A Novel Augmented Reality Framework for Museum Exhibits

Julien Li-Chee-Ming, Zheng Wu, Randy Tan, Ryan Tan, Naimul Mefraz Khan, Andy Ye, Ling Guan
Department of Electrical and Computer Engineering, Ryerson University
Toronto, Canada
{julien.li, zheng.wu, randy.tan, ryan.tan, n77khan, agye, lguan}@ryerson.ca

***Abstract*—This paper presents a novel augmented reality (AR) framework that runs on Android mobile devices (smartphones and tablets). The AR framework uses the mobile device's camera and inertial measurement unit (IMU), and does not rely on external infrastructure (e.g., WiFi or Bluetooth beacons, targets, etc.). The proposed AR solution combines a vision-based object recognition algorithm called bag of words (DBoW2 [1]) with a simultaneous localization and mapping (SLAM) solution called ORB-SLAM [2]. The potential of the AR solution is demonstrated in a museum application. With the objective of enriching the learning experience, typical museum artifacts have been enhanced with interactive and engaging 3D AR animations.**

Keywords—augmented reality; SLAM; object recognition

Fig. 1 shows the Cheerful Oak Stove artifact (circa 1920) at the Canada Science and Technology Museum in Ottawa, Canada. When the user views the artifact using the mobile device's camera, DBoW2's object recognition module automatically retrieves the corresponding AR content and ORB-SLAM map from the database. The AR content allows the user to virtually interact with the artifact by dragging logs of wood into the stove via the mobile device's touch screen. The user can then place a lit match in the stove to set the wood on fire. ORB-SLAM's camera localization allows the AR content to maintain alignment with the artifact as the user moves the camera to view the artifact from different perspectives.

There were 8 AR experiences in total. Table I shows average result from 10 trials at each experience. Where each trial involved the user opening the application, going through the AR experience, then closing the application. The results include the average time taken for DBoW2 to recognize the object, properties of the ORB-SLAM map (i.e., the number of keyframes, the number of map points, the file size, and the average load time on the Google Pixel phone), the average time taken for ORB-SLAM to localize with respect to the map, and the average reprojection error upon localization. The reprojection error is the image distance between a projected point belonging to the 3D AR content and its corresponding 2D point in the image. 5 check points, evenly spread throughout the object, were measured to calculate each scene's reprojection error with an image size of 640x480 pixels. DBoW2 performed uniformly across all experiences, with 100% precision and recall, and approximately equal recognition times. Larger artifacts required larger areas to be mapped, resulting in larger map file sizes and map load times. To view the full extent of the artifact in the camera's field of view, these experiences required farther camera-to-object distances. The experiments revealed several effects as camera-to-object distance increases: Error in the map increased, the map points were sparser, and less ORB features were extracted from smoother images. These effects manifested as longer localization times and larger reprojection errors. The results also indicate that the localization time is too long for real-time augmented reality applications. Further, tracking often failed because of large camera movements between consecutive localization updates. To solve this issue, the device's IMU-based rotational tracker was used to update the camera orientation in between ORB-SLAM's localization updates. The IMU provided an orientation estimate every frame (20 Hz), however the pose drifted over time and the rotational tracker did not estimate changes in the device's position. ORB-SLAM's localization was called in a separate thread at 0.5 Hz, based on the results in Table 1, to provide a periodic drift correction.

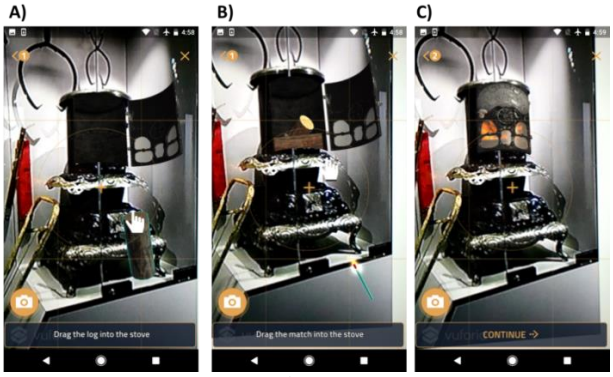

Fig. 1. Stove artifact with AR content.

REFERENCES

[1] D. Gálvez-López and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188-1197, 2012.

[2] R. Mur-Artal, J. M. Montiel and J. D. Tardós, "Orb-slam: a versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147-1163, 2015.

TABLE I. Average result from 10 trials at each AR experience. The results include the time taken to recognize the object, properties of the ORB-SLAM map (i.e, the number of keyframes, the number of map points, the file size, and the load time, the time taken to localize the camera, and the reprojection error.

Recognition time (s)	# key frames	# map points	File size (Mb)	Map load time (s)	Localization time (s)	Reprojection error (pix)
0.47 ± 0.32	89.75 ± 35.16	1067.25 ± 299.91	4.56 ± 2.09	1.83± 0.58	1.85± 0.69	41.43 ± 16.84

*This research was financially supported through an NSERC Collaborative Research and Development (CRD) Grant (#507333-16).