Figure 1: Integrated sensors and the calibration cube.

ABSTRACT

We present a method for *in situ* modeling of polygonal scenes, using a laser rangefinder, an IMU and a camera. The main contributions of this work are a well-founded calibration procedure, a new hybrid, driftless orientation tracking method and an easy-to-use interface based on natural interactions.

Index Terms: I.2.10 [Vision and Scene Understanding]: Modeling and recovery of physical attributes—Motion

1 INTRODUCTION AND RELATED WORK

3D models of scenes are often required in the field of Augmented Reality (AR), whether for the purpose of camera tracking, virtual objects positioning or occlusion handling. *In situ* modeling is generating increasing interest in the community as it makes it possible to build AR applications in unprepared environments. In this paper, we propose a new method for interactive modeling of polygonal scenes. The user is physically present within the scene to be modeled, equipped with a tablet PC, a webcam, an inertial measurement unit (IMU) and a single-point laser rangefinder (figure 1). The scene is observed through the camera from a fixed position, and polygonal regions are outlined on the images using a mouse or directly by pressing the touchscreen. Laser scanning is then performed inside the regions, providing depths and orientations of the polygons in the camera coordinate system.

Several *in situ* modeling methods have already being proposed. Some of them are purely vision-based, but they require either to generate sufficient motion parallax to get depth information [2], which can be problematic in large-scale environments, either to use an incremental approach [5], which requires that the polygons in the scenes are connected. Moreover, the scene is generally described up to a scale unless objects of known dimensions are detected in the images. A laser rangefinder coupled with a vision-based orientation tracker was used in [7]. 3D point clouds are first acquired from which planar regions are automatically detected. However, automatic detection of planes has some issues and requires relatively dense point clouds. In [4], a "3D cursor" made of a laser and a IMU is used to define 3D vertices of polygonal objects. However, using an IMU alone to get the camera orientation lacks of accuracy. Moreover, the object boundaries are often difficult to aim at with the laser beam as confusions with the background can easily occur. Finally, [7, 4] require careful physical alignments between the sensors, whose relative positions and orientations are measured manually.

In this paper, we argue that integrating basic user interactions in the modeling process is likely to provide more reliable results while
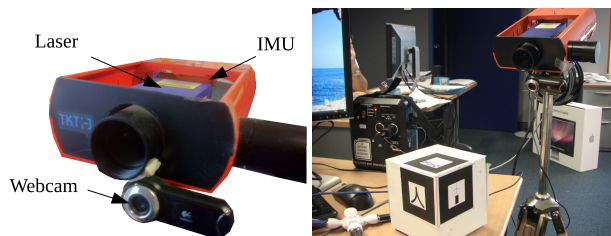
*e-mail: christel.leonet@gmail.com
†e-mail: gilles.simon@loria.fr
‡e-mail: marie-odile.berger@loria.fr

allowing sparser laser measurements. The main contributions of the work presented here are: (i) a well-founded calibration method which provides the orientation of the IMU as well as the origin and the direction of the laser beam in the camera coordinate system (section 2), (ii) a new hybrid, driftless orientation tracking method inspired by the tracking-by-synthesis algorithm [5]. This method takes advantage of both the IMU robustness and the computer vision accuracy, without having to use environment mapping that has several issues as detailed in section 3; (iii) an easy-to-use interface based on natural interactions. Visual hints are provided to help the user get the best possible accuracy (section 4).

2 HARDWARE AND CALIBRATION

Our mobile system is composed of a tablet PC Dell Latitude XT2, a camera logitech 9000, an IMU Xsens MTi and a Laser Jenoptik LDM 42 supplied with a portable 12V battery (figure 1). A plastic cube covered with three ARToolkit markers is used to calibrate the device and express all sensor measurements in the camera coordinate system. The IMU/camera relative orientation is obtained using standard handeye calibration [1]. For the laser/camera calibration, we marked a point at a known position on the cube. The cube and/or the laser/camera combined device are moved so that the laser light spot hits the marked point; the 3D coordinates of this point are obtained in the camera frame using marker detection. Repeating this operation several times, we get a set of 3D points belonging to the laser beam, expressed in the camera frame and associated with a laser distance. A PCA is then performed to get the principal axis of this set of points, that is the direction of the laser beam in the camera coordinate system. The origin of the laser beam is then obtained by propagating the laser distances along this axis.

It must be noticed that the system calibration is done once and for all, on the understanding that all sensors stay rigidly mounted on their common support.

3 ORIENTATION TRACKING

Several methods have been proposed in the past for hybrid [1] or purely vision-based [3, 6] orientation tracking. In [3], sparse optical flow is used for frame-to-frame tracking and heavyweight landmark features (SURF descriptors) are used to correct the drift. However, continuous laser scanning requires accurate temporal estimate of
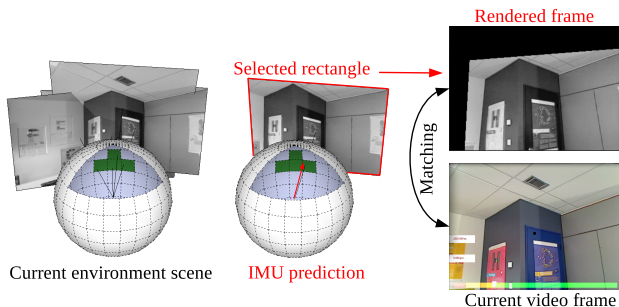
Figure 2: The orientation tracking process.

the laser beam pointing direction while landmarks detection cannot be done every frame due to performance issues. In [6], a cylindrical environment map is built on the fly, in which FAST corners are detected and matched to the current frame. However, flattened cylinders have discontinuities on the left and right borders and suffer from nonlinearities along the vertical axis. Moreover, merging images acquired at different times into a single map often generates geometric and/or photometric misalignment which can significantly disturb the detection and matching process.

To tackle these issues, we propose a new method inspired by the tracking-by-synthesis algorithm [5], adapted to pure rotational motions (figure 2): (i) a sphere is sampled in the spherical coordinate space, (ii) the sphere area corresponding to the camera orientation predicted by the IMU is selected, (iii) if this area has not been visited yet, a textured rectangle fronto-parallel to the camera is added to an "environment scene" by back-projecting the current video frame, (iv) if the area has already been visited, the rectangle associated with this area is rendered using the camera orientation predicted by the IMU; the rendered frame