

영상 특징 검출 기반의 실시간 실내 장소 인식 시스템

송복득¹, 신범주^{1,a}, 양황규²

¹ 부산대학교 생명자원과학대학 IT응용공학과

² 동서대학교 컴퓨터정보공학부 소프트웨어공과

A Real-time Indoor Place Recognition System Using Image Features Detection

Bok-Deuk Song¹, Bum-Joo Shin^{1,a}, and Hwang Kyu Yang²

¹ Department of IT Application Engineering, Busan National University, Milyang 627-706, Korea

² Division of Computer and Information Engineering, Dongseo University, Busan 617-716, Korea

(Received December 20, 2011; Revised December 24, 2011; Accepted December 24, 2011)

Abstract: In a real-time indoor place recognition system using image features detection, specific markers included in input image should be detected exactly and quickly. However because the same markers in image are shown up differently depending to movement, direction and angle of camera, it is required a method to solve such problems. This paper proposes a technique to extract the features of object without regard to change of the object scale. To support real-time operation, it adopts SURF(Speeded up Robust Features) which enables fast feature detection. Another feature of this system is the user mark designation which makes possible for user to designate marks from input image for location detection in advance. Unlike to use hardware marks, the feature above has an advantage that the designated marks can be used without any manipulation to recognize location in input image.

Keywords: Image features detection, Place recognition, Place recognize marker

1. 서 론

오늘날 정보통신 기기 및 소프트웨어 기술의 발달로 인간의 편리함을 위해 로봇 기술에 대한 기술들이 각광받고 있다. 이러한 환경에서 로봇을 이용한 자율 항해 및 장소의 인식 영역에 대한 많은 연구가 수행되었다. 장소 인식에 대한 여러 가지 연구 방법 중 실내에 위치한 특정 장소를 인식하는 방법은 크게 센서를 통해 인식하는 방법과 카메라에 입력되는 영상을 통해 인식하는 방법이 있다 [1,2]. 센서를 이용하는 방법은 센서의 특징이나 환경의 변화에 따라 인식

률에 차이가 날 뿐 아니라, 센서 설치에 따른 경제적, 시간적인 면에서 비효율적이다 [3-8]. 반면 영상 인식 기술을 사용하는 방법은 카메라와 입력 영상을 처리할 수 있는 컴퓨터 기기 외에 추가 하드웨어가 필요치 않기 때문에 가장 활발하게 연구되고 있다.

영상 인식에 기반 한 장소 인식 연구는 크게 형태 정보, 색상정보, 인공 마커 (artificial marker)를 사용하는 방법, 그리고 영상 내에 존재하는 특징들을 검색하여 장소를 인식하는 방법으로 구분된다 [8-13]. 형태 정보 기반의 장소 인식 방법은 특정 마커의 형태에 대한 템플릿을 만들어서 정합하는 방법으로 복잡한 배경의 경우 오인식율이 높다. 그리고 색상 정

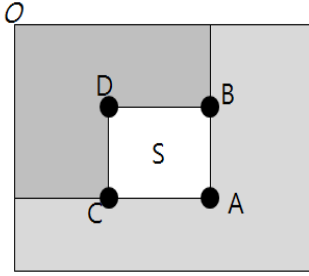


Fig. 2. Integral operation in rectangle area using integral image.

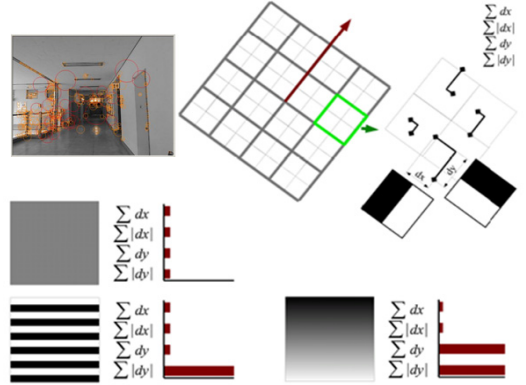


Fig. 4. Creation of SURF descriptor and applying it to contrast image.

