Egocentric Place Recognition for AR Wayfinding Assistive Technology

Charbel Azzi, John S. Zelek
University of Waterloo
Waterloo, ON, Canada
cazzi/jzelek@uwaterloo.ca

Abstract

Wayfinding is something we take for granted whether it is from within our vehicles or by foot. We usually know where we are by our position to some landmarks in the context of some destination. GPS technology has helped augment our wayfinding in unfamiliar surroundings. If someone is perceptually (i.e., blind) or cognitively (i.e., dementia) challenged, then precise localization - more precise than a GPS - is essential for simple tasks like knowing where the curbs are or when crossing a road. The advent of geocoded image streetview data that provides a database for us to georeference to from an ego-centric worn camera when GPS is not accurate enough or non existent such as indoors. The method of augmentation of the experience does not necessarily lend itself to being visual with the lack of visual or cognitive senses. We have explored haptic augmentation due to its primitive reactive responses. Most of our current focus now are on the challenges associated with place recognition including the building of a map from streetview imagery and the indexing and searching with this geocoded database of images. The tasks include navigation, obstacle avoidance and accurate geo-referencing.

1. Introduction

The number of elderly people in the world is increasing in proportion to the total number of people. As people get older, their cognitive and perceptual faculties decline in functionality and sometimes fail entirely; for example, manifesting themselves in disease or illness such as blindness or Alzheimers disease. Approximately 14 million people in North America are affected by blindness and 7 million people suffer from Alzheimers. Technology can help alleviate issues of confinement, security and safety as well as empowering people who feel constrained by their condition. One such way is to ensure that they are still able to conduct their daily activities by providing them with technological navigational tools.

The ability to perceive the local and immediate environment for daily living activities is needed by all people. People with visual impairments as well as people with cognitive impairments require technology that identifies obstacles in their immediate path both inside and outside of buildings and assists in navigational tasks. Unanticipated factors such as construction work, streets and walkways in need of repair, garbage placement, poor weather, irregular traffic, other pedestrians and bicycles can make the basic challenge of navigating streets and sidewalks difficult. Improved localization as well as obstacle detection and avoidance capabilities (that augment existing aids such as the long cane or guide dog) will support safe independent travel, increasing access to school, work and the community overall as well as increasing self-esteem and independence.

We have designed, built and tested a tactile navigational belt that translates navigational orientational information (from a GPS positional fix differentiated towards a desired goal location) into tactile nudges along the waist. A GPS can tell people where they are and how to get somewhere but the interface can be complicated for the elderly. Field trials with persons with dementia demonstrated the success of the tactile interface [**?**], [**?**] (See Figure **??**. It appeared that the nudging of the waist was almost like a sixth sense in that no or very little cognitive processing was needed, almost a reactive type of response. We also reported upon a similar device used with people who were blind where a hand tactile device was used to convey orientation to avoid obstacles as sensed by a worn stereo vision camera [**?**] These two tactile devices show that tactile feedback as a local or global orientation guide is an effective feedback mechanism. GPS as a localization device is only accurate to a few meters (i.e., 3 m) and is not available indoors. Also the environment may contain obstacles such as bumps in the sidewalks. An egocentric camera can provide a finer localization fix when fused with GPS if available. Also, the sensed images can indicate obstacles and hazards that are to be avoided.

Route navigation in spaces that are too large to view from a single perspective is achieved through comparison of perceptual information about our orientation and motion

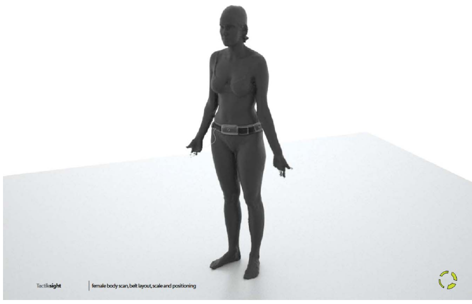

Figure 1. A rendering of the tactile belt that was used for dementia experiments. The actual device used was not as sleek.

to memorial representations of relevant spatial landmarks [?] and the relationships between landmarks. Those who are older, visually impaired, or with dementia all, to varying degrees and for different reasons, experience challenges in constructing and accessing accurate memorial representations of the environments in which they travel. Consequently, for these individuals, navigation can often be disorienting and frustrating.

