
불변하는 스케일-아핀 특징 점을 이용한 평면객체의 위치 추정

Planar-Object Position Estimation by using Scale & Affine Invariant Features

이석준, Seokjun Lee, 정순기, Soon Ki Jung
경북대학교 컴퓨터공학과, 가상현실 연구실

요약 카메라로 입력되는 영상에서 객체를 인식하기 위한 노력은, 기존의 컴퓨터 비전분야에서 좋은 이슈로 연구되고 있다. 영상 내부에 등장하는 객체를 인식하고 해당 객체를 포함하고 있는 전체 이미지에서 현재 영상의 위치를 인식하기 위해서는, 영상 내에 등장할 객체에 대한 트레이닝이 필요하다. 본 논문에서는 영상에 등장할 객체에 대해서, 특징 점을 검출(feature detection)하고, 각 점들이 가지는 픽셀 그라디언트 방향의 벡터 값들을 그 이웃하는 벡터 값들과 함께 DoG(difference-of-Gaussian)함수를 이용하여 정형화 한다. 이는 추후에 입력되는 영상에서 검출되는 특징 점들과 그 이웃들 간의 거리나 스케일의 비율 등의 파라미터를 이용하여 비교함으로써, 현재 특징 점들의 위치를 추정하는 정보로 사용된다. 본 논문에서는 광역의 시설 단지를 촬영한 인공위성 영상을 활용하여 시설물 내부에 존재는 건물들에 대한 초기 특징 점들을 검출하고 데이터베이스로 저장한다. 트레이닝이 마친 후에는, 프린트된 인공위성 영상내부의 특정 건물을 카메라를 이용하여 촬영하고, 이 때 입력된 영상의 특징 점을 해석하여 기존에 구축된 데이터베이스 내의 특징 점과 비교하는 과정을 거친다. 매칭 되는 특징 점들은 DoG로 정형화된 벡터 값들을 이용하여 해당 건물에 대한 위치를 추정하고, 3차원으로 기 모델링 된 건물을 증강현실 기법을 이용하여 영상에 정합한 후 가시화 한다.

핵심어: Image recognition, Position estimation, Satellite image

1. 서론

카메라로부터 입력되는 영상은 다양한 정보를 내포하고 있다. 이러한 디지털 영상을 다루는 컴퓨터 비전 연구에서는 카메라로부터 입력되는 영상으로부터 의미 있는 정보를 추출하기 위해 다양한 형태의 노력을 기울여 왔다. 특히 영상 매칭(image matching)분야는 다양한 접근법과 과정으로 연구되어왔고, 영상 매칭을 이용하는 다양한 응용분야의 발전을 가져왔다. 영상 매칭은 객체인식, 스테레오정합, 다중영상으로부터 3차원 구조물 복원, 동작추적, 장면인식 등의 분야에 활용되어 왔다.

본 논문에서는 영상 특징 점(feature point)[1] 기반의 객체 인식 시스템을 제안한다. 이는 영상데이터로부터 스케일-아핀 불변(scale-affine invariant)의 특징 점을 추출하고, 새로 입력되는 영상으로부터 같은 방법으로 추출한 특징 점과의 유사도 검사를 통해 가장 높은 카운트를 가지는 위치에 해당하는 객체를 인식하는 과정을 거친다.

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 검색 시스템의 구현내용과 실험결과에 대해 기술하고, 5장에서는 본 논문의 내용을 정리하면서 향후 연구에 관하여 언급한다.

2. 관련연구

디지털 영상을 사용하는 컴퓨터 비전의 다양한 분야에서는 주어진 영상을 컴퓨터가 인식하고 해석하게 하기 위해서 특징 점이라고 일컫는, 영상내의 특징을 추출하는 과정을 반드시 거치게 된다. 따라서 특징 점은 디지털 영상처리에서 주요한 입력요소로서 작용하며, 마커리스(marker-less)한 환경에서 영상의 특징을 찾는 데 필수적인 요소이다. 본 논문에서는 이러한 영상의 특징 점을 이용하여 영상을 분석하고 인식하는 영상기반 매칭기법을 활용한다.