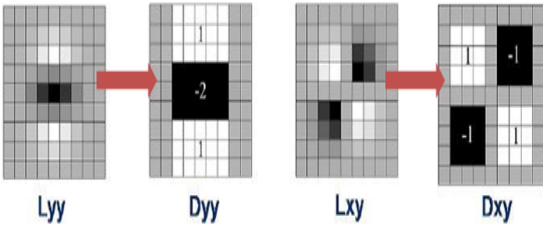


Fig. 3. The second order Gaussian partial derivative and approximated box filter.

위치의 픽셀값을 의미한다. 그림 2는 계산된 적분영상에서 사각영역의 적분값 연산을 보여주는 예로 D 영역의 적분값을 구하는 과정을 보여준다. S의 영역은 A-B-C+D를 통해 구해질 수 있다. 이와 같이 적분영상을 이용하여 4개의 포인트만 알면 그 영역의 면적을 간단하게 계산할 수 있으므로 실시간 입력 영상의 처리 속도가 빨라진다.

$$I_f(x,y) = \sum_{i=0}^x \sum_{j=0}^y I(x,y) \quad (1)$$

식 (2)에 정의된 헤시안 매트릭스에 기반을 둔 특징점 추출 알고리즘은 속도와 정확도면에서 좋은 성능을 보인다. 식 (2)에서 L_{xx} 는 x, y 위치의 입력영상과 σ 의 분산을 갖는 가우시안의 x방향의 2차 미분값과의 컨볼루션한 값을 의미하고 나머지도 x, y방향으로 미분 그리고 y방향으로의 2차 미분된 가우시안 필터와 컨볼루션한 값을 의미한다. 이는 x축, y축, 그리고 x, y축 (대각)에 대해 4개의 입력점을 헤시안으로 계산하면 관심점 (interest point) 후보자들이 선정된다.



Fig. 5. Extracting image features using SURF.

$$H = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \quad (2)$$

그림 3과 같이 입력된 영상은 9x9 박스필터를 사용하는데 계산을 좀 더 간편하고 단순하게 하기 위해서 근사화된 D_{xx} , D_{yy} 박스필터를 사용한다.

이러한 근사화된 박스필터를 사용함으로써 SIFT보다 빠른 처리 시간을 제공한다.

그림 4는 SURF의 표현자 생성 및 콘트라스트 영상 적용 과정을 나타낸다. SURF는 이웃한 점들과 비교하여 중심점이 가장 클 경우 대응점 후보자로 뽑고 표현자를 생성하기 위해서 관심점 주위에 20s 크기의 윈도우를 구성한다. 표현자 윈도우를 4x4로 다시 나누고 각각의 구역은 Haar wavelet으로 계산한다. 그리고 x, y축에 대한 응답값을 절대값으로 합한 값도

이용함으로써 대조 및 대비 영상에 강하다는 것을 간략히 보여주고 있다.

그림 5는 실시간 입력되는 영상에 SURF를 이용한 대응점(keypoint) 및 표현자를 생성하여 영상 특징점을 추출한 결과 화면이다.

2.2.2 마커 후보 영역 검출

SURF를 이용하여 실시간으로 입력되는 영상의 특징 추출 분포에 따라 장소 인식 마커와 매칭작업을 이용하여 장소인식 마커의 후보영역을 검출한다. SURF를 이용한 장소인식 마커 후보영역 추출은 2.2.1절에 소개한 콘트라스트 영상에서의 x축과 y축에 대한 절대값을 합한 값을 이용하여 단순 매칭작업을 수행할 수 있다. 그리고 이전에 헤시안 매트릭스에서 계산한 라플라시안 부호를 비교해서 간단하게 추출할 수 있으며 추가적인 연산 비용이 들지 않고 표현자의 성능 감소 없이 장소 인식 마커의 후보영역 검출 속도가 향상되는 것을 알 수 있다. 그림 6은 영상 분석을 위해 사용되는 영상 스케일링 과정을 나타낸다. SURF에서 사용되는 영상 스케일링 과정은 SIFT와 비슷하게 영상 피라미드를 구성하는데 SIFT에서는 다른 크기의 이미지에 적용하면서 스케일 불변을 보장하는 반면에 SURF에서는 박스 필터 크기를 변경해가면서 적용한다.