2. Visual Place Recognition

Whether it is indoors or outdoors, one of the precursors for being able to localize based on a visual image is the need for a map with 3D information. With a collection of overlapping images of the same scene (as with streetview imagery), it is possible to acquire 3D camera pose (position and orientation) as well as 3D scene points. When images are registered against scene images, it is then possible to localize the person who is wearing a camera. This problem has been typically cast as an image retrieval problem where collections of features are matched against a database collection of geocoded images. There is typically a subsequent spatial verification step that helps refine the matching images to a smaller set, hopefully a single image and location. However there can be sets of visual elements appearing similar geometric configurations in unrelated database photos which can complicate the process [?]. Visual location recognition is also used in loop closure in robotic's processes to create maps (i.e., SLAM, SFM), landmark recognition, visual navigation and image based navigation. Some of the challenges in visual place recognition include that the appearance of a place can vary drastically, multiple places in an environment may look similar, places may not always be revisited from the same viewpoint (i.e., pose) and there may be perceptual aliasing (i.e., 2 different locations look the same) [?]. Ideally a place's appearance should be distinctive relative to nearby locations relative to some sensory information.

We assume that global context in addition to the uniqueness of a particular places leads to place recognition and localization. We have explored using the GIST global feature descriptor [?]. GIST descriptors are used to search for candidates for localization. Subsequently 3D points (characterized by the SIFT points used in building the 3D map) are used to localize purely on visual data. Our results on standard datasets show that our system can achieve better localization accuracy than the state of the art at a fraction of the computational cost.

3. Challenges

There are many challenges at various levels. For example a person with Alzheimer's may forget to put the belt on so it would be best to design the device into their clothing which is easier said than done. Google, Microsoft and others (i.e., open source Mapilliary) provide Streetview imagery that is updated every couple of years. In order to localize it is essential to first build a 3D map from the Streetview imagery. Using Marr's principle of least commitment, methods such as Visual SFM are preferred as the images, points matched and camera poses are all kept as opposed to a voxel or similar dense 3D map. Locating place at a particular time is useful for localization however it can also be used to detect obstacles or deviations from a normal street scene. For people who are cognitively or perceptually challenged, safety concerns such as those associated with falls are paramount. For example, the video feed from a worn camera can also be used to predict falls from deviations from a normal gait pattern. Currently our focus is on place recognition and localization. The 3D map we build can be based on SFM or SLAM pipelines however in order to improve accuracy all steps in the pipeline need to be evaluated as errors propagate in the pipeline.

References

[1] C. Azzi, D. Asmar, A. Fakih, and J. Zelek. Gist guided 3d keypoints selection in image-based localization. In *submitted to IROS 2016*, 2016. 2

[2] L. E. Grierson, J. Zelek, I. Lam, S. E. Black, and H. Carnahan. Application of a tactile way-finding device to facilitate navigation in persons with dementia. *Assistive Technology*, 23:108–115, 2011. 1, 2

[3] L. E. M. Grierson, J. Zelek, and H. Carnahan. The application of a tactile way-finding belt to facilitate navigation in older persons. *Aging International*, 2009. 1

[4] S. Lowry, N. Sunderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford. Visual place recognition: A survey. *IEEE Transactions on Robotics*, 32(1):1–19, 2016. 2

[5] T. Sattler, M. Havlena, K. Schindler, and M. Pollefeys. Large-scale location recogntion and the geometric burstiness problem. In *to appear in CVPR 2016*, 2016. 2

[6] J. S. Zelek, S. Bromley, D. Asmar, and D. Thompson. A haptic glove as a tactile-vision sensory substitution for wayfinding. *Journal of Visual Impairment and Blindness*, 97(10):621–632, 2003. 1