일반적으로 특징 점을 이용한 영상매칭에 관한 연구로는 여러 가지가 있는데, 특히 Haralick과 Shapiro[2]의 저서에서 영상 매칭을 위해 필요한 차별적인(distinctive) 특징에 대해 다음과 같은 다섯 항목으로 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 특징 점을 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]는 주어진 영상에서 추출한 특징 점들의 세트를 구성하여 프로세싱 하는 주요 과정을 SIFT(scale invariant feature transform)로 정의하였다. SIFT알고리즘의 첫 번째 단계는 영상의 모든 스케일(scale)과 위치(location)들을 찾아서 정의한 후, DoG(difference-of-Gaussian)함수를 이용하여 카메라의 시점 변경에 의한 스케일-아핀 변화에 영향을 받지 않는 점들을 식별해낸다. 식별된 점들은 각각의 키 포인트가 위치하고 있는 영역의 주변영역에 대한 정보와 함께 계산해 둘 수 있다. 이것을 정 방향 거리(square distance)라 하는데, 이는 영상의 그라디언트 방향의 벡터들을 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 키 포인트의 유형을 정의할 수 있다고 한다. 이러한 접근법은 영상 내에서 추출한 포인트(corner-like points)나, 영역들의 그룹을 정립된 몇 개의 특징 점으로 표현함으로써, 영상의 스케일이나 아핀 특성의 변화가 있을 경우에도 해당하는 특징 점을 찾을 수 있다는 장점이 있다.

3. 특징 점 추출과 객체인식

디지털 영상을 사용하는 컴퓨터 비전의 다양한 분야에서는 주어진 영상을 컴퓨터가 인식하고 해석하게 하기 위해서 영상내의 특징을 추출하는 과정을 반드시 거치게 된다. 본 절에서는 인공위성 영상으로부터 특징 점을 추출하고, 추출한 특징 점의 특성을 활용하여 영상매칭과 인식, 카메라의

움직임 분석에 대해 서술토록 하겠다.

3.1. 시스템 개요

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

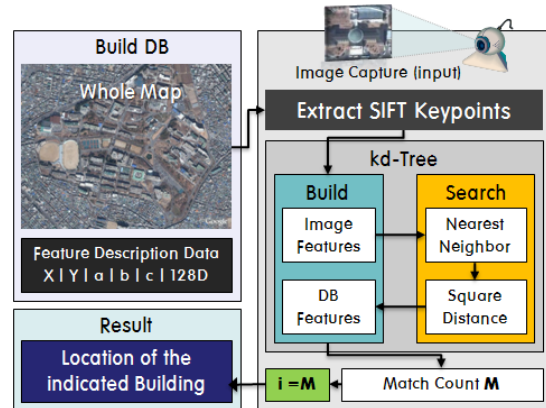


그림 1. 시스템 도표.

본 시스템은 PC용 웹캠으로부터 입력된 영상으로부터 특징 점을 추출하고, 기 추출하여 데이터베이스화 된 전체 인공위성 지도영상에 대한 특징 점의 해석 데이터와 비교하는 과정을 거쳐 영상 매칭을 하는 작업을 수행한다. 특징 데이터에 대한 구조화와 검색과정은 kd-tree를 사용하였고, 특징 데이터 비교과정에서 가장 많은 카운트를 가지는 특징 데이터 클러스터의 인덱스가 가리키는 좌표가 해당 위치라고 판단하게 된다.

