도시 영상에서의 Inlier 선택과 Database Redundancy 감소 기법

*안하은 **유지상
광운대학교
*mysco226@kw.ac.kr

Inlier selection and Database Redundancy Reducing Method in Urban Environment

*Ahn, Ha-eun **Yoo, Jisang
Kwangwoon University

요약

특징점 기반 건물인식 시스템에서는 강건한 특징점을 추출하는 것이 인식률 향상에 바로 직결되는 중요한 요소이다. 영상에서 특징점들이 너무 많이 추출되는 경우 인식이나 학습단계에서의 알고리즘 수행 시간을 증가시키는 원인이 된다. 또환 중요하지 않은 특징점(배경이나 가려짐 영역, 기타 객체에서 추출된 특징점)이나 조명 변화에 민감한 영역에서 임의로(arbitrarily) 추출된 특징점은 인식률을 저하시키는 문제를 발생시킨다. 특히 도시환경에서 촬영된 영상의 특징점을 추출할 때 이러한 문제 현상들이 빈번하게 발생한다. 본 논문에서는 이러한 문제를 해결하고자 multi-view 영상에서 건물의 homography를 기반으로 정확히 정합된 특징점인 inlier만을 선택하는 알고리즘을 제안한다. Inlier로 분류된 특징점들은 건물 인식 시스템을 구성하기 위해 사용되고 조명 변화에 민감한 영역에서 임의로 추출된 특징점들은 영역 기반 특징을 추출하여 건물 인식 시스템의 인식률을 높인다. 또한 이를 이용하여 인식하고자 하는 건물과의 상관관계가 적은 잉여 영상들을 DB에서 제거하는 방법도 제안한다. 실험을 통하여 제안하는 기법의 우수성을 보였다.

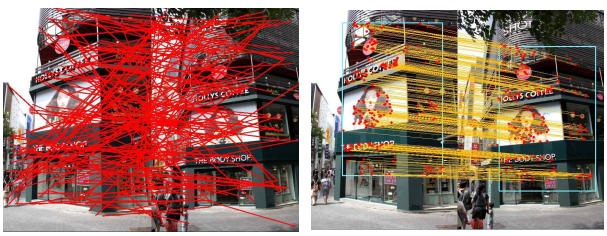
1. 서론

도시 환경에서 여러 가지 사물을 인식하는 기술은 다양한 컴퓨터 비전 응용에 적용될 수 있는 핵심 기술이다. 최근에는 사물 인식 기반 증강현실의 형태로 사물의 정보를 사용자에게 제공하는 서비스가 크게 주목받고 있다. 건물은 도시 환경에서 가장 많이 존재하는 객체 중 하나이며 따라서 건물 인식과 관련된 연구가 많이 진행되었으며 지금도 인식률을 높이기 위한 연구가 활발하게 진행되고 있다.

건물 인식은 주로 여러 가지 종류의 특징을 이용하는 특징 기반으로 연구되어 왔다[1]. 최근에는 특징점 추출 방법과 vocabulary tree를 이용하여 건물을 인식하고 spatial consistency를 측정하여 인식률을 향상시키는 방법[2], 추출된 특징을 바탕으로 기계학습을 이용하는 방법[3] 등이 제안되었다. 이러한 방법은 대규모의 database에 적용될 경우 기존 기법들보다 우수한 인식률을 보인다. 하지만 가려짐 영역이 발생하거나 건물 이외의 객체가 다수 포함된 경우 특징점이 오정합된 outlier가 많이 발생하여 인식률이 낮아지는 문제가 있다.

특징점 기반의 건물인식 방법에서는 강건한 특징점을 추출하는 것이 인식률 향상에 가장 중요한 요소이다. 특징점이 너무 많이 추출되는 경우, 인식이나 학습단계에서의 프로세싱 시간이 증가되는 원인이 된다. 또한 중요하지 않은 특징점(배경이나 가려짐 영역에서 추출된 특징점)이나 임의로 추출된 특징점(텍스트 영역 등에서 추출된 특징점)은 인식률 저하에 영향을 미친다. 특히나 상업단지나 도시 환경에서 촬영된 건물 영상은 가려짐 영역이나 배경에서 많은 특징점이 추출된다. 특정 상표나 간판에 존재하는 텍스트 영역이 많아 임의로 추출된 특징점도 많다. 이런 영역에서 추출된 특징점들은 인식률을 저하시키는 주된 원인이 된다. [4]에서는 특징영역을 추출하여 건물을 인식하는 방법을 제안하였다. 특징점보다 강건한 특징영역을 추출함으로써 인식률을 크게 증가시켰다. [5]에서는 영역특징을 트래킹 하는 방법들도 제안되었다. 본 논문에서는 이에 영감을 받아 조명 변화에 민감한 영역에 대해서는 영역기반의 특징을 추출하여 inlier를 선택하는 방법을 제안한다.

