

음향-영상-위치 통합정보의 앙상블학습을 이용한 장소인식

Place Recognition using Ensemble Learning of Audio-Vision-Location Integrated Information

이충연, 이범진, 온경운, 하정우, 강우성, 장병탁

한국컴퓨터종합학술대회 (KCC 2014)

2014. 06. 13

서울대학교 컴퓨터공학부 바이오지능연구실

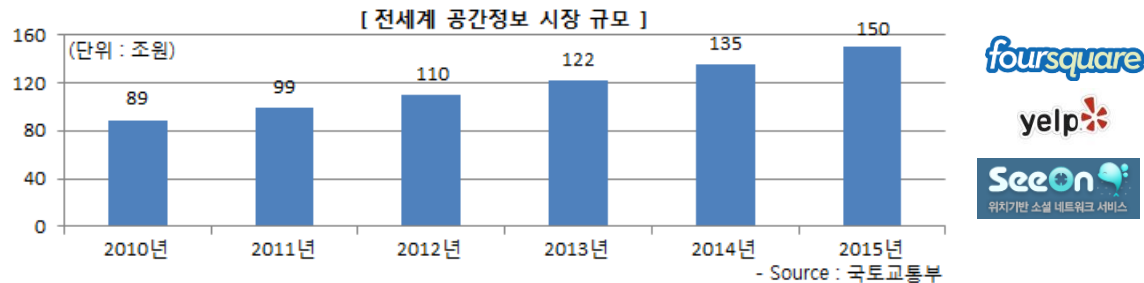
목차

- 연구 목적 및 필요성
- 연구 방법
- 실험 결과 및 논의

연구 목적 및 필요성

■ 위치기반서비스 (Location-Based Services, LBS)

- 2010년 이후 스마트폰 시장과 맞물려 급속 성장
- 위치정보와 커머스, 엔터테인먼트가 결합한 하나의 플랫폼으로 진화할 것으로 전망
- SNS, 증강현실, 모바일 게임, 광고 등의 서비스와 연계
- 위치정보를 활용한 긴급구조 등 사회 안전망으로서의 활용성 증대