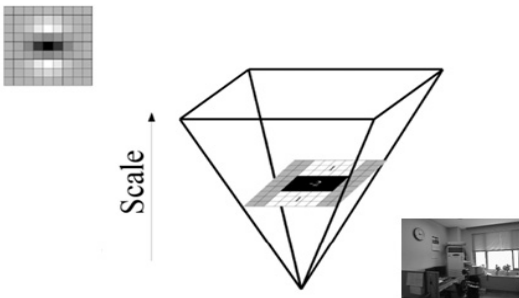


Fig. 6. The image scaling process used for scale analysis.

그림 7은 입력되는 영상에서 검출하고자 하는 인식 마커의 스케일 및 시점 변화에도 장소인식 마커의 후보 영역을 검출하는 결과화면이다. 동작하는 카메라의 위치에 따라 인식 마커의 스케일 및 시점이 변

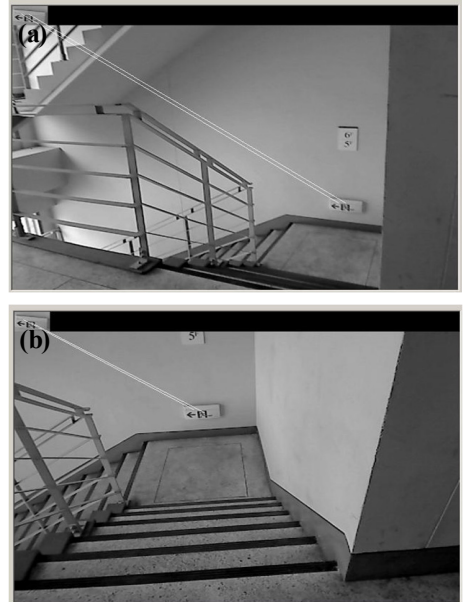


Fig. 7. Detection of same feature from two different images. Image (a) and (b) are shown different images because of rotation and different view point.

하여도 장소 인식 마커 후보영역을 검출하는 것을 확인할 수 있다.

2.2.3 인식 마커를 이용한 장소 인식

장소 인식 마커에 대해 사전에 SURF를 이용하여 영상 특징점을 추출한 후, 추출된 특징점들에 대한 모델링을 수행한다. 장소 인식 마커의 모델링 방법은 직접 인식의 대상이 되는 장소로부터 획득한 영상에서 추출한 특징점의 좌표 및 크기 등의 특징을 통해 이루어지며, 이를 데이터베이스에 저장한다.

인식 마커를 이용한 장소 인식은 인식의 대상이 되는 장소를 이동하는 카메라로 부터 실시간으로 영상을 획득한다. 획득된 영상으로 부터 특징을 추출하고, 추출된 특징은 미리 구성된 인식 마커 데이터베이스를 이용하여 레이블링 된다. 추출된 특징을 X로 표현할 때 NaiveBesian에 의해, 특정 장소 인식 마커 ω 일 확률은 베이시안 규칙에 의해 식 (3)과 같이 표현된다 [17].

$$P(\omega|X) = \frac{P(X|\omega) \cdot P(\omega)}{P(X)} \quad (3)$$



Fig. 8. Recognition of place marker from real time image.

식 (3)에서 $P(X)$ 는 항상 같기 때문에 각 레이블링 된 각 특징들에 서로 연관성 없이 출현한다고 가정하면 $P(\omega|X)$ 는 식 (4)와 같이 계산될 수 있다.

$$P(\omega|X) = \prod_i P(x_i|\omega) \quad (4)$$

식 (4)에서 X 는 위치 인식 마커의 특징 집합으로 하고 사전에 만들어진 장소인식 마커의 대표적인 특징으로 바꾸어 기술할 수 있다.

장소인식의 결과는 식 (5)에서와 같이 미리 모델링 된 모든 장소인식 마커를 대상으로 사후확률을 계산 하여 최고의 사후확률을 보이는 장소로 인식한다.

$$\text{Classify}(X) = \text{argmax}\{P(\omega = \omega_i|X)\} \quad (5)$$

3. 결과 및 고찰

3.1 시험 환경

본 논문의 시스템을 시험하기 위한 환경은 다음과



Fig. 9. The marker images used for experimental place recognition.



