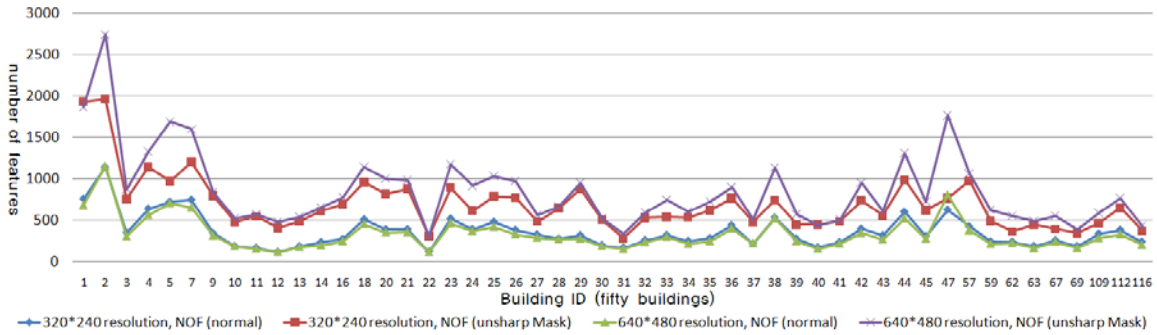


그림 6. 원천영상의 해상도와 필터여부에 따른 추출된 특징 개수(NOF)의 변화

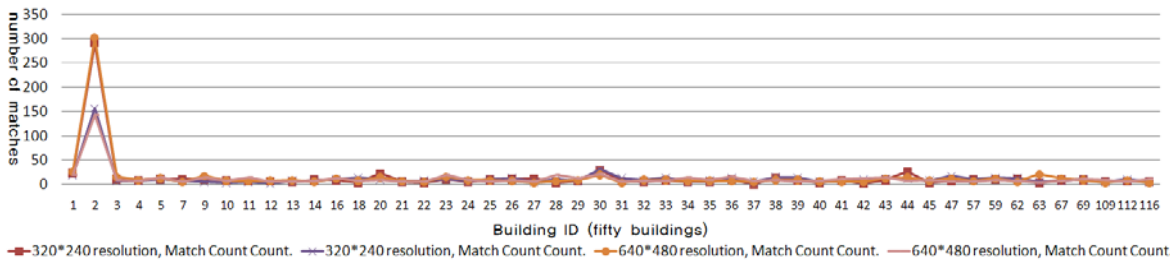


그림 7. 원천영상의 해상도에 따른 응답률(match count) 비교(입력영상이 2번 건물일 경우).

4.2 실험방법

실험은 표 1에서 나타낸 바와 같이 입력되는 영상에 변화를 가할 수 있는 다양한 조건들의 변화에 따른 응답률(matching count)을 산정하는 방식으로 진행하였다.

표 1. 실험 파라미터 내용.

조건변화	조건상세	
입력영상	카메라의 위상변화	방위(orientation)각 변화(45 degree) 회전(rotation)각 변화(45 degree)
	촬영대상	인쇄 매체(대형 인쇄) 디지털 디스플레이(LCD 모니터)
원천영상	해상도	320×240 640×480
입력기기	영상 필터	un-sharp Mask filter
	USB 웹캠 / 모델: VIJETALK CCD	
	핸드폰 카메라 / 모델: Samsung SPH3200	
검색 파라미터	HD 캠코더 / 모델: SONY HD CAMCORDER	
	KDTREE_BBF_MAX_NN_CHKS (100/200/300) NN_SQ_DIST_RATIO_THR (0.19/0.49/0.89)	

4.3 영상 조건변화에 따른 비교

영상의 조건변화에 따른 실험은 초기에 영상에 대한 특징 데이터베이스를 구축할 때 사용되는 원천영상의 변화에 따른 비교와 사용자가 원하는 정보를 입력할 때 사용하게 되는 입력영상의 변화에 따른 비교로 나눌 수 있다. 원천영상은 데이터베이스를 구축하기 위해 사용하는 영상데이터인데,

초기 특징데이터의 정보량에 따른 응답률을 알아보기 위함이다. 입력영상에 따른 비교과정은, 본 논문에서 제안하는 시스템이 사람이 들고 사용하는 모바일 기기를 대상으로 하고 있으므로, 영상데이터가 입력되는 상황에서 발생 가능한 다양한 변수들 중에서 예상 가능한 변수들을 정리하여 비교해 보는 과정이다.