 iBeacon, Spotsetter, Hopstop, Embark, BroadMap

 Google Now, Skybox Imaging, Titan Aerospace

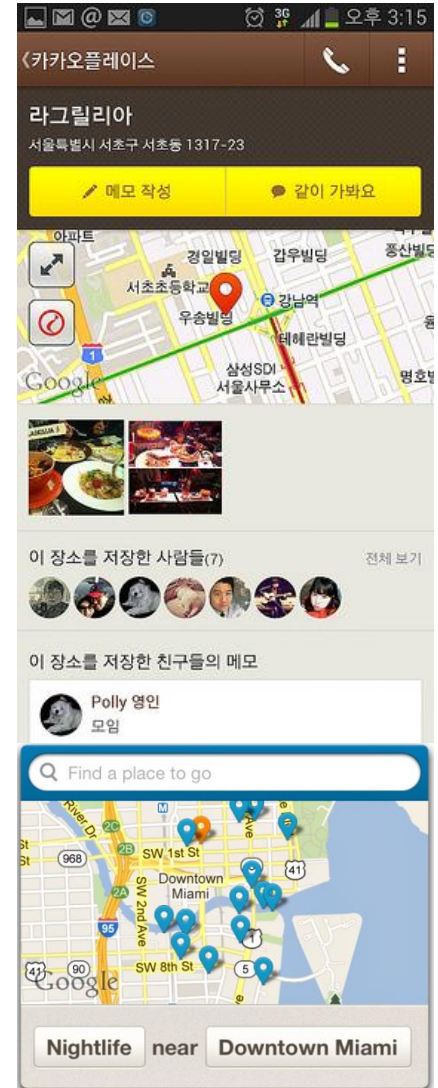
 Autonavi

 Nearby Friends

 Navinfo, Dianping

 Shopcloud

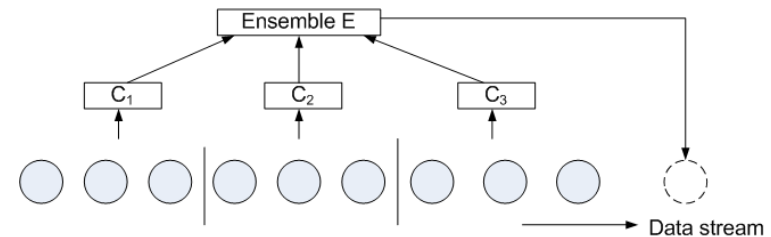
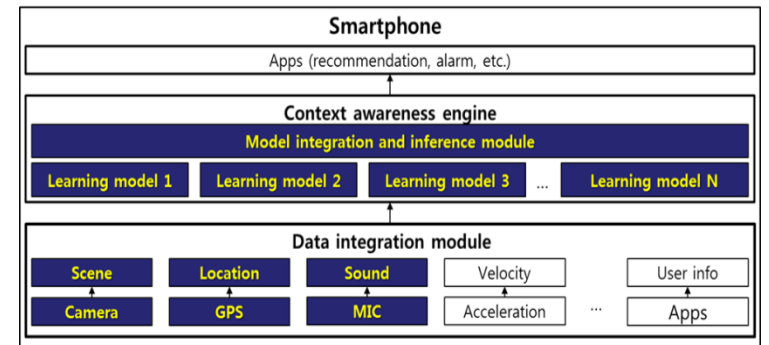
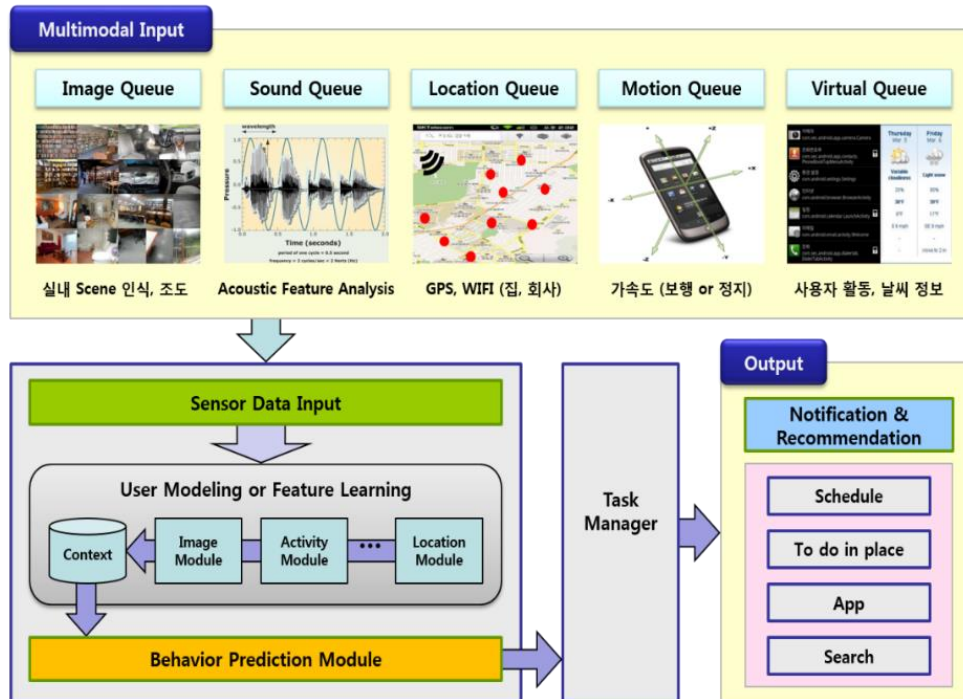
- 위치측위 오차가 크고, 실내 및 GPS 신호 음영지역에서 측위 불가능
- 실제 서비스에서 이용하기 위해 해당 위치의 논리적 정보의 개별적 맵핑 필요