Fig. 10. The indoor place images used for experimental place recognition.

같다. 시험 환경은 영상인식 프로그램을 수행할 노트북 1대와 영상을 획득하기 위한 웹카메라 1대로 구성된다. 노트북의 운영체제는 WindowsXP이며, 카메라는 Logitech WebCam Pro9000이다.

장소인식을 위한 영상은 초 당 30 프레임으로 입력되며, 800×600의 해상도를 갖는다. 카메라의 성능에 따라 위치인식 마커의 특징에 영향을 받지 않도록 하기 위해 일반적으로 많이 사용되는 웹카메라를 선정하였다. 웹카메라는 카메라 지그에 고정형으로 부착하지 않고 노트북에 연결하여 사람이 손에 들고 직접 걸어 다니면서 동작한다.

그림 9는 본 시험을 위해 사용된 인식 마커들을 나타낸다. 인식 마커는 인식하고자 하는 장소에 임의의 사물로 지정해서 장소 인식 마커를 만들 수 있다.

그림 10은 본 논문의 시스템을 시험하기 위하여 사용된 실내 장소를 나타낸다. 실내 장소는 복잡한 사물이 배치되어 있는 사무실 내부 (상단)와 계단 (중앙), 복도 (하단) 등의 장소로 구분하여 시험하였다.

3.2 시험 결과

그림 11은 장소 인식 마커를 이용한 장소 인식 결과 화면을 나타내고 있다. 각 장소 별로 장소 인식 마커를 이용하여 장소가 인식된 결과 화면을 나타내고 있다. 그림 11의 (a),(c),(e)와 같이 장소 인식 마커의 후보영역 검출은 라인으로 표시를 했으면 그림 11의 (b),(d),(f)는 인식된 장소 인식 마커에는 사각형으로 표시를 하였다.

그림 11의 (a)의 경우 장소 인식 마커와 유사한 물체가 있어도 인식이 잘되는 결과를 확인할 수 있었다. 복잡한 사물이 있는 사무실 내부 외에 복도, 계단 등에서 본 논문의 시스템을 적용한 결과 실시간 처리되고 입력되는 과정에서 원활하게 장소 인식 마커를 찾고 장소를 인식할 수 있음을 볼 수 있었다.

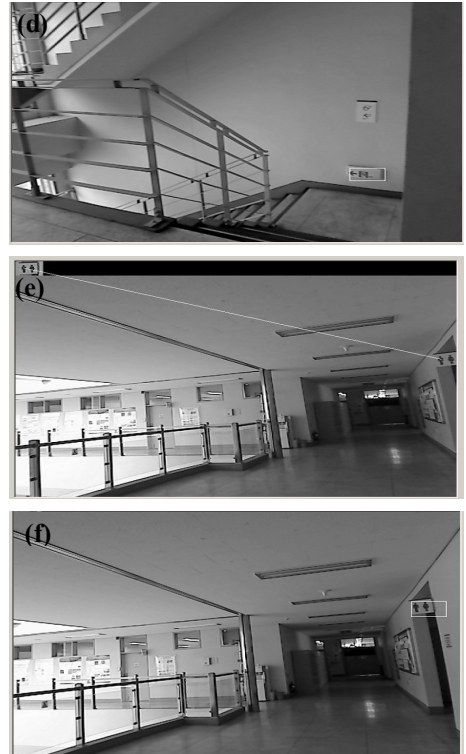
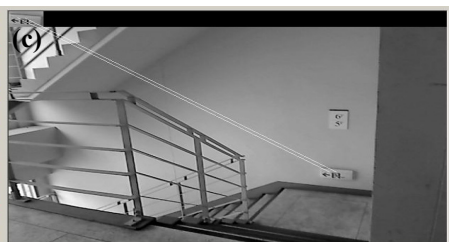


Fig. 11. Experimental images for place recognition using marker (a) recognizing region including one of candidate markers on the inside of office, (b) recognizing one of candidate markers on the inside of office, (c) recognizing region including one of candidate markers on the inside stair of building, (d) recognizing one of candidate markers on the inside stair of building, (e) recognizing region including one of candidate markers in indoor hallway, (f) recognizing one of candidate markers in indoor hallway.



