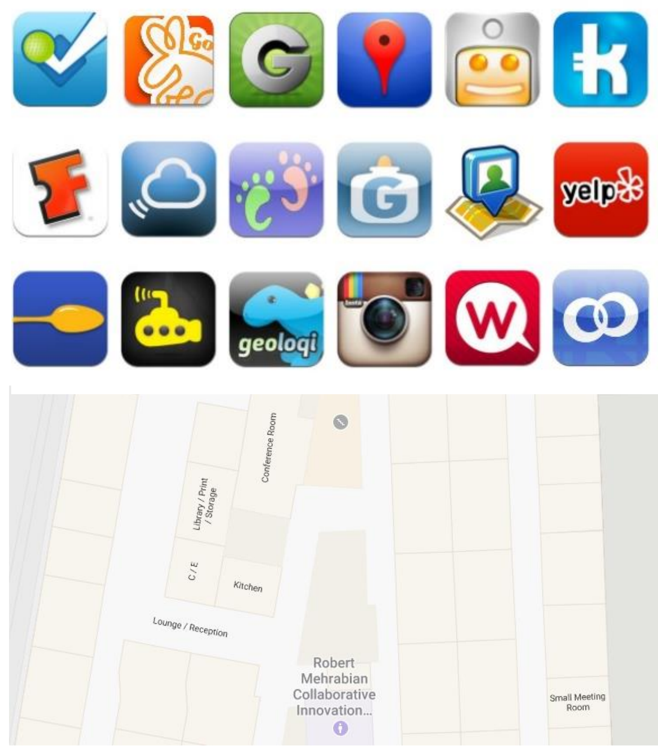
ShareLock: A Crowdsourced System for Automated Semantic Tagging of Indoor Floorplans

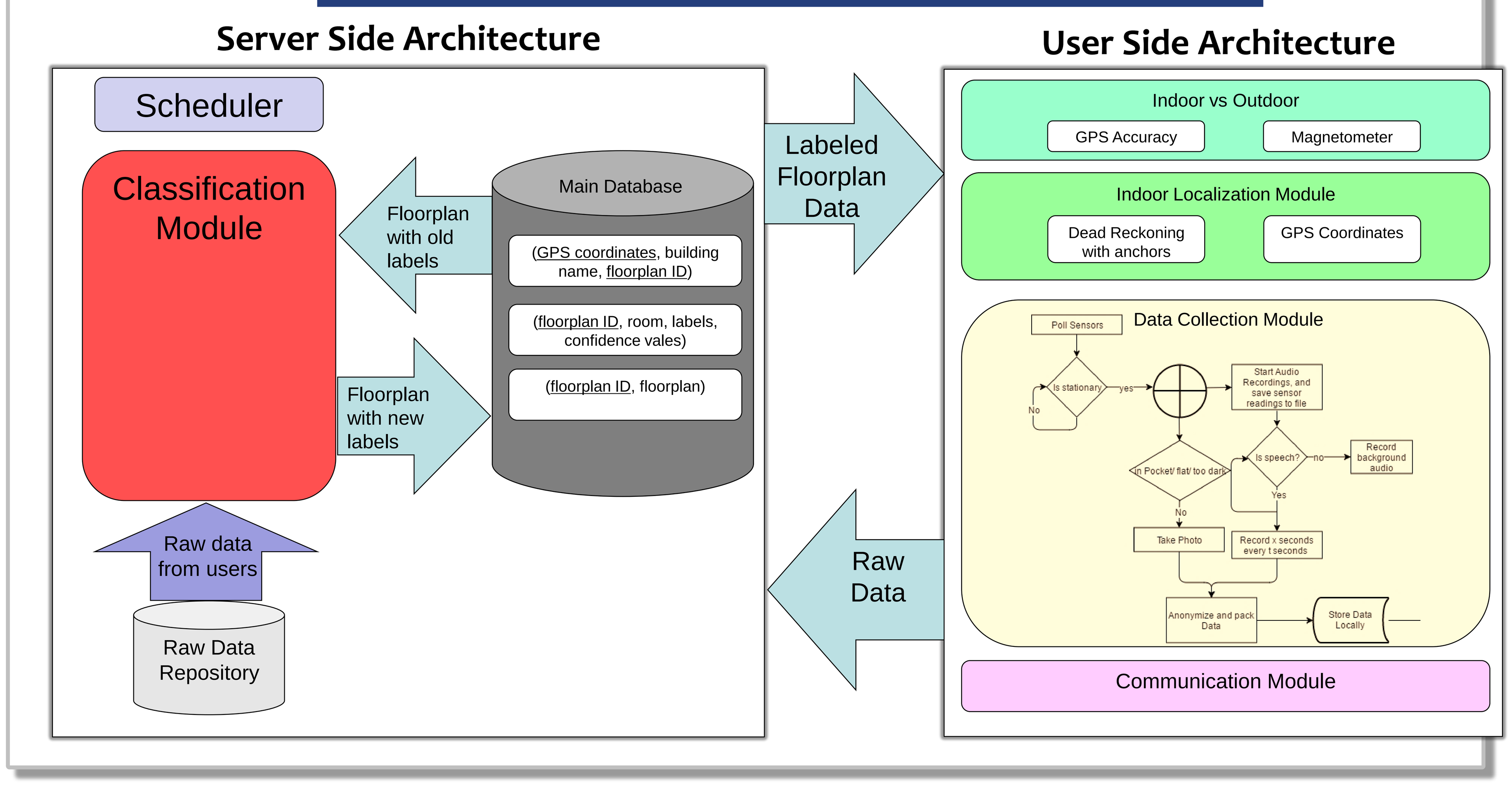
Muhammad A. Shah+ , Khaled A. Harras+, Bhiksha Raj+
+Carnegie Mellon University