연구 목적 및 필요성

■ 연구 목적

- 좌표 정보 외에 해당 위치에서의 시각 및 청각 신호 등 멀티모달(Multi-modal)의 환경 정보들을 이용하는 사람의 위치 인식 메커니즘을 모방한 음향-영상-위치 통합 정보 기반의 장소 정보 인식 방법
- 지도상의 단순 좌표 인식이 아닌 논리적 위치 정보, 즉 장소(Place) 인식을 수행
- 이종(Heterogeneous) 데이터들을 함께 이용하기 위해 앙상블 학습 방법 이용

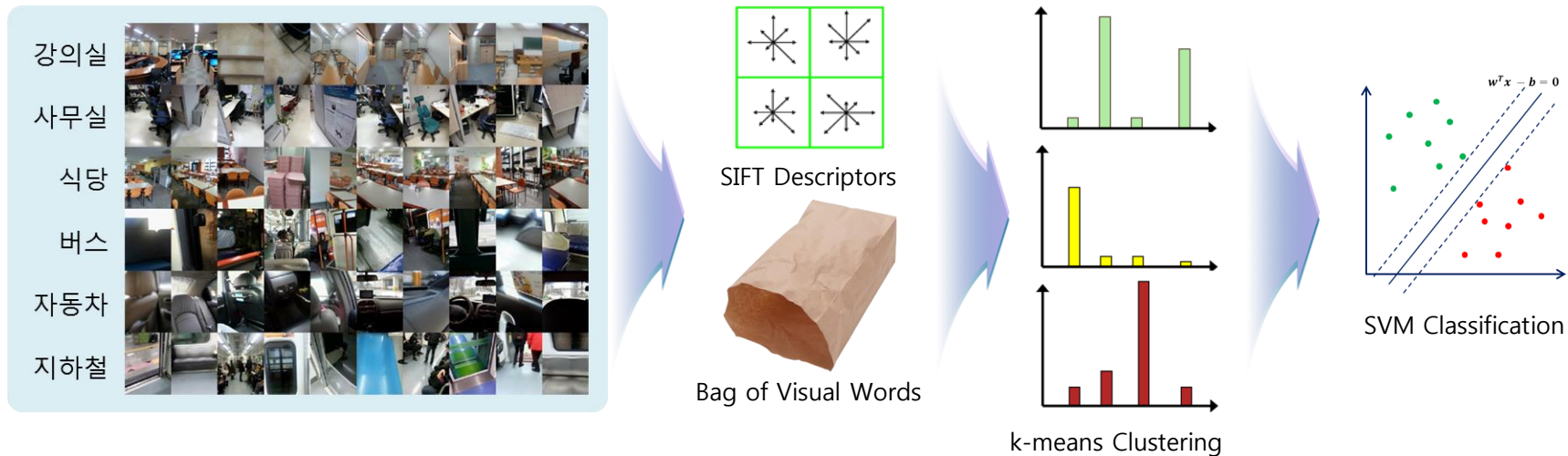


연구 방법 (1/4)

■ 영상 데이터 전처리 및 특징 추출

- 6개 장소에서 3,000개 이미지 획득 (480×640 pixels, 72dpi)
- SIFT 기술자 추출 및 BoW 히스토그램 생성
- k-means 클러스터링 (k=200)
- SVM Classification (RBF kernel, cost=1, $\gamma=0.01$, 10-fold cross validation)

→ 영상 데이터 기반 장소인식 정확도: 64.17% (80% 학습, 20% 검정)