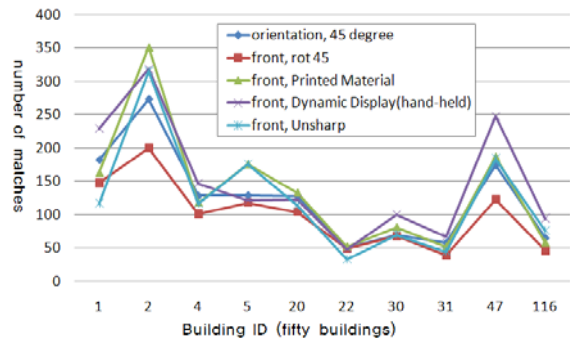


그림 8. 입력영상의 조건변화에 따른 응답률 비교 (10개 건물).

4.4 영상입력기기에 따른 비교

영상에서 픽셀간의 그라디언트의 대비를 더욱 크게 하면 특징을 찾고 그 특징적인 요소들을 보다 확연하게 구분 할 수 있을 것이라는 가정 하에, 언-샤프 마스크 필터(un-sharp mask filter)를 사용하였다. 언-샤프 마스크 필터는 영상의 에지(edges)를 검출하는 필터가 아니라, 비슷한 명암을 가진 픽셀을 둘러싼 다른 명암 값의 차이를 조절 하는 필터이다. 이는 영상에서 저주파 성분을 걸러내고, 보다 선명한 결과를 얻을 수 있도록 해준다. 표 2에서는 각 입력기기에 따른 마

스크 필터 적용 결과를 보여주는데, 사람의 눈으로 보기에는 보다 선명하고 구분하기 용이한 결과임을 알 수 있다. 그림 6의 실험결과를 보면 언-샤프 마스크 필터를 적용하였을 경우에 추출된 특징의 양이 많아 진 것을 확인할 수 있고, 이는 정보의 양이 많아 졌다고 해석할 수 있다. 하지만, 실제로 영상매칭 과정을 거쳐서 매칭 응답률을 비교해 보면 오히려 매칭과정에서는 응답률을 낮추는 역반응이 나타난다는 것을 확인 할 수 있다(그림 9).

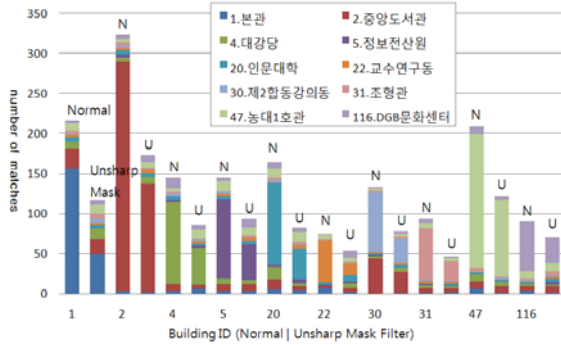


그림 9. 마스크필터 적용여부에 따른 매칭 응답률 비교.

입력기기에 따라서 해상도, 샘플링 디바이스의 컬러 스케일링 범위, 렌즈의 특성 등에서 획득된 영상에 차이가 발생할 수 있다. 그림 10을 보면 영상을 획득하는 장치에 따라 매칭 응답률에 차이가 발생할 수는 있으나, 한 장치를 대상으로 보았을 때, 건물을 인식하는 데는 크게 무리가 따르지 않음을 알 수 있다.

표 2. 영상획득 기기에 따른 원본영상과 필터 적용결과 영상.

구분	원 본	Unsharp Mask
구글 어스로부터 저장된 영상 (Web)		
SONY HD Camcorder 로 획득한 영상 (720×480)		
VIJETALK CCD WebCam으 로 획득한 영상 (640×480)		

휴대전화의 카메라가 매칭 응답률의 차이가 큰 것은, 카메라 지원 해상도가 매우 작기 때문에 발생한 문제이므로 현재 생산되는 휴대전화 카메라의 해상도가 꾸준히 질이 좋아지

고 있음을 감안하면 크게 문제되지 않을 부분이다.