3.2. 스케일-아핀 불변 특징 점 추출

본 연구에서는 SIFT 정점 추출기를 사용하여 특징 점을 추출하였다. SIFT는 크게 검출기와 분석기로 나뉘는데, 검출기는 MSERs[4]과 같은 원리로 원형의 영역내부의 그라디언트 값을 이용하여 특징 점을 찾고, 특징 점을 둘러싼 원형의 스케일과 회전정도를 파악하여 정규화 시킨다[5]. 찾아진 특징 점들은 DoG극값과의 조화정도를 이용하여 스케일 공간 극값(scale-space extrema)을 찾는다. 찾아진 값은 특징 점 점(feature point)으로 정의되며, 위치를 측정하고 필터링을 하여 사용하기에 좋지 않은 특징 점들을 개선하는 과정을 거친다. 그림 2는 구글 어스로부터 획득한 경북대학교 위성사진의 특징 점 추출 예이다.

3.3. 특징 점 데이터베이스 구축

특징 점 데이터베이스는 특징 점에 대한 기하학적정보와 해당 특징 점들의 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

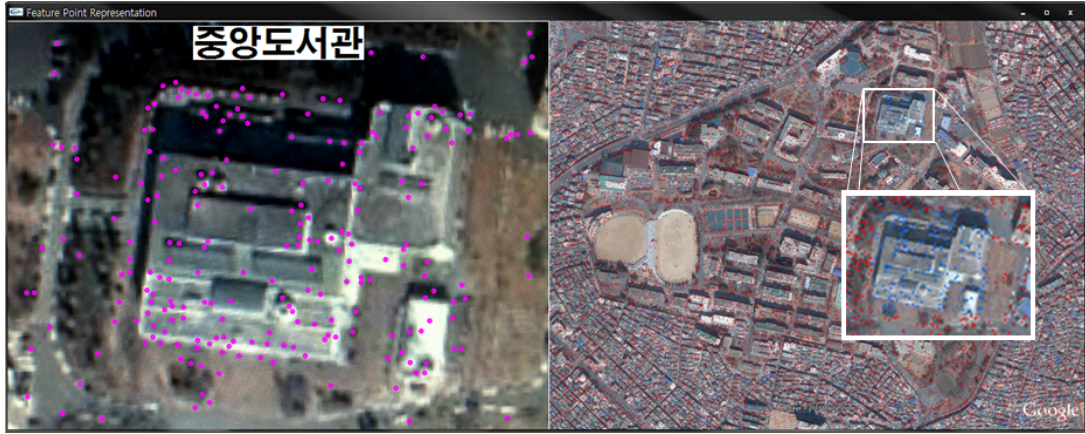


그림 2. 건물인식 예, 입력된 장면이 전처리과정에서 트레이닝 된 키포인트와 매칭 되는 점(좌측, 붉은 점)과 전체 인공위성 영상에서 해당 건물의 매칭 점(우측, 파란 점), (1280×999×24bit, 24,245개의 특징 점, 경북대학교, 위성영상 출처: 구글 어스).

ID와 TAG를 할당하였다. ID는 객체를 정의하기 위해서 할당하고, TAG는 해당 객체에 대한 정보를 삽입하기 위해서인데, 여기서는 시설물의 이름으로 정의하였다. 특징 점 분석 데이터는 기본적으로 특징 점의 2차원 영상내의 좌표 X, Y와 영역에 대한 변환관계 a, b, c(affine region parameter), 그리고 각 특징 점마다 고유한 description 데이터가 있다. SIFT를 이용한 본 실험에서는 한 지역에 대한 영상의 그라디언트의 벡터가 각각이 8방향의 오리엔테이션으로 이루어진 4×4배열로 구성하였으므로, 각각의 특징 점마다 4×4×8=128개의 키포인트 벡터들로 이루어진다.

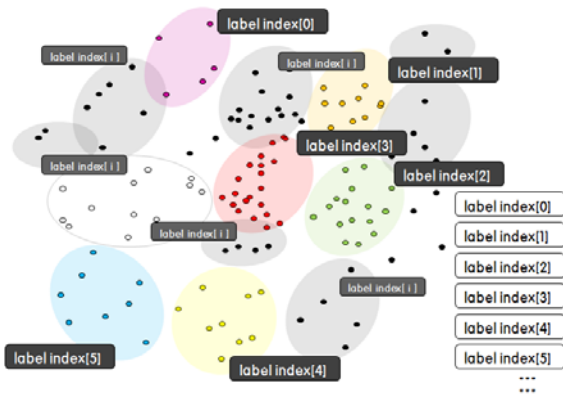


그림 3. 키포인트 클러스터링과 라벨 표시.