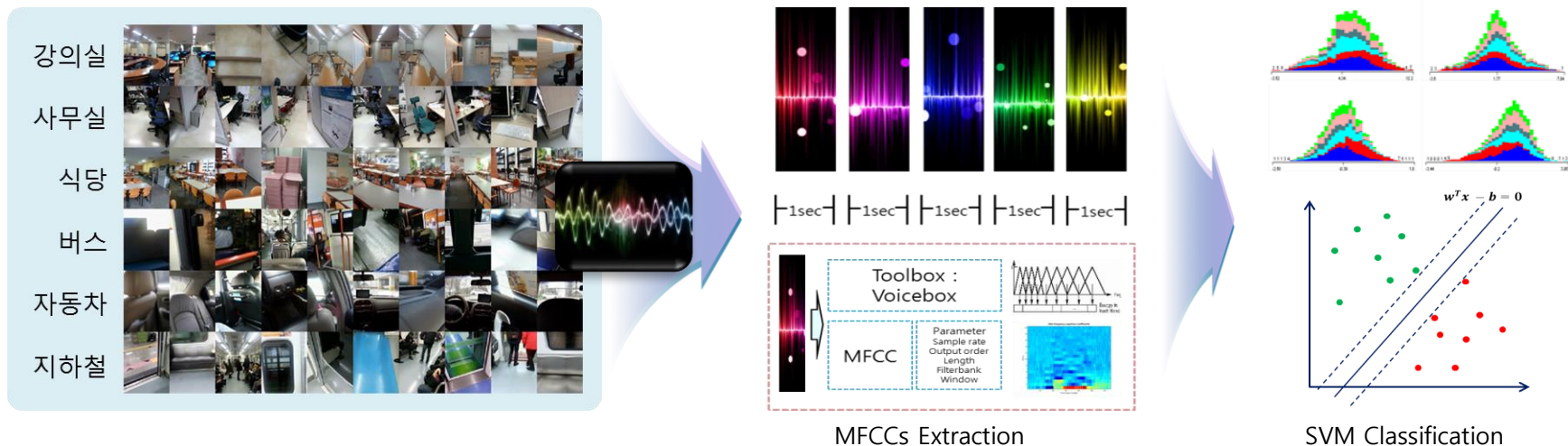


연구 방법 (2/4)

■ 음향 데이터 전처리 및 특징 추출

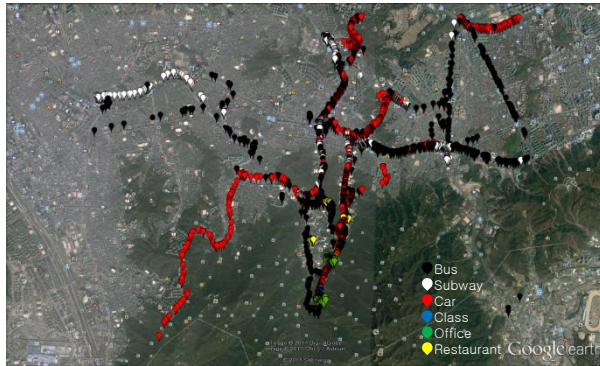
- 6개 장소에서 3,000개 데이터 획득 (44100 Hz 주파수, 1초 길이)
- MFCC 계수 추출 (13개 계수)
 - 2048개 샘플로 구성된 프레임을 1024개 간격으로 이동시키면서 추출 (50% 중첩, 총 42개)
 - Hamming windowing (불연속 지점 오류 최소화), FFT (주파수 영역 변환)
 - Mel-scale filterbank (파워스펙트럼 추출), DCT (중첩된 필터뱅크 출력데이터 간의 상관관계 제거)
- SVM Classification

→ 음향 데이터 기반 장소인식 정확도: 81.83% (80% 학습, 20% 검정)

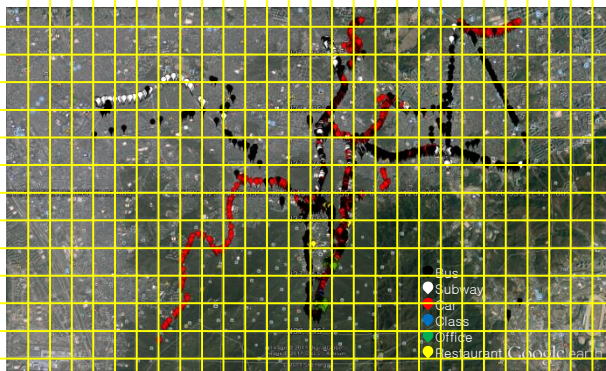


연구 방법 (3/4)