4.5 검색 파라미터에 따른 비교

본 논문에서 제안하는 시스템은 영상으로부터 추출한 특징 데이터를 kd-tree를 사용하여 구조화하고, 중심이 되는 특징과 이웃특징간의 거리를 계산한 값을 이용하여 비교하는 과정을 거친다. 여기서 특징에 대한 최 근사 이웃들에 대한 해석데이터를 근삿값이 큰 순으로 정렬하게 되는데, kd-tree로 정렬된 해석데이터를 탐색해 가는 깊이를 얼마나 할 것인지를 결정하는 파라미터를 설정하였다(표 1, *KDTREE_BBF_MAX>NN_CHK*S). 이 값이 커지면 커질수록 최 근사 이웃 특징들의 후보들을 체크하는 범위가 넓어지므로 더 많은 후보들을 검색해 볼 수 있다. 또한 최 근사 이웃들에 대한 거리 값을 이용하여 이웃하는 특징이나 아니냐를 판별하게 되는데, 판별 시 사용하는 문턱 값(threshold)의 크기에 따른 검색범위에 대한 파라미터도 설정하였다(표 1, *NN_SQ_DIST_RATIO_THR*). 이 값이 커질수록 해석 값의 범위를 넓게 잡을 수 있다. 그림 12는 1번 건물에 해당하는 지역을 카메라로 촬영하였을 경우, 검색 파라미터 변화에 따른 결과를 나타내었다. 결과적으로 1번 건물에 대한 특징 데이터와 매칭 응답률이 가장 높게 나타났으며, 이는 곧 1번 건물을 인식하였다는 것을 의미한다. 하지만, *NN_SQ_DIST_RATIO_THR* 값이 커지면 매칭 카운트의 수가 늘어나는 효과 이외에는 별다른 차이를 발견 할 수가 없었으며, 건물 인식하고 식별하는 데는 크게 영향을 미치지 않았다.

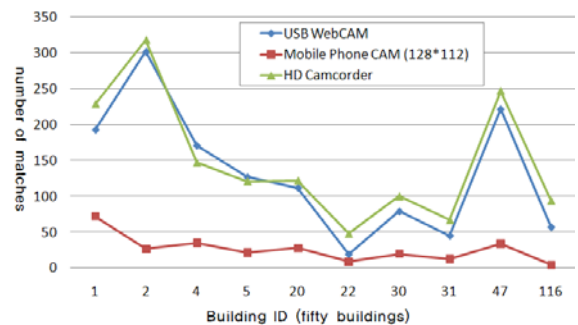


그림 10. 입력기기에 따른 응답률 비교(10개 건물).

5. 결론 및 의의

본 논문에서는 인공위성 영상을 기반으로 영상의 변화조건을 설정하여 영상매칭에 얼마나 영향을 미치는가를 살펴 보았다. 영상 변화조건에 대한 요소로써, 입력되는 영상의 조건변화, 데이터베이스를 구축할 때 사용하는 원천영상의 변화, 영상 입력기기에 따른 변화, 트리를 검색할 때 얼마만큼의 검색범위를 지정할 것인지에 대한 실험과 검증은 해보았