키포인트 추출과 벡터 데이터로 표현하는 작업이 완료되면 전체 지도에 대한 키포인트들에 대해 지역적으로 묶어주는 과정인, 클러스터링을 수행한 후 각 클러스터마다 인덱스에 상응하는 라벨을 붙여 준다[6]. 이는, 차 후 프로세스에서 키포인트 벡터 값 비교로 매칭 되는 지역에 대한 위치를 파악하는데 사용 된다(그림 3).

3.4. 특징 점 색인과 검색

특징 점 색인과 검색과정은 기본적으로 kd-tree를 이용

하여 영상 특징 점의 근사 k개의 최 근사 이웃들(k-nearest neighbors)을 찾고, 각 이웃 특징 점들의 분석 거리(distance)가 증가하는 순서대로 정렬한다. 본 논문에서는 2차원 평면 영상의 특징 점들을 kd-tree[8]로 구성하였으므로, 유클리디안-공간(euclidean-space)상에서 최 근사거리 검색(approximate Nearest Neighbor search)[7]을 통하여 같은 특징 점들을 검색할 수 있고, 각 특징 점들의 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.

3.5. 특징 점 기반의 모션필드(motion field) 분석

그림 4를 보면 한 건물에 대한 카메라의 움직임에 따른 매칭 점의 이동경로를 파악 할 수 있다. SIFT알고리즘으로 추출된 특징 점들은 스케일-아핀 불변한 특징을 가지고 있으므로, 대체로 매우 정확한 정도의 영상 매칭 결과를 이끌어 낼 수 있다. 따라서 카메라가 이동할 때 입력되는 영상으로부터 매칭경로를 추적해 보면 카메라의 움직임을 추정 할 수 있다.

표 1. 카메라 동작 추정을 위한 모션 필드의 형태.

타겟중심회전(R_{tr})	평행이동(T_{tr})	다가감(Z_{in})
타겟중심회전(R_{td})	평행이동(T_{td})	멀어져감(Z_{out})