본 논문에서 사용하는 SURF 알고리즘의 우수성을 보이기 위해 SIFT 알고리즘을 사용한 경우와 처리 시간을 비교하였다. 그림 12는 SIFT 알고리즘을 이용한 장소인식 마커인식 결과 화면이다.

표 1은 그림 10에 나타난 실내 장소별로 본 논문의 시스템에 대한 장소인식 마커인식을 및 1 프레임 당 영상인식 시간을 측정된 결과를 나타낸다. 인식 장소 별 장소인식 마커 인식율은 1,800 프레임 기준 90% 이상의 인식을 하는 것을 볼 수 있고, 영상 1 프레임 당 처리시간이 0.5초 이하로 처리가 되는 것을 확인할 수 있기 때문에 실시간 장소인식으로 동작 가능함을 알 수 있다. 장소인식 마커의 인식율은 각 장소 별 조명 환경 및 장소인식 마커의 위치 및 인식 마커가 포함되지 않은 영상으로 인식율 차이가 발생하였다.

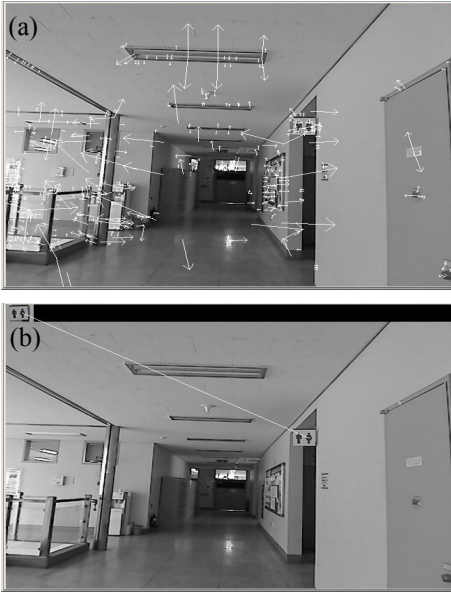


Fig. 12. Experimental image when using SIFT (a) showing detection of image feature, (b) showing detection of one of candidate markers.

Table 1. The comparison of recognition rate and time in each different places.

Places	Recognition rate of place recognize marker (1,800 frame)	Average recognition time (1 frame/sec)
place1 (Inside the office)	98%	0.5
place2 (Inside the office)	95%	0.3
place3 (Corridor)	92%	0.3
place4 (Stair)	90%	0.2
place5 (Corridor)	95%	0.3

표 2는 표 1을 위해 사용된 위치에서 서로 다른 1 frame의 전체 영상에서 SIFT와 SURF를 사용했을 때 특징 추출에 소요되는 시간을 나타낸다. 표에서 볼 수 있듯이 SURF 알고리즘에 비해 SIFT 알고리즘이 10배 정도의 시간을 소요함을 볼 수 있다.

표 3은 SURF를 이용할 때 각 장소 별 영상 특징을 추출하는 데 소요되는 시간을 나타낸다. 복잡한 물체들이 있는 사무실 내부에서 최대 처리 시간을 나타내고 있음을 볼 수 있다. 그러나 이 경우에도 1 frame 당 0.3초 미만의 시간이 소요되므로 실시간 시스템으로 사용하는 데 무리가 없음을 확인할 수 있다.

Table 2. The comparison of recognition time between SURF and SIFT in each different places.

Places	SIFT(msec)	SURF(msec)
place1	2261.5	241.4
place2	2533.7	193.7
place3	1937.7	191.2
place4	1907.6	173.5
place5	2010.3	172.6

Table 3. The comparison of max/min processing time for image feature extraction in each different places when using SURF.