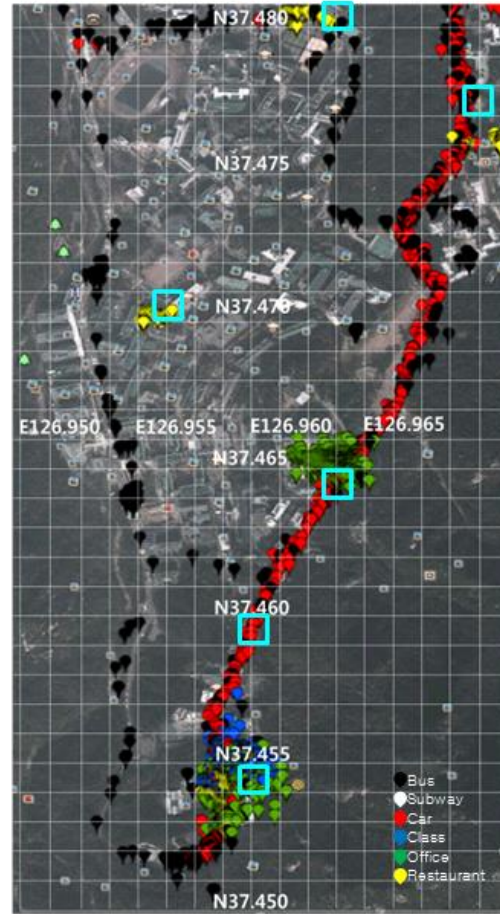
■ 위치 데이터 전처리 및 특징 추출



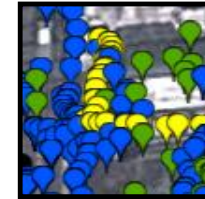
Test GPS data



Grid-Partitioning (Degree Unit: 0.001)



Cropped data for major region



Place	Probability
Bus	0.00 %
Subway	0.00 %
Car	0.00 %
Class	61.18 %
Office	16.89 %
Restaurant	21.93 %



Place	Probability
Bus	22.22 %
Subway	0.00 %
Car	61.11 %
Class	0.00 %
Office	16.67 %
Restaurant	0.00 %



Place	Probability
Bus	7.70 %
Subway	46.15 %
Car	46.15 %
Class	0.00 %
Office	0.00 %
Restaurant	0.00 %