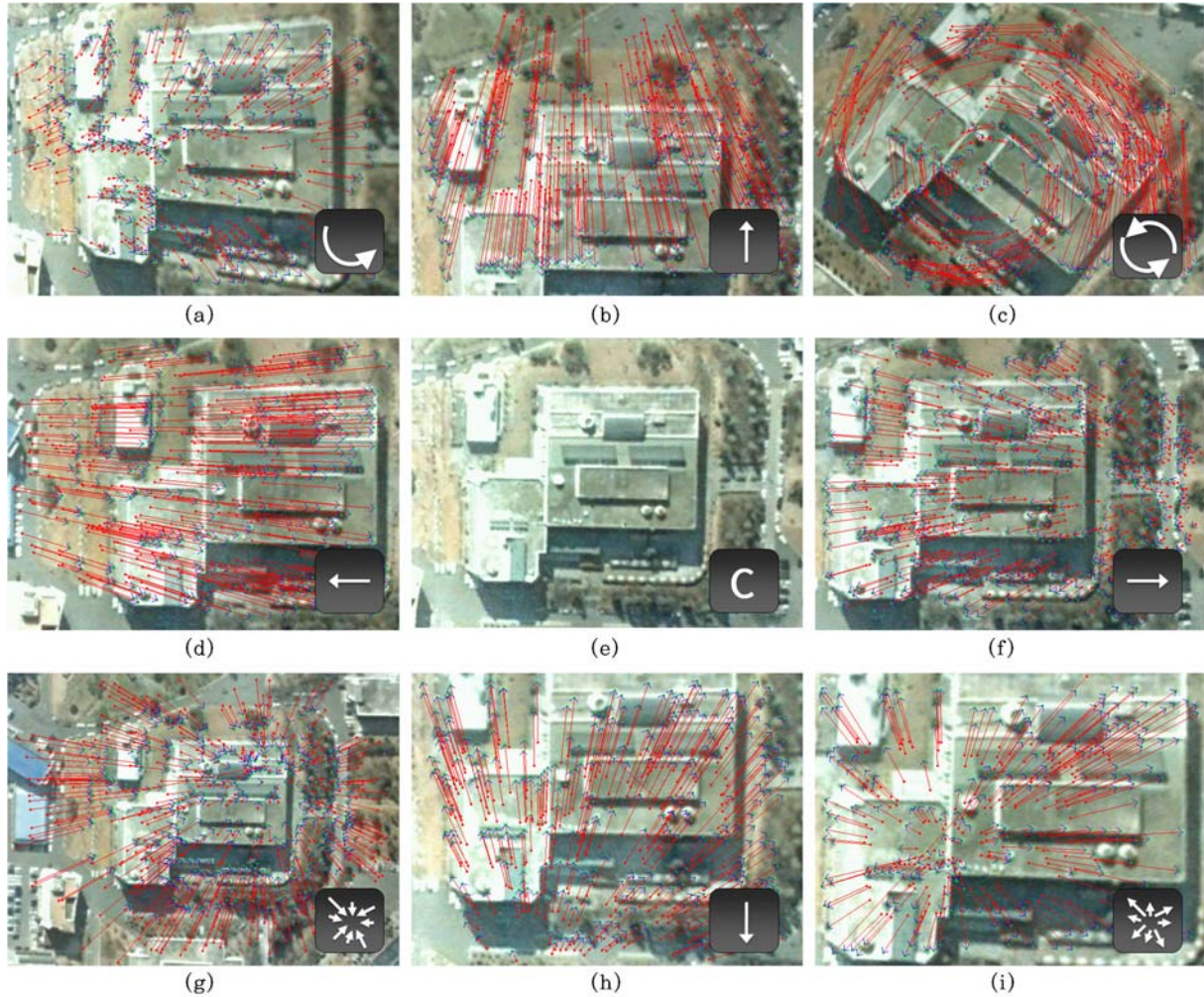


그림 4. SIFT 키포인트 매칭 점(matching point)으로 생성된 모션필드(motion field)로부터 카메라의 움직임 추정할 수 있다. (a)정 상 위 방향에서 우측으로 회전, (b)위쪽으로 이동, (c)정 회전, (d)좌측으로 이동, (e)기준 영상, (f)우측으로 이동, (g)멀어져감(zoom out), (h)아래쪽으로 이동, (i)다가감(zoom in).

그림 4는 아래 표 1의 카메라의 대표적인 인터랙션 대한 모션 추정결과를 SIFT 알고리즘을 활용한 특징 점 기반의 영상 매칭으로 파악할 수 있음을 보여준다.

4. 구현 및 실험결과

본 논문에서는 SIFT알고리즘을 기반으로 하여, 인공위성 으로부터 획득한 위성영상에서 영상의 부분을 인식하는 시스템을 구현하였다. 일반적으로 시스템을 구현할 때, 시스템의 응답률을 높이기 위해 트레이닝과정을 거치거나 빌드하기 위한 데이터의양을 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 목적이므로 여타의 트레이닝 과정은 거치지 않도록 실험환경을 설정하였다.

4.1. 시스템 구현환경

본 논문에서 제안한 시스템은 인쇄된 인공위성 영상을 PC용 USB웹캠을 이용하여 위성 영상의 특정 한 부분을 촬영하고, 카메라로 입력된 영상 시퀀스로부터 특징 점을 추출한 후, 기 구축된 특징 점 DB로부터 해당 영상 내부에 존재하는 건물을 인식하는 작업을 수행한다.