그림 1은 도시 환경에서 촬영한 두 장의 영상에서 특징점을 추출한 뒤 특징점 정합을 수행한 결과이다. 그림 1(a)는 가려짐 영역과 배경에서 너무 많은 특징점이 추출되었기 때문에 다수의 outlier를 inlier로 오정합하고 있다.



(a) (b)

그림 1. 도심 환경 영상에서 수행한 특징점 정합
(a) 기존 방법, (b) 제안하는 방법

본 논문에서는 기존의 건물 인식 기법들의 다양한 문제점들을 해결하고 도시 환경에서의 건물 인식률을 향상시키기 위해 inlier만을 선

택하는 방법을 새로이 제안한다. 제안하는 기법에서는 multi-view 영상 간에 같은 객체의 homography 변환을 활용하여 inlier만을 선택한다. 텍스트 영역이나 반복적인 건물의 패턴을 가지는 영역에서는 특징점 추출의 반복성이 떨어지기 때문에 해당 영역에 대해서는 영역 기반 특징점 추출 방법을 이용한다. 또한 inlier가 적은 잉여 영상은 database에서 제거하여 효율적인 인식 시스템을 구성하는 방법도 제안한다.

2. 본론

도시환경에서 촬영된 건물 영상들은 일반적으로 다른 건물들, 가로수, 각종 표지판 등 복잡한 배경과 오토바이, 자동차, 보행자 등 건물 이외의 객체를 포함하고 있다. 특히 상가건물인 경우 보행자, 오토바이, 표지판, 가로수, 간판 등 원하지 않는 객체가 많이 존재한다, 따라서 이런 경우 추출되는 outlier들은 건물의 인식률을 저하시키는 원인이 된다. 제안하는 기법에서는 multi-view 영상에서 추출된 특징점들의 정합 쌍을 찾고 이들을 이용하여 건물의 homography 변환 행렬을 구한다. 특징점 정합 쌍에는 가려짐 영역이나 배경에서 추출된 outlier들에 의해 오정합된 쌍들도 다수 존재하기 때문에 건물의 정확한 homography 변환 행렬을 구하는 것이 쉽지 않다.

제안하는 기법에서는 양질의 정합 쌍만을 선별하여 homography를 정의한 뒤 이를 이용하여 multi-view 영상에서 추출된 특징점들의 정합 쌍을 다시 정의하게 된다. 여기서 양질의 정합 쌍은 정합된 특징점 쌍의 두 특징점의 descriptor vector가 서로 유사한 경우로서 대체적으로 영상에서 건물 객체 등 중요한 영역에서 추출되는 특징점들이 이에 해당되며 특징점 추출의 반복성이 강한 특징을 가지고 있다. 이러한 특징점들을 이용하여 homography를 정의할 경우 객체 영역에서의 변환관계를 잘 표현할 수 있고 동시에 배경이나 가려짐 영역 그리고 기타 객체들이 존재하는 영역을 배제하여 inlier들이 존재할만한 후보 영역만을 찾을 수 있다. 식 (1)에서 정의된 정합된 특징점 쌍의 정합도(distance)를 계산함으로써 양질의 정합 쌍을 찾을 수 있다.

$$Distance^i = \sum_{d=1}^{D} \left(p_d^i - q_d^i\right)^2, \text{ for } all\ i \quad (1)$$

여기서, $Distance^i$는 i번째 특징점 쌍의 정합도, D는 특징점 descriptor vector의 차원, p_d^i와 q_d^i는 각 multi-view 영상에서 추출된 i번재 특징점 descriptor vector의 d번째 element를 나타낸다. 제안하는 방법에서는 $Distance^i$ 배열에 대하여 소팅(sorting) 과정을 수행한 후 정합이 가장 잘된 특징점 쌍들을 선택하여 homography를 찾는다. 실험을 통하여 평균적으로 상위 30개의 정합 쌍을 사용하였을 때 정확도가 높은 homography를 찾을 수 있다는 것을 확인하였다. 이 이상의 정합 쌍을 사용하면 homography의 정합도가 포화상태가 된다.

