Demo: Real-time Object Tagging and Retrieval

Puneet Jain
Duke University

Romit Roy Choudhury
University of Illinois, Urbana-Champaign

ABSTRACT

We propose an augmented reality system on off-the-shelves smartphones which allows random physical object tagging. At later times, such tags could be retrieved from different locations and orientations. Our approach does not require any additional infrastructure support, localization scheme, specialized camera, or modification to smartphone's operating system. Designed and developed for current generation smartphones, our application shows promising initial results with retrieval accuracy of 82% in indoor environment without noticeable impact on the user experience.

If made commercially available, such system could be used in city tourism, infrastructure maintenance, and enabling new kind of social interactions.

Demonstration Setup

We expect to demonstrate our system on Android smartphones in MobiSys 2014 main conference. The users shall be provided with smartphones and would be asked to tag any random object at the conference venue. We shall simultaneously allow other users to retrieve same objects from different location, orientation, and phones. The demonstration does not require any additional setup help except stable WiFi connectivity from the conference committee.

We have prototyped our system on Android 4.3 and tested it with Samsung Galaxy S4 phones. We have used Amazon EC2 for the data storage and offloading of computationally intensive tasks.

Challenges and Related Work

Object tagging and retrieval in real-time is a challenging problem for several reasons:

1. GPS-Compass-Accelerometers are too noisy – it is hard to infer precise pose of the tagger, triangulation based techniques on short distances do not work.
2. Smartphones still have limited processing power – image processing and computer vision algorithms are computation hungry.
3. Application layer end-to-end network latency is too high – per frame cloud offloading is not feasible.
4. The retriever continuously scans the environment for tags using viewfinder – no two video frames may be similar, selective frame processing may not work.

Our approach to solving this problem is hybrid: leverage smartphone sensors such as Accelerometer, Compass, GPS, and Gyroscope to estimate quick and rough location-orientation of the phone and incrementally applying heavy-weight computer vision algorithms to reduce the search space of the matching tags. The former is used to build a rough and continuously changing in-device cache of tags in the vicinity, while the later is used to affirm the object matching using low to high fidelity algorithms.

Our system leverages a rich body of work in both the Mobile Computing and Computer Vision. Specifically, we use various sensor stabilization, filtering, triangulation, pose-estimation, and dead-reckoning techniques to continuously infer the rough pose using noisy inertial sensors. Ultimately to increase our object retrieval confidence, we have modified and customized several state-of-the-art image and scene analysis algorithms to run on Android operating system with minimal processing latency.

Real-time in-phone augmented reality is coming into realization with the projects like Google Tango [1] and Kinectfusion [2]. While promising, these projects are in still in prototype stages, require specialized hardward and sensors such as depth camera and GPUs. Our approach to solving this problem is inexpensive, practical, and compatible to wide range of phones.

Evaluation

We have evaluated our system by tagging random 20 objects in the Coordinated Science Lab, UIUC. The objects are later retrieved 5 times each, from different locations and orientations. Each trial retrieves three best matched tags from the repository. Among 100 retrieval trials, 54 trials resulted best matched tag as the desired tag, while in 16 and 5 trials, same appeared on the second and third positions. The object confusion matrix is shown in Figure 1.

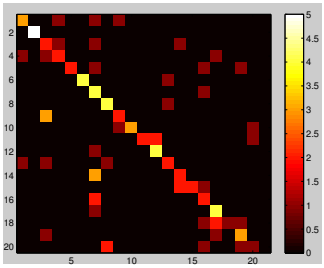

Figure 1: Tag retrieval accuracy in the form of confusion matrix – Tag X is confused with Y proportionate to intensity value

1. REFERENCES

[1] Project tango. https://www.google.com/atap/projecttango/, 2014.

[2] S. Izadi et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *UIST*. ACM, 2011.

MobiSys'14, June 16–19, 2014, Bretton Woods, New Hampshire, USA.
ACM 978-1-4503-2793-0/14/06.
http://dx.doi.org/10.1145/2594368.2601472 .