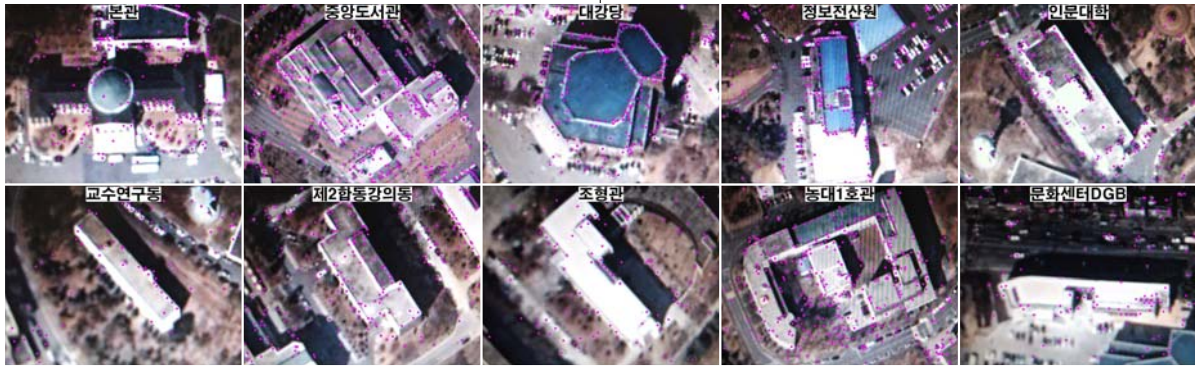


그림 11. 10개 건물에 대한 매칭 결과영상

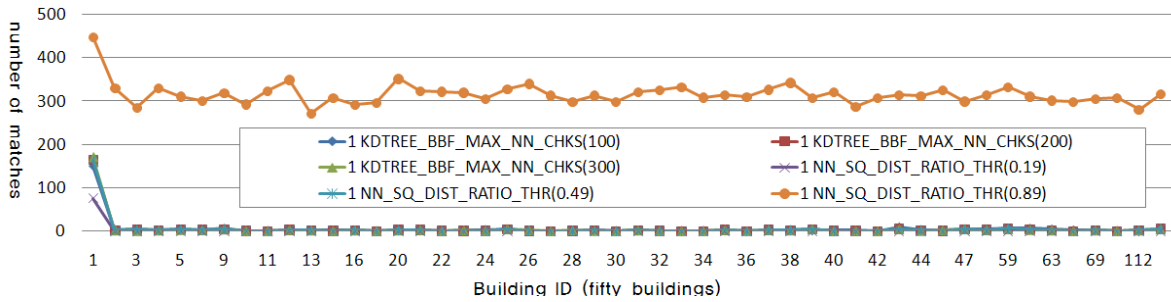


그림 12. 검색과라미터 변화에 따른 매칭 응답률(1번 건물을 입력영상으로 하였을 경우).

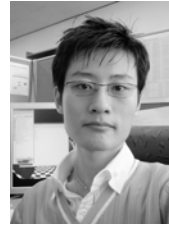
다. 또한 영상매칭 결과를 활용하여 모바일 기기를 활용한 영상기반 검색시스템을 구축하여 위성영상을 통한 시설물의 정보를 제공할 수 있도록 하였다.

본 논문에서 제안한 영상기반 검색시스템은 기존의 SIFT를 활용하여, 불변하는(invariant) 영상특징들을 추출하고 매칭하는 과정을 기반으로 구동되기 때문에 프로세스의 양이나 시간적인 차원에서 보다 효율적인 방향으로 개선될 필요가 있다. 먼저 위성영상에 적합한 특징들에 대해, SIFT에서 사용하는 MSERs 특징 디텍터만 사용할 것이 아니라, 속도는 느리지만 스케일에 따른 응답률이 비교적 높게 나오는 Harris-Laplacian 디텍터[20]를 속도문제를 개선하여 사용한다거나, 속도는 빠르지만 응답률은 다소 떨어지는 Lepetit[9]의 특징검출법을 사용해 보는 것도 먼저 시도해 볼 것이다. 이는 모바일 기기에서 사용하는 시스템에서는 정확도와 처리시간이 trade-off 관계에 있으므로, 그 사이의 효율적인 최적조건을 찾는 것이 앞으로 풀어야 할 문제일 것이다. 또한 영상매칭을 하는데 사용하는 불변특징들의 기하학적 변환관계(affine transformation)를 잘 활용하여, 본 정보검색 시스템의 가시화 단계에서는 AR기법을 적용하여, 보다 높은 차원정보들을 가시화 할 수 있도록 연구되어야 할 것이다.