본 실험에서 사용한 시스템 환경은 퍼스널 컴퓨터로 사용하는 PC사양에서 구현되었고 테스트 되었다. 카메라로 입력되는 영상 시퀀스에서 나타나는 지도상의 특징 점을 추출하고, 스케일-아핀 불변의 키포인트를 정리하는 과정에는 SIFT알고리즘을 사용하였다. 카메라와 영상을 컨트롤하기 위하여 인텔 OpenCV 컴퓨터비전 라이브러리를 이용하였고, 영상을 입력 받기위해서 640x480의 비디오 해상도와 초당 25프레임을 지원하는 USB웹캠을 사용하였다.

4.2. 실험결과

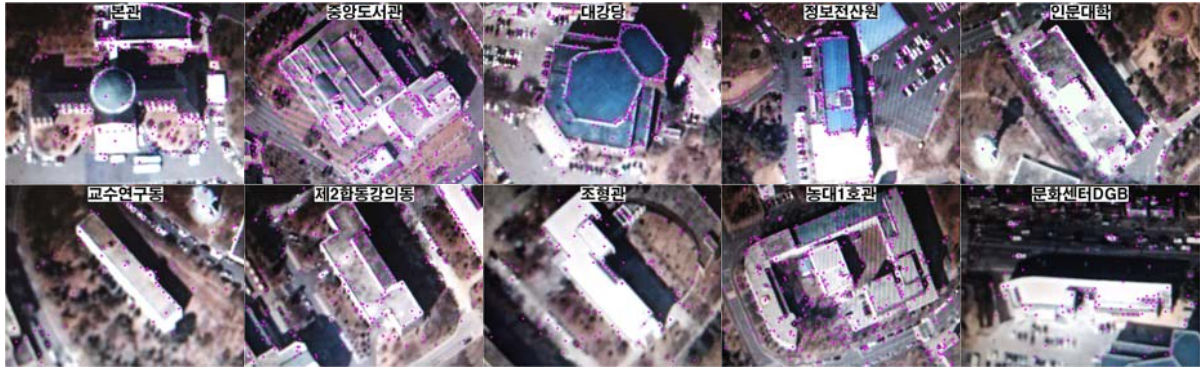


그림 5. 10개 건물에 대한 매칭 결과영상.

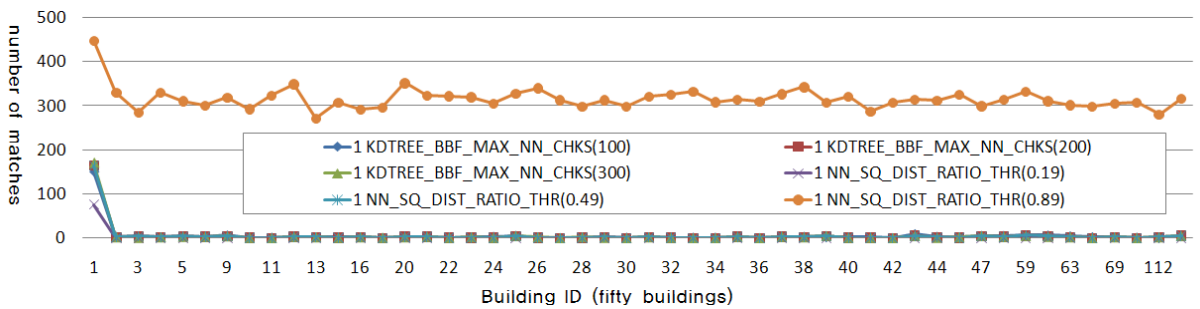


그림 6. 검색파라미터 변화에 따른 매칭 응답률(1번 건물을 입력영상으로 하였을 경우).