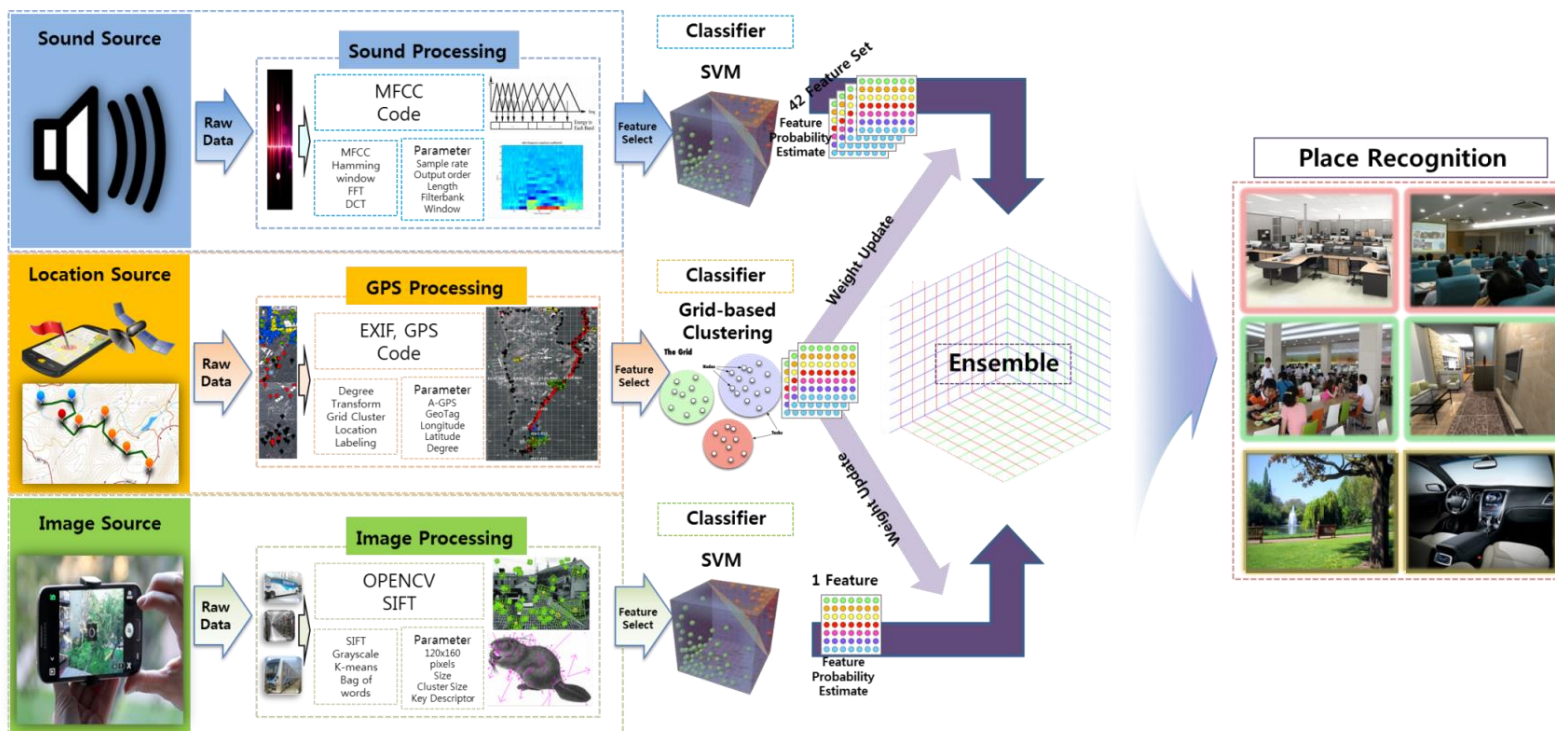
Place	Probability
Bus	80.00 %
Subway	0.00 %
Car	20.00 %
Class	0.00 %
Office	0.00 %
Restaurant	0.00 %

Probability for place in each grid cluster

연구 방법 (4/4)

■ 음향-영상-위치 통합정보의 앙상블학습을 이용한 장소인식

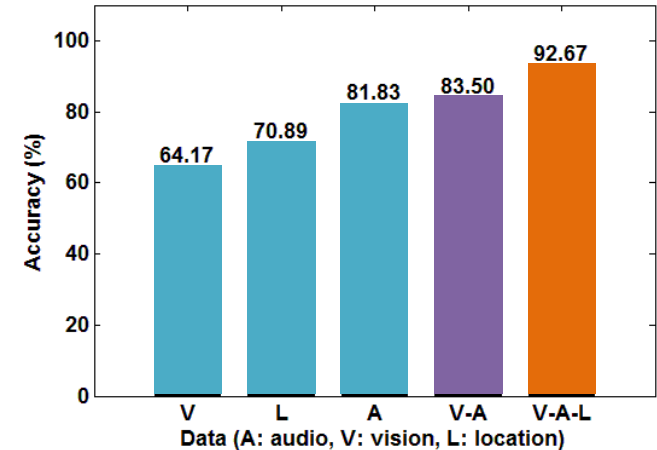
1. 음향과 영상 데이터의 특징 벡터들로부터 각각 SVM 분류기를 이용하여 장소 분류
2. 음향 데이터로부터 42개의 서로 다른 분류 결과들 중 가장 높은 빈도의 장소를 선택하고, 이 장소에 대한 분류 결과들의 추정확률(probability estimates, [T. F. Wu, *JMLR* 2004]) 평균을 계산
3. 위치 데이터의 영역별 장소 확률을 음향과 영상 데이터 분류 결과에 대한 가중치로 사용
4. 음향과 영상 데이터의 장소 분류 결과 중 추정확률과 위치 데이터 가중치의 곱이 큰 것을 선택



