Emerging Techniques in Vision-based Indoor Localization

PhD Student: Feng HU
Advisor: Prof. Zhigang ZHU
Committee Members: Prof. Jizhong XIAO
Prof. Jianting ZHANG
Outline

• Introduction
• Indoor Localization: An Overview
• 2D Image-based Approaches
• 3D Model-based Approaches
• Challenging Issues and Emerging Solutions
• Conclusion
Outline

• Introduction
• Indoor Localization: An Overview
• 2D Image-based Approaches
• 3D Model-based Approaches
• Challenging Issues and Emerging Solutions
• Conclusion
Indoor Localization

• Definition
  – Use one or more sensors, such as cameras, magnetic sensors, IMUs, RGBD sensors, etc., to automatically determine the location of a robot or a person in real-time in an indoor environment.

• Special focuses
  – Visual-sensor-based localization
    • E.g. camera, RGBD sensor, especially omnidirectional cameras.
  – Assistive applications
    • Assisting Visually Impaired People (VIP) with performing indoor navigation tasks.
Motivation

• VIPs’ memory burden in indoor navigation
  – 285 million VIP (4% of world population); 39 million blind. Updated Aug. 2014 from WHO.
  – How VIP navigate in indoor environments? [Paisios 2012]
    • By memorization. Lots of information to memorize: turn-by-turn, points of interests (i.e. landmarks), office number, etc.
  – Computer vision techniques, probably accelerated by GPUs, on daily-used mobile/wearable devices and on servers on the cloud, have the potential to achieve accurate, real-time, and robust indoor localization.
Working Modes

• Two modes a vision-based indoor localization system can be used

  – VIP call the application of a smart-phone or wearable device and find out the current location;

  – System automatically sends notification to the users if they reach a specific location, even though they have not realized it.
Outline

• Introduction
• Indoor Localization: An Overview
• 2D Image-based Approaches
• 3D Model-based Approaches
• Challenging Issues and Emerging Solutions
• Conclusion
Why not GPS?

- Good performance in the outdoor
- Limited in indoor environments
  - Signal attenuation by construction materials, such as, roofs, walls, etc.
- Indoor assistive localization requires higher accuracy (usually, <1m)
Other Non-visual Solutions?

• **Bluetooth based localization [SPREO 2014]**
  – Bluetooth beacons + mobile receiver(s).
  – Trilateration.
  – Determine a receiver’s coordinate, given at least 3 beacons’ coordinates.

• **Radio Frequency Identification (RFID) based localization**
  – Use electromagnetic field, to wirelessly transfer data, for automatically identifying and tracking RFID tags attached to objects.
  – E.g. [Chumkamon 2008] embedded tags into stone blocks, and put them onto a footpath. The blind users, carrying a portable RFID reader, can receive the signal and decode the location and other encoded information.
  – **Near Field Communication (NFC)** is—a technology evolved from RFID—enables smart-phones or other devices to establish radio communication (13.56M Hz) with each other by touching them together or bringing them into proximity, typically at a distance of 10cm or less with the maximum speed of 4.24Kb/s.

• **Disadvantages**
  – Extra modifications to the environments;
  – Cost proportionally increases when localization area expands;
  – Arise aesthetic and legal issues
Other Non-visual Solutions? (cont’d)

• **Wi-Fi based localization**
  – Store the localization area’s routers’ SSIDs, whose geo-locations are pre-calculated, into a server offline;
  – A user’s device, e.g. a smart-phone, sends all detected SSIDs into the server;
  – The server looks up into the SSID database, finds the associated routers’ locations, and triangulates the device’s location.
  – The server sends the location back to the device.

• **Disadvantages**
  – At least three routers are required for a query, which is not always satisfiable;
  – Need to maintain the SSID database dynamically for adding or deleting a router.
  – No other contextual information, besides a coordinate, is provided, which is not sufficient in many assistive applications.
Visual Sensor Based localization

• Why visual sensors?

  – Use visual information to perceive the environment is more natural for human beings and includes more rich information than just coordinate, names, etc.;

  – Mobile device cameras and consumer level 3D sensors are becoming cheap and widely used;

  – Especially, an omnidirectional camera system can provide with a 360*180 degree Field Of View (FOV), far larger than a normal camera’s 50*40 degree, which can save many troubles where narrow FOV would cause serious problems.
Omnidirectional vision-based localization

• Three modules in a typical system
  – Omnidirectional imaging
    • Process of using one or multiple omni-systems to move along the indoor area where we want to provide service, capture and store visual information.
  – Environment mapping or model building
    • Process of constructing one-to-one or many-to-one relationships between the 3D points in the physical space and the points in the digital space.
  – Feature extraction and searching
    • Process of extracting representations of local images or 3D points used to construct the digital space model and finding the correspondence in the pre-built digital space.
Outline

• Introduction
• Indoor Localization: An Overview
• 2D Image-based Approaches
• 3D Model-based Approaches
• Challenging Issues and Emerging Solutions
• Conclusion
2D Image-based Approaches

• Definition
  – To use original 2D images, transformations of the images, or their 2D features, to localize an input image within a pre-built area database.