참고문헌

- [1] Fabio Remondino, Detectors and Descriptors for Photogrammetric Applications, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Commission III Symposium, Vol. XXXVI, part 3, pp. 49-54, Bonn, Germany, 2006.
- [2] Robert M. Haralick and Linda G. Shapiro, Computer and Robot Vision (Vol. II), Prentice Hall, 630 pages, June 2002.
- [3] David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, Vol. 60, Issue 2, pp. 91-110, 2004.
- [4] Matthew Brown and David G. Lowe, Automatic Panoramic Image Stitching using Invariant Features, International Journal of Computer Vision, Vol. 74, Issue 1, pp. 59-73, 2007.
- [5] Autostitch, <http://cs.ubc.ca/~mbrown/panorama/panorama.html>
- [6] David Nister and Henrik Stewenius, Scalable Recognition with a Vocabulary Tree, Proc. of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), Vol. 2, 2006.
- [7] M. Donoser and H. Bischof, Efficient Maximally Stable Extremal Region(MSER) Tracking. Proc. of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.

- [8] Josef Sivic and Andrew Zisserman, Video Google: A Text Retrieval Approach to Object Matching in Videos, Proc. of the 9th IEEE International Conference on Computer Vision (ICCV 2003), Vol.2, 2003.
- [9] V. Lepetit and P. Fua, Keypoint Recognition using Randomized Trees, Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, Nr. 9, pp. 1465–1479, 2006.
- [10] Hansheng Lei and Venu Govindaraju, Direct Image Matching by Dynamic Warping, Proc. Conference on Computer Vision and Pattern Recognition Workshop, Vol. 5, 2004.
- [11] 3D-Equalizer, <http://3dequalizer.com>.
- [12] Voodoo, <http://www.digilab.uni-hannover.de>.
- [13] SynthEyes Camera tracker, <http://ssontech.com>.
- [14] PFTrack, <http://thepixelfarm.co.uk>.
- [15] PFHoe, <http://www.pfhoe.com/hoe.htm>.
- [16] 2d3 boujou, <http://www.2d3.com>.
- [17] Realviz Matchmover, <http://sfx.realviz.com>.
- [18] Volker Gaede and Oliver Gunther, Multidimensional Access Methods, ACM Computing Surveys, Vol. 30, Issue 2, pp. 170–231, 1997.
- [19] K. Mikolajczyk and C. Schmid, Scale & Affine Invariant Interest Point Detectors, International Journal of Computer Vision, Vol. 60, Issue 1, pp. 63–86, 2004.
- [20] K. Mikolajczyk and C. Schmid, Indexing based on Scale Invariant Interest Points, Computer Vision, 2001, ICCV 2001, Proc. of the 8th IEEE International Conference on, Vol. 1, pp. 525–531, Vancouver, BC, Canada, 2001.
- [21] P. Indyk and R. Motwani, Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality, Proc. of the 30th Annual ACM Symposium on Theory of Computing (STOC), pp. 604–613, 1998.
- [22] J. Kleinberg, Two Algorithms for Nearest-Neighbor Search in High Dimensions, Proc. of the 29th Annual ACM Symposium on Theory of Computing (STOC), pp. 599–608, 1997.
- [23] Andrea Vedaldi, Source code and Manual: An implementation of SIFT detector and descriptor, <http://vision.ucla.edu/vedaldi/code/sift/sift.html>, VisionLab, UCLA.



이 석 준

1999년 3월 ~ 2004년 2월 대구대학교 정보통신공학부 졸업(공학사). 2004년 3월 ~ 2006년 2월 경북대학교 컴퓨터공학과 졸업(공학석사). 2006년 ~ 현재 경북대학교 컴퓨터공학과 박사과정. 관심분야는 HCI, Image Recognition, Mixed Reality 등임.



정 순 기

1986년 3월 ~ 1990년 2월 경북대학교 컴퓨터공학과 졸업(공학사). 1990년 3월 ~ 1992년 2월 한국과학기술원 전산학과 졸업(이학석사). 1992년 3월 ~ 1997년 2월 한국과학기술원 전산학과 (공학박사). 1997년 ~ 1998년 University of Maryland (Research Associate). 2001년 ~ 2002년 IRIS, University of Southern California(Research Associate). 1998년 ~ 현재 경북대학교 컴퓨터공학과 부교수. 1999년 ~ 현재 (주)아이디스 기술고문. Virtual Reality, Artificial Intelligence, Computer Vision, Image Processing, Computer Graphics 등임.