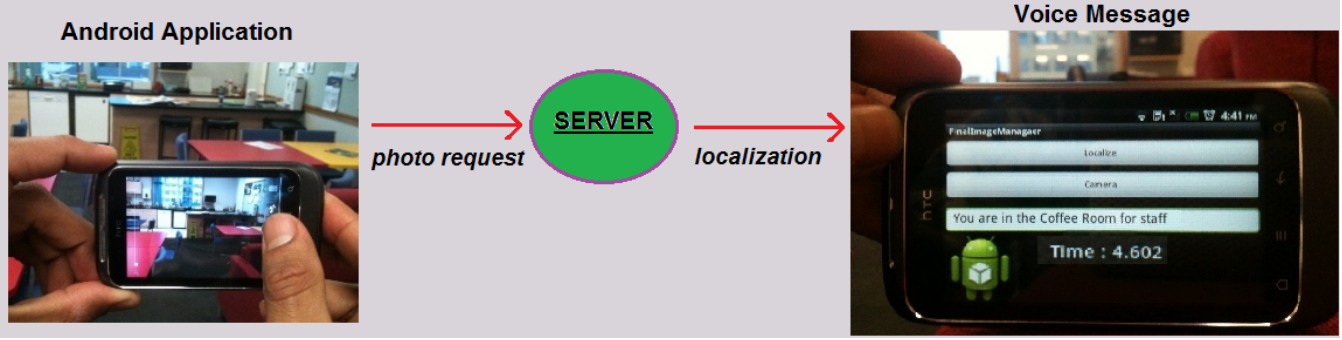
# Vision based Indoor Scene Localization via Smart phone

Nabeel khan
Computer Science Department
University of Otago
nabeel@cs.otago.ac.nz

Brendan McCane
Computer Science Department
University of Otago
mccane@cs.otago.ac.nz

Steven Mills
Computer Science Department
University of Otago
steven@cs.otago.ac.nz




## ABSTRACT
Scene localization without GPS in indoor environments is a challenging problem. Robust indoor scene localization is particularly useful for visually impaired people and robots during navigation. We present a prototype Android application which is intended to perform indoor localization and convey location information to a blind person. The application is in its early stages and we are looking to extend its functionality in the future.

## Categories and Subject Descriptors
I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Shape*; H.3.3 [**Information Systems**]: Information search and retrieval—*Retrieval methods*

## General Terms
Verification, Performance.

## Keywords
Visual Bag of Words, SIFT features, Speech synthesis

## 1. INTRODUCTION
Precise localization is needed for navigation systems because wrong location identification may lead to a navigation failure. Vision based localization in an office type indoor environment remains a challenging problem due to the visually similar nature of many locations within a typical office building. The proposed work aims to aid visually impaired people while they are indoors. The proposed smart phone application can be used by a blind person to take a photo of the current scene and then determine their location information whenever they want.

## 2. SYSTEM
The proposed system works in a client server paradigm. The client side refers to an Android application currently running on an *HTC Wildfire S* (2.3 Gingerbread). The framework of our system consists of two modules:-

1. **Offline:-** A server contains a database image of the building sorted by location. The Visual Bag of Words (BoW) is used for matching on the server. SIFT features are extracted from trained images followed by generation of visual words for every trained image. The server develops an inverted index and then waits for the client request.

2. **Online:-** The smart phone application sends the taken photo to the server and waits for an answer. When the server gets the request, it uses the index to retrieve top 200 similar images and then ranks the top images by BoW. If the top three images refer to the same place then the server simply returns the corresponding location. Otherwise a fundamental matrix based verification is performed for the reliable match. The smart phone then uses a voice message to indicate the current location.

## 3. EXPERIMENTS
The system is tested on indoor images taken from the Computer Science building at Otago University. It is a standard office building with many self-repetitive confusing patterns. The trained images were taken in the evening while query photos are captured during the night and the morning. A good localization rate of 86% is observed in reasonable time from the server.