• No 3D models are built or maintained; 3D location is obtained via geolocation-tagged images/features database

• Three majors issues discussed:
  – feature extraction
  – modeling via indexing
  – localization through retrieving
Three Major Issues

• Feature extraction
• Modeling via indexing
• Localization through retrieving
Feature Extraction

• Direct feature methods
  – Global features
    • Global feature summarize the global information of an image, usually, in the form of a single vector.
    • [Chapoulie 2013] does Spherical Fourier Transform to omnidirectional image, and uses spherical harmonic parameters to represent an omnidirectional image.
Global Feature (cont’d)

– [Oliva 2001] proposes a **holistic image representation** by modeling the image shapes.

• By classifying image holistic representations, with five **criteria** - naturalness, openness, roughness, expansion, and ruggedness, we can classify the scenes.
Feature Extraction

• Direct feature methods
  – Local features
    • Popular features, e.g., Scale Invariant Feature Transform (SIFT) [Lowe 2004] and Speeded Up Robust Features (SURF) [Bay 2006].
    • Variants in applying detected features to robustly represent images.
      – Combining spatially related features together [Johns 2014]
Variants (cont’d)

– [Li 2010] assigns different weights to different SIFT features, and relies the feature matching algorithm’s robustness on the high weighted feature points.
Variants (cont’d)

- [Shen 2013] groups SIFT features into physical planes and matches the images according to the number of planes matched instead of number of total SIFT features matched.
- Due to large non-linear distortion, SIFT matching algorithm on omnidirectional images only generates a few matches. [Carufel 2011] converts omnidirectional images matching problem into normal matching problem by cut omnidirectional images into several sub-images (tangent planes), and increases the matched feature’s number.
Feature Extraction

• Why pooling methods?
  – Acquire features with better invariance to image transformation
  – More compact image representation
  – More robust effect to noise and clutters

• How?
  – Create a feature book (a M*K matrix) by clustering the features
  – Convert the matrix into a K dimensional vector: max pooling and average pooling

\[
X = [x_1, x_2, ..., x_M]^T \in \mathbb{R}^{M \times D}
\]

\[
V = [v_1, v_2, ..., v_K]^T \in \mathbb{R}^{K \times D}
\]

\[
\alpha_{i,j} = \begin{cases} 
1, & \text{if } j = \text{arg} \min_k \|x_i - v_k\|_2 \\
0, & \text{otherwise}
\end{cases}
\]
Three Major Issues

• Feature extraction
• Modeling via indexing
• Localization through retrieving
Modeling via Indexing

• Identifies useful images or features and relates them to the physical locations in the space.
• Once and offline.
• Modeling difference for metric localization and topological localization.
Metric Localization Maps

• Metric localization provides detailed coordinates of a user in a given coordinate system
  – Use single global Cartesian coordinate system [Zhang 2014]
    • The entrance coordinate of a building can be used as the initial location, instead of brute-force searching.
Metric Localization Maps (cont’d)

- Use existed floor plan and directly label images.

- Image types in modeling and retrieving processes can be different.
Topological Localization Maps

- Topological localization does not provide absolute coordinates, but provide discrete location categories.
  - E.g. [Anjum 2011][Chaoulie 2013]
Three Major Issues

• Feature extraction
• Modeling via indexing
• Localization through retrieving
Localization through Retrieval

• Retrieval process includes
  – Capture one or more new images
  – Search them against the scene database
  – Find the most likely matched ones
  – Output localization results under certain matching criteria
One-step Methods

- **Topological retrieval:**
  - Support Vector Machine (SVM) [Altwaijry 2014], Nearest Neighbor (NN) [Murillo 2007], or Pyramidal Matching [Murillo 2007].

- **Metric retrieval:**
  - Trifocal tensor method [Murillo 2007], or Look-up table method [Zhang 2014].
Multi-step methods

- First classify input images into coarse categories, and then refine the localization result.
- Mathematically, divide search space into several subspaces.
  - Database size too large to put into memory, especially if implemented in mobile devices.
  - Save searching time
Outline

• Introduction
• Indoor Localization: An Overview
• 2D Image-based Approaches
• 3D Model-based Approaches
• Challenging Issues and Emerging Solutions
• Conclusion
3D Model-based Approaches

- Real physical space is three dimensional, even though it is more challenging to obtain 3D models
- It’s natural to sense the world with 3D sensors
- There are quite a few methods to obtain depth information
- Two classes of methods are covered:
  - Structure from Motion (SfM) based methods
  - Direct 3D sensor based methods
3D Approaches

- SfM from crowdsourced images
- SfM from self-collected images
- RGB-D sensor based methods
- Depth sensor only based methods
SfM-based localization

• SfM is the process of estimating three dimensional structure from two dimensional image sequences.

• Depending on the sources of images used:
  – Crowdsouced images localization
  – User-captured images localization

• How?
  1. Given an input image, extract 2D features;
  2. Match 2D features with 3D points in scene database;
  3. Use perspective **N-points (pNp) algorithm** to determine camera location and orientation
SfM with Crowdsourced Images

• Emergence of digital images / videos uploaded into Web.