I. MOTIVATION

- Growth in the popularity of location-aware services.
- Lack of a scalable indoor mapping system is holding back the introduction of location-aware services in indoor environments.
- Automated floorplan generation systems do not provide room labels
- **Challenges**: Determining relevant data sources, learning room features from data, scalable means of collecting data, localization.
- **Solution:** A crowdsourced system for automated semantic tagging of indoor floorplans.



II. SHARELOCK ARCHITECTURE



III. PROBLEM & PROPOSED SOLUTION

Problem

- The current work generates floorplans devoid of labels. [1]
- Limited usefulness for user-centric applications, eg. Navigation, etc.
- Currently manual entry is required to make labelled indoor maps available.

Solution:

- Crowdsourced data collection with minimal user interaction
- Continuous Acoustic Monitoring.
- Images Classification.

Crowdsourced Data Collection

- Runs in the background and collects data when it is suitable to do so.
- Preset triggers (see II) determine the suitability of data collection and the frequency at which it is obtained.
- Data is opportunistically uploaded to the server based upon the internet connectivity of the client device

Continuous Acoustic Monitoring

- Confidence in a label is determined as an increasing amount of data is made available to the classifier.
- Classification approaches used:
  - Gaussian Mixture Models (GMM) [2]
  - GMM Supervectors with Support Vector Machines [3]

Image Classification

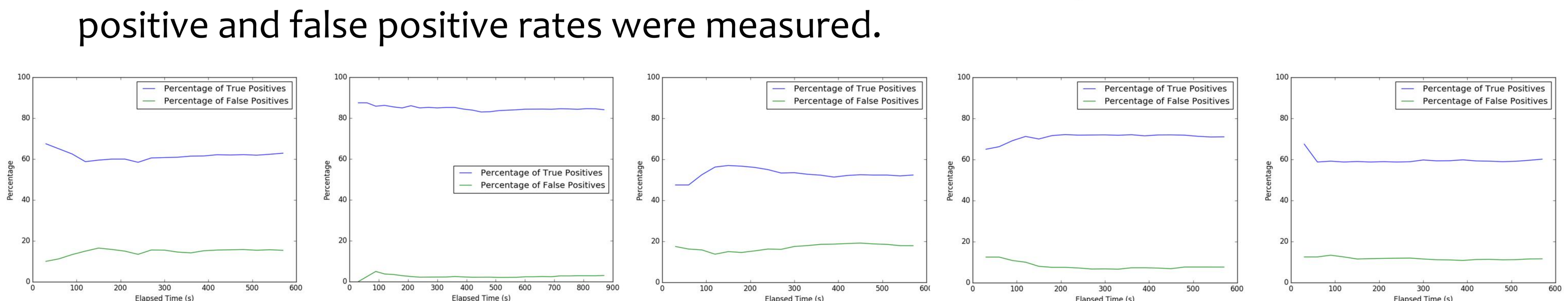
- Image classification is done using Google's vision API.[4]

III. DATASET

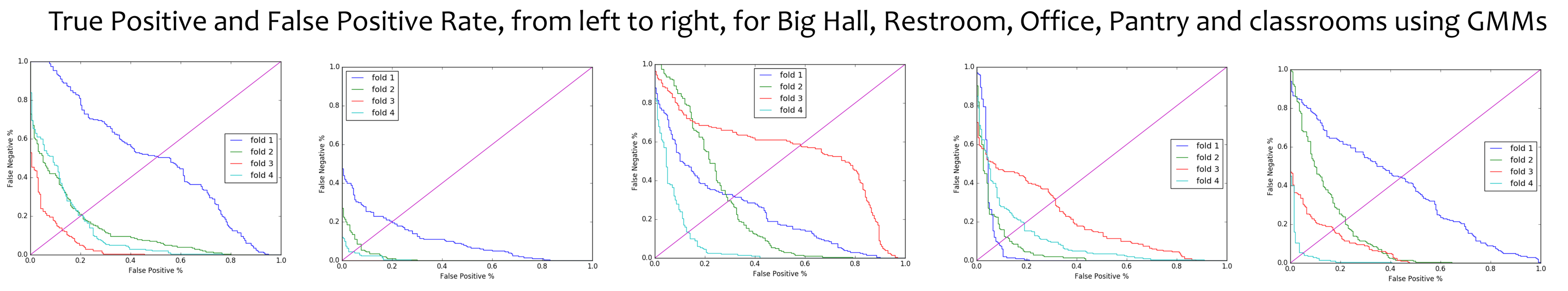
- **22,200 seconds** of single-speaker speech recordings (7400 recordings, 3 seconds each)
- The speakers were **2 male undergraduate students**
- Recordings were obtained using **2 smartphones**, one in hand at chest height and other in pocket, simultaneously.
- The recordings were obtained in **5 offices, 5 restrooms, 4 pantries, 4 lecture halls and 4 classrooms.**
- The speakers repeated **4 sentences, each 5 times at 5 different locations** within each room.
- The recordings were sampled at **44.1 kHz** and saved as uncompressed .wav files.
- The recordings were done when rooms were **unoccupied**.
- **20-D MFCCs** were extracted using 64ms frames with 16ms overlap.
- The dataset was divided into **4 folds for cross validation.**