Places	SURF	
	Max. process time (msec)	Min. process time (msec)
place1	299.9	203.5
place2	220.8	174.2
place3	221.4	169.4
place4	205.9	161.2
place5	165.1	186.8

4. 결론

본 논문은 SURF를 사용하여 영상 특징을 추출함으로써 실시간으로 장소를 인식할 수 있는 시스템을 구현하였다. 본 논문의 시스템은 특정인식 마커를 사람이 고정하거나 만드는 것이 아니라 입력되는 영상에서 특정 사물에 대해 사전에 특징을 추출한 후 데이터베이스화 하는 방법을 사용하기 때문에 유연성을 제공하는 장점이 있다. 시험 결과에서와 같이 실시간 입력되는 1,800개 영상에서 장소인식 마커 인식율은 90% 이상 되었고 SIFT 알고리즘과의 영상 특징점 추출처리 시간 비교에서 10배 정도 빠른 처리속도로 처리되는 것을 볼 수 있었다. 그리고 사물이 복잡하게 위치해 있는 사무실 내부의 최대 처리시간이 300 msec이므로 실시간으로 처리하는 시스템에 적용 가능할 것으로 예상된다. 시험에 사용된 카메라는 일반적으로 사용되는 저가의 웹카메라를 사용하였음에도 무리없이 동작하는 것을 확인하였다.

그러나 본 논문에서 사용된 영상 특징 추출 기법인 SURF는 흑백 영상으로부터 스케일, 회전 변환에 불변하는 특징점을 찾고 표현자를 구성하지만, 밝기 변화 및 부분 폐색에 약하다는 단점을 가지고 있다. 또한 흑백 공간의 정보만 이용하기 때문에 특징을 구분하는데 컬러 공간의 유용한 정보를 제대로 활용하지

못한다는 단점도 있다. 이러한 단점들을 개선하여 밝기 변화 및 부분 폐색 영역이 많은 실내외 및 움직임이 많은 기기에 적용할 수 있는 개선된 실시간 장소인식 시스템에 대한 연구는 향후 과제의 하나이다.

감사의 글

이 논문은 부산대학교 자유과제 학술연구비(2년)지원에 의하여 연구되었습니다.

REFERENCES

- [1] V. Kulyukin, C. Gharpure, J. Nicholson, and S. Pavithran, *Int. Conf. Intell. Robot. Syst.*, 1979 (2004).
- [2] Y. Hu, B. J. Shin, and C. W. Lee, *KSCI*, **13**, 52 (2008).
- [3] B. Nissanka, Anit Chakraborty, Hari Balakrishnan, *Proc. ACM Mobicom*, 32 (2000).
- [4] A. Harter and A. Hopper, *IEEE Pers. Commun.*, **4**, 42 (1997).
- [5] R. Want, V. Falcao, and J. Gibbon, *ACM Trans. Inf. Syst.*, **10**, 91 (1992).
- [6] P. Bahl and V. Padmanabhan, *Proc. IEEE Infocom.*, **2**, 26 (2000).
- [7] I. Ulrich and I. Nourbakhsh, *IEEE International Conference on Robotics and Automation*, **2**, 1023 (2000).
- [8] A. Briggs, D. Scharstein, and S. Abbott, *Algorithmic and Computational Robotics: New Directions, in Fourth International Workshop on Algorithmic Foundations of Robotics, WAFR*, 349 (2000).
- [9] G. Dudek and D. Jugessur, *IEEE International Conference on Robotics and Automation* (San Francisco, USA, 2000) p. 1030.
- [10] A. Torraba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, *IEEE Int. Conf. Comput. Vis.*, **1**, 273 (2003).
- [11] D. Jung and H. Kim, *IEA/AIE '2005 Proceedings of the 18th International Conference on Innovations in Applied Artificial Intelligence*, 36 (2005).
- [12] F. Li and J. Kosecka, *IEEE Int. Robotics and Automation*, 3405 (2006).
- [13] H. Bay, T. Tuytelaars, and L. V. Gool, *European Conference on Computer Vision*, **3951**, 404 (2006).
- [14] D. Lowe, *Int. J. Computer Vision*, **60**, 91 (2004).
- [15] L. Ledwich and S. Williams, *In Australian Conference on Robotics and Automation, ACRA* (2004).
- [16] D. D. Lewis, *In Proceedings of ECML-98, 10th European Conference on Machine Learning* (Germany, 1998) p. 4.