특징점이 정확하게 정합되었다는 것은 특징점 정합 쌍의 descriptor vector가 서로 유사하다는 것을 의미한다. 이는 각 특징점 정합 쌍의 특징점들이 객체의 동일한 위치에서 추출된 경우이다. 본 논문에서는 이러한 특징에 착안하여 multi-view 영상에서 추출된 특징점에 대하여 보다 신뢰도 높은 특징점 정합 쌍을 찾는 방법을 제안한다. 첫 번째 multi-view 영상에서 추출된 특징점($P_{(x,y)}^k$)을 두 번째 multi-view 영상 좌표계로 투영한 위치에서 추출되는 특징점은 $P_{(x,y)}^k$ 점과 정합 쌍일 가능성이 높다. 제안하는 기법에서는 첫 번째 multi-view 영상에서 추출된 특징점 $P_{(x,y)}^k$ 점을 두 번째 영상으로 투영한 뒤 투영 좌표의 주변영역에서 추출되는 특징점들을 정합 쌍 후보군으로 설정한다. 특징점 $P_{(x,y)}^k$와 후보군들의 descriptor vector를 비교하여 유사한 값을 가지는 경우에 대하여 특징점 정합 쌍을 정의한다. 그림 2는 후보군들 중 특징점 $P_{(x,y)}^k$와 가장 유사한 descriptor vector를 가지는 특징점을 찾아서 해당 특징점이 $P_{(x,y)}^k$의 올바른 특징점 정합 쌍인지 확인하는 과정을 나타낸다.

$$\text{if } \left(Distance^k \le \frac{1}{N}\sum_{i=1}^{N} Distance^i\right)$$
$$\rightarrow inlier$$
$$else$$
$$\rightarrow outlier$$

그림 2. Inlier 선택 방법에 대한 의사 코드

그림 2에서 $Distance$는 식 (2)의 특징점 정합도를 나타내고 N은 homography를 찾을 때 사용된 특징점의 개수를 나타낸다. 올바른 정합 쌍으로 판별되는 경우에는 이를 inlier로 정의한다. Outlier로 판별된 특징점들이 밀집된 영역에서는 영역 기반 특징점 추출 방법을 적용한다. Outlier로 판별되는 특징점들이 밀집된 영역은 난반사가 심한 유리 외벽, 외부 조명이 존재하는 영역이나 텍스트가 존재하는 영역, 조명 변화에 따른 화소 값 변화가 심한 영역 등이며 이러한 영역에서는 특징점이 임의로 추출되는 문제가 존재한다. 제안하는 기법에서는 이러한 영역에 대해 dense SIFT를 이용하여 특징을 추출하는 방법을 제안한다.

그림 3. 특징점이 임의로 추출되는 영역의 예

그림 3은 조명변화에 민감한 영역이나 상업 단지의 상표, 표지판 등 텍스트 영역에서 추출된 특징점을 보여준다. 이러한 영역에서는 특징점이 임의로 추출되기 때문에 정합이 제대로 이루어지지 않는다. 특히 건물에서 추출되는 특징점 임에도 불구하고 특징점들의 descriptor vector가 서로 상이하기 때문에 outlier로 분류된다. 제안하는 기법에서는 이러한 문제를 해결하기 위하여 특징점이 임의로 추출되는 영역에서는 MSER(Maximally Stable Extremal Region)을 기반으로 dense SIFT를 추출한다. 그림 4는 MSER을 이용하여 추출된 영역들을 보여준다. 특징점이 임의로 추출되는 영역에서는 특징점들의

descriptor vector가 서로 다르기 때문에 특징점 정합이 발생할 수 없다. 따라서 제안하는 방법에서는 MSER을 추출한 뒤 해당 영역에 타원을 피팅(fitting) 한 뒤 dense SIFT를 추출하여 inlier들을 선택한다.

그림 4. 조명변화에 민감한 영역에서 추출된 MSER

그림 5는 MSER에서 dense SIFT를 추출하기 위하여 local patch를 지정하는 방법을 보여준다. SIFT에서 특징점의 dominant orientation으로 local patch를 지정하는 것과 유사하게 MSER에 타원을 피팅 시킨 후 이를 감싸는 사각형을 local patch로 지정하여 dense SIFT를 계산한다. MSER은 특징점보다 저조한 특징 추출 반복성을 가지기 때문에 각 MSER마다 계산된 dense SIFT들은 별도의 특징 정합과정 없이 inlier로 선택할 수 있다.

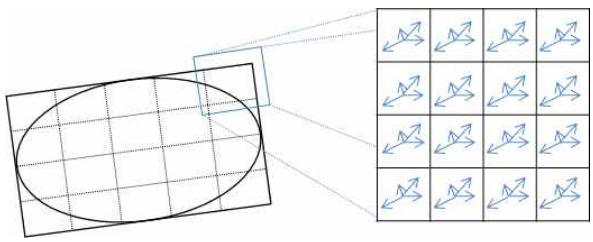


그림 5. MSER에서 dense SIFT를 추출하는 방법

3. 실험 결과

일반적으로 DB 크기의 증가는 인식률 개선에 도움이 된다고 알려져 있지만 과도하게 방대한 양의 DB에서 특징을 추출하여 인식 시스템을 구성하는 것은 효율적이지 않은 방법이다. 또한 인식하고자 하는 건물과 상관관계가 떨어지는 영상들은 DB에서 제거하는 것이 시스템 구성 시 시간적인 측면이나 메모리 관리 측면에서 유리하다.