실험 결과 및 논의 (1/2)

■ 실험 결과 및 논의

- 학생 및 연구원 6명이 스마트폰을 통해 데이터 획득
- 1 Sample = BoW Histogram, 42 MFCCs, 3 GPS Coordinates
 - 사진 3000장, 오디오 3000초, **GPS 좌표** 9000개
 - **실내**: 사무실, 강의실, 식당
 - **실외**: 버스, 자가용, 지하철



- 실험결과: **92.67%** (556/600 of test data)

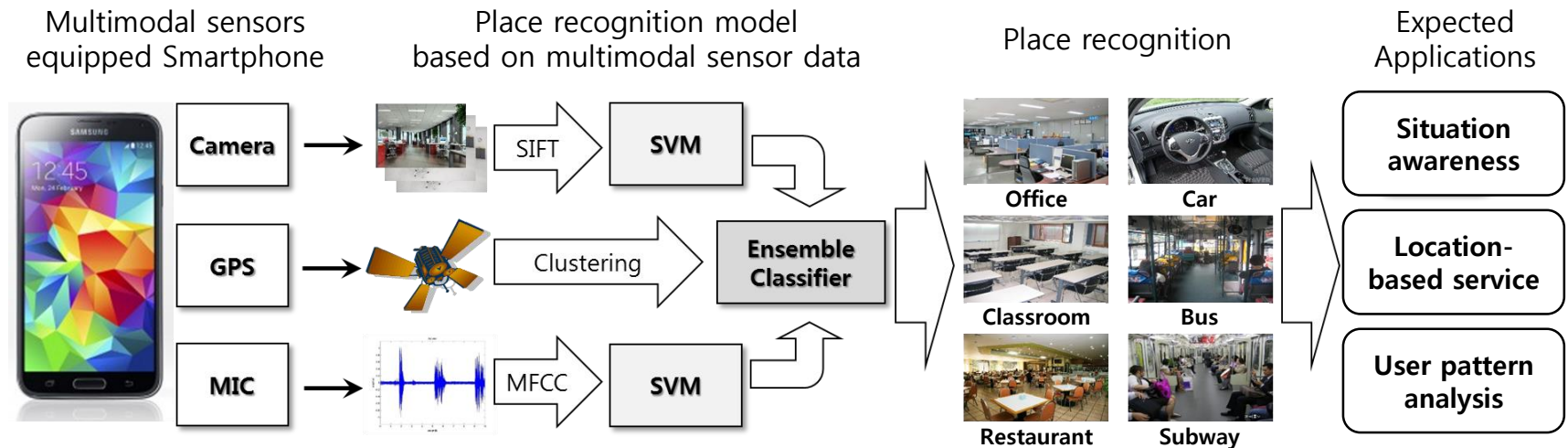
단일 데이터 분류 < 영상&음향 데이터 분류 < 영상&음향&위치 데이터 분류

- 제안한 방법은 **단일 데이터 분류시 문제가 되는 요인들을 상보적으로 완화시켜** 인식률이 향상된 것으로 보임
 - 영상 데이터: 조도 변화나 흔들림 등 외부 요인에 의한 잡음 포함 (image variability)
 - 음성 데이터: 서로 다른 장소의 데이터에 유사한 음향 신호(e.g. 사람 목소리)가 포함 (perceptual aliasing)
 - 위치 데이터: 동일 건물 내에 위치한 사무실, 강의실, 식당의 분류가 어려움 (signal unavailability)

실험 결과 및 논의 (2/2)

■ 향후 연구 방안 및 활용

- 영상 기반 장소인식 확률을 높이기 위해 Pyramidal Matching 사용
- 지도(map) 정보 서비스, SNS, 사용자 행동 정보와 연계하여 장소인식 성능 고도화
- 사용자 상황인식 서비스 및 실내 · 외 위치기반 추천 서비스 등에서 활용



감사합니다!