• Rapid development of SfM techniques and software packages:
  – Commercial: Microsoft’s Photosynth, Autodesk’s 123D Catch, and 3DFlow’s Zephyr;
  – Open sourced: Insight 3D, SFMToolkit, Bundler, VisualSFM, and OpenSLAM.

• Advantages
  – Number,
  – Poses variants,
  – Imaging conditions.
2D-to-3D matching

• An example. [Sattler 2011]

• How?
• Database size problem
  – 1.5M SIFT points, 700M
• Solutions
  – Quantize SIFT descriptors [Irschara 2009], Sub-sampling [Sattler 2012], and model generating [Irschara 2009].
3D-to-2D

- 3D-to-2D method is an extension of 2D-to-3D method to improve feature matching and searching efficiency. [Li 2010][Sattler 2012]
3D Approaches

- SfM from crowdsourced images
- SfM from self-collected images
- RGB-D sensor based methods
- Depth sensor only based methods
SfM with self-collected Images

• Why self-collected images?
  – Localization area may not be popular public places; Web data is not sufficient or dense enough.

• Omnidirectional camera based SfM
  – Reduce number of images needed
  – Lead to a fast and efficient data acquisition process
    • Examples
Normal camera based SfM
3D Approaches

• SfM from crowdsourced images
• SfM from self-collected images
• RGB-D sensor based methods
• Depth sensor only based methods
Direct 3D sensor based methods

• 3D environment information can be directly obtained with 3D sensors, such as Microsoft Kinect, ASUS Xtion, etc.

• Depending on which part of 3D data is used, we have two classes of methods:
  – RGB-D based methods
  – Depth-only methods
RGBD based methods

• Create 3D occupancy map for the navigated area. [Lee 2014]
RGBD based methods
3D Approaches

- SfM from crowdsourced images
- SfM from self-collected images
- RGB-D sensor based methods
- Depth sensor only based methods
Depth only methods

• When floor plans are distinctive enough, depth data only can achieve localization. [Biswas 2012]
Outline

• Introduction
• Indoor Localization: An Overview
• 2D Image-based Approaches
• 3D Model-based Approaches
• Challenging Issues and Emerging Solutions
• Conclusion
Challenging Issues and Emerging Solutions

• Challenging Issues
  – Accuracy standard and sensors capacity
  – Computational cost and real-time performance
  – Evaluation of working systems

• Emerging solutions
  – Mobile platforms
  – GPU acceleration
Mobile platforms

• Many popular platforms are capable of implementing assistive localization systems
  – iOS/Android smart-phones/tablets
  – Smart watches
  – Wearable glasses

• Examples[Middleberg 2014]
GPU Acceleration

- **Graphics Processing Units (GPUs)**, originally designed to manipulate and alter memory to accelerate the rendering of images in a frame buffer intended for output to a display, can be extended to carry out many computer vision tasks.
  - SiftGPU [Wu 2015]
  - SURF GPU [Cornelis 2008]
  - GPU accelerated KinectFusion [Newcombe 2012]
  - Etc.
City College Visual Computing Laboratory

---

**Graph 1:**
- X-axis: Image size
- Y-axis: Speed (ms)
- Legend:
  - glsl
  - cuda
  - -fo -1 -glsl
  - -fo -1 -cuda
  - Parameter:
    - -n -s -fo 0
    - Image size, glsl, cuda
    - 320x240, 33.0 45.2
    - 640x480, 23.0 27.1
    - 800x600, 19.2 21.2
    - 1024x768, 16.3 16.8
    - 1280x1024, 13.2 12.9
    - 1600x1200, 9.9 9.0
    - 2048x1536, 7.0 X

**Graph 2:**
- X-axis: Feature Extraction
- Y-axis: Ms
- Legend:
  - Descriptor Generation Rotation Assignment
  - Image Loading
  - FPS: Frames Per Second
  - GF8800GTX Desktop
  - QFX1600M Laptop
  - 35 FPS
  - 103 FPS
  - 14.0 ms
  - 5.4 ms
  - 5.5 ms
  - 0.5 ms
  - 0.5 ms

---

2/4/2015
Outline

• Introduction
• Indoor Localization: An Overview
• 2D Image-based Approaches
• 3D Model-based Approaches
• Challenging Issues and Emerging Solutions
• Conclusion
Conclusion

• Conclusion
  – State-of-the-art methods and systems in non-visual based indoor localization.
  – 2D image based approaches
    • feature extraction
    • modeling via indexing
    • localization through retrieving
  – 3D model based approaches
    • SfM based localization
    • Direct 3D sensor based methods
  – Emerging mobile platform and GPU acceleration
Future Work

• Omnidirectional vision-based mobile front end plus GPU-enabled server end work together, by querying large scale pre-built geo-located scene databases, to provide an accurate, robust and real-time localization service.

• Different map building techniques, feature extraction or image representation methods, and searching methods combine together to achieve an optimal solution

• GPU acceleration and global optimization
Publications/Presentations


Thank you!

* This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 11337172. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.