Inlier가 많이 선택되는 영상은 인식하고자 하는 건물에서 양질의 특징점이 다량 검출 되었다는 것을 의미한다. 동시에 inlier가 적게 선택되는 영상들은 건물보다 배경이나 기타 객체에서 특징점들이 추출되었음을 의미하고, 인식률 증가에 큰 영향을 미치지 못한다. 따라서 제안하는 기법에서는 inlier가 적게 선택되는 영상은 DB에서 제거하여 인식 시스템을 구성한다. 참조 multi-view 영상에서 선택되는 inlier들의 개수를 파악하여 소팅한 후 database 활용율을 조절하여 인식 시스템을 구성한다. 본 논문에서는 database 활용율에 따른 에러 측정실험을 진행하였다.

표 1은 DB database 활용율 변화에 따른 에러를 보여준다. 그림 6은 제안하는 방법을 이용하여 database 활용율을 조정한 결과와 무작위로 database 활용율을 조정한 결과의 정확도 차이를 보여준다. 그림 6의 검은 점선은 제안하는 방법을 이용하여 inlier의 개수가 적고 건물의 특징을 잘 반영하지 못하는 영상을 database에서 우선적으로 제거하여 구성한 인식 시스템의 에러를 보여준다. 붉은 점선은 무작위로 database를 감소시켜 구성한 인식 시스템의 에러를 보여준다.

표 1. 빌딩 인덱스와 database 활용율에 따른 에러

Building Index	Top-1 err (100%)	Top-1 err (80%)	Top-1 err (60%)
A	21.4	25.4	34.3
B	14.7	18.5	28.6
C	11.8	16.5	27.7
D	20.6	24.3	30.9
E	19	21.8	29.7
F	19.1	20.7	31.1
G	27.3	29.5	34.8
H	7.6	9.3	27.8
I	8.5	13.2	28.5
J	21.7	24.1	37.3

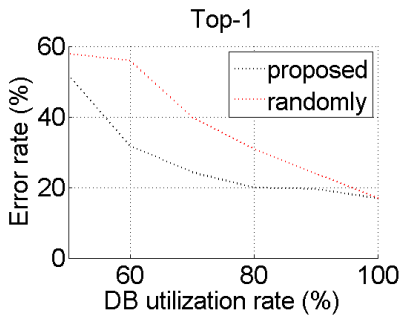


그림 6. Database 활용율 변화에 따른 인식률 차이
(검은 점선: 제안하는 시스템의 에러, 붉은 점선: 무작위로 database를 감소시켜 구성한 시스템의 에러)

3. 결론

본 논문에서는 multi-view 영상에서 배경이나 가려짐 영역 혹은 외부 객체에서 추출되는 outlier들을 제거하고, 건물 객체에서 추출되는 inlier를 효율적으로 선택하는 기법을 제안하였다. Multi-view 영상에서 높은 신뢰도를 가지는 특징점 정합 쌍을 이용하여 homography 변환 행렬을 구하고 이를 이용하여 특징점 정합 쌍을 새로 정의하였다. 또한 벽면의 유리와벽이나 텍스트 영역같이 조명 변화에 따라 화소 값의 변화가 심한 영역에서는 MSER(maximally stable extremal regions) 기반 dense SIFT를 추출하여 특징의 반복성을 높이는 효과를 보였다. 참조 영상에서 획득한 inlier들을 이용하여 건물 인식 시스템을 구성하고 제안하는 기법을 이용하여 database 활용율을 조절한 결과와 무작위로 database 활용율을 조절한 결과를 비교하여 제안하는 기법이 우수하다는 것을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.R0132-16-1005, 온-오프라인에서의 콘텐츠 비주얼 브라우징 기술개발)

참 고 문 헌

[1] J. Li, W. Huang, L. Shao and N. Allinson, "Building recognition in urban environments: A survey of state-of-the-art and future challenges", Information Sciences, vol. 277, no. 1, pp. 406-420, Sept. 2014

[2] S. H. Said, I. Boujelbane and T. Zaharia, "Recognition of urban buildings with spatial consistency and a small-sized vocabulary tree", 2014 IEEE Fourth International Conference on Consumer Electronics, Berlin, pp. 350-354, Sept. 2014.

[3] J. Li and N. Allinson, "Building recognition using local oriented features", Industrial Informatics, vol. 0, no. 3, pp. 1697-1704, Aug. 2013.

[4] Y. Chung, T. X. Han and Z. He, "Building Recognition Using Sketch-Based Representations and Spectral Graph Matching", IEEE 12th International Conference on Computer Vision Sept, 2009.

[5] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (MSER) tracking", Computer Vision and Pattern Recognition, vol. 1, pp.17-22, June. 2006