본 논문에서는 대형 지도 내부에 존재하는 객체의 위치를 추정하기 위해서 SIFT 알고리즘을 사용하였으며, 이는 실험 대상으로 사용된 전체 지도 내에 존재하는 특징 점을 검출하고, 카메라로 입력되는 영상의 특징 점과 비교하여 서로간의 매칭관계를 비교하여 카메라로 입력되는 영상에 해당하는 객체의 위치를 추정하는 방법을 사용했다. 그림 2의 우측 영상과 같이 전체 지도에 대한 특징 점을 미리 검출 해 둔 뒤, 입력되는 영상에서 추출된 특징 점과의 매칭 카운트가 가장 높은 객체에 해당하는 특징 점의 세트를 가진 인덱스를 출력하도록 하였다. 결과는 그림 5에서 확인할 수 있으며, 이는 경북대학교 위성사진 상에 존재하는 50여개 이상의 건물 중에 10개 건물에 대한 결과 영상이다. 마찬가지로 그림 6에서는 1에서 112의 인덱스 번호를 가진 50개의 건물을 대상으로 1번 인덱스를 가진 건물에 대해 매칭 결과를 비교해 보면, 1번 건물에 대한 매칭 카운트가 가장 높게 나타남을 알 수 있다. 그림 6의 그래프에서는 각각 2가지 파라미터에 대한 3단계 변화에 따른 검색결과(매칭 카운트)를 비교하였으며, 어떠한 파라미터의 변화에도 관계없이 검색결과가 옹게 나타남을 볼 수 있다.

또한 그림 4를 보면 SIFT특징 점의 매칭관계를 이용하여 카메라의 동작을 추정할 수 있다는 가능성을 보여주었으며, 이는 차후 연구에서 카메라를 활용한 장면 내비게이션이나 카메라의 위치나 자세를 추정하는데 활용할 수 있을 것이다.

5. 결론 및 의의

본 논문에서는 인공위성 영상을 기반으로 SIFT알고리즘을 활용하여 카메라로 입력되는 영상의 위치를 추정하는 방법에 대해 실험 하였다. 본 논문에서 제안한 영상기반 검색 시스템은 기존의 SIFT를 활용하여, 불변하는(invariant) 영상특징 점들을 추출하고 매칭 하는 과정을 기반으로 구동되기 때문에 프로세스의 양이나 시간적인 차원에서 보다 효율적인 방향으로 개선될 필요가 있다. 먼저 위성영상에 적합한 특징 점들에 대해, SIFT에서 사용하는 MSERs특징 점 디텍터만 사용할 것이 아니라, 속도는 느리지만 스케일에 따른 응답률이 비교적 높게 나오는 Harris-Laplacian 디텍터[5]를 사용하여 속도문제를 개선한다거나, 속도는 빠르지만 응답률은 다소 떨어지는 Lepetit[8]의 특징 점 검출법을 사용해 보는 것도 먼저 시도해 봄직하다. 이는 프로세스가 낮은 기기에서 사용하거나 처리속도의 개선을 위해서, 보다 효율적인 최적조건을 찾는 것이 앞으로 풀어야 할 문제일 것이다. 또한 영상매칭을 하는데 사용하는 불변특징 점들의 기하학적 변환관계(affine transformation)를 잘 활용하여, 본 정보검색 시스템의 가시화 단계에서는 AR기법을 적용하여, 보다 높은 차원정보들을 가시화 할 수 있도록 연구되어야 할 것이다.

감사의 글

본 논문은 Brain Korea 21 정보기술 인력양성사업 과제에 의하여 지원 되었습니다.

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] 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.
- [5] K. Mikolajczyk and C. Schmid, Scale & Affine Invariant Interest Point Detectors, International Journal of Computer Vision, Vol. 60, Issue 1, pp. 63–86, 2004.
- [6] Volker Gaede and Oliver Gunther, Multidimensional Access Methods, ACM Computing Surveys, Vol. 30, Issue 2, pp. 170–231, June 1997.
- [7] 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.
- [8] 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.
- [9] Andrea Vedaldi, Source code and Manual: An implementation of SIFT detector and descriptor, <http://vision.ucla.edu/vedaldi/code/sift/sift.html>, VisionLab, UCLA.