IV. EXPERIMENTS WITH GMMS

- We performed **binary classification** using GMMs:
  - 2 GMMs per room type, 1 trained on positive instances of the room the other on the negative
- Each GMM had **64 mixture** components.
- The test set was divided into 30 second chunks (10 recordings/chunk) and the true positive and false positive rates were measured.



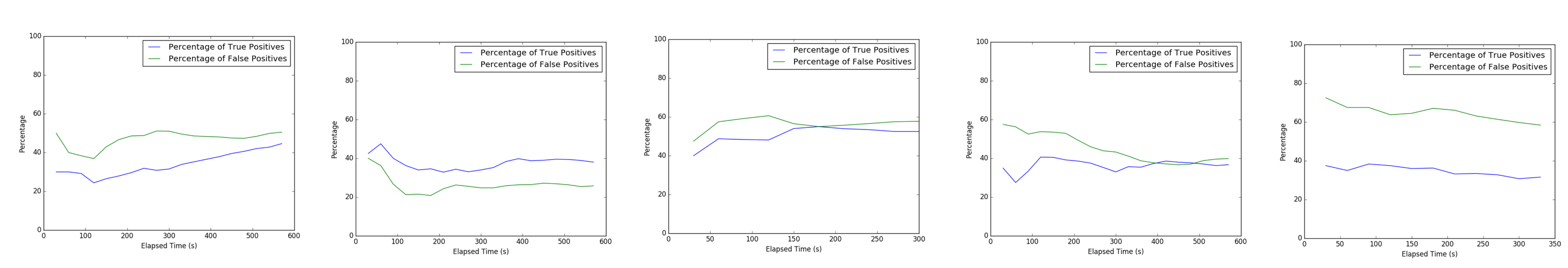
True Positive and False Positive Rate, from left to right, for Big Hall, Restroom, Office, Pantry and classrooms using GMMs



DET Curves, from left to right, for Big Hall, Restroom, Office, Pantry and classrooms using GMMs

V. EXPERIMENTS WITH GMMS

- Trained a Universal Background Model GMM with 64 components
- Used MAP-Adaptation to extract supervectors from recordings
- Trained a binary SVM for each room type using liblinear
- Setup the SVM to output a score instead of +1/-1 results
- The test set was divided into 30 second chunks (10 recordings/chunk) and the true positive and false positive rates were measured.



True Positive and False Positive Rate, from left to right, for Big Hall, Restroom, Office, Pantry and classrooms

VI. EXPERIMENTS WITH IMAGE CLASSIFICATION

- Added noise and motion blur, and reduced the brightness of the image of auditorium to test how degrading quality impacts the results of Google Vision API [4].

Confidence in the label "Auditorium" for the given image. "n/a" refers to the case when the confidence in the "Auditorium" label was less than 50%.

VI. CONCLUSION & FUTURE WORK

Conclusions

- Acoustic monitoring is a promising means to identify room type.
- GMM supervectors with SVMs are not very effective
- Google Vision API is fairly robust to blur and noise so is a good candidate as a supporting modality.

Future and Ongoing Work

- Exploring other learning and classification techniques (currently exploring i-vectors and non-negative matrix factorization
- Improve the prototype android application to deploy to test users.
- Collect data from user trials

References

1. Alzantot, Moustafa, and Moustafa Youssef. "Crowdinside: automatic construction of indoor floorplans." Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, 2012.
2. 10. Rajapakse, Menaka, and Lonce Wyse. "Generic audio classification using a hybrid model based on GMMs and HMMs." 11th International Multimedia Modelling Conference. IEEE, 2005.
3. 15. Campbell, William M., Douglas E. Sturim, and Douglas A. Reynolds. "Support vector machines using GMM supervectors for speaker verification." IEEE signal processing letters 13.5 (2006): 308-311.
4. https://cloud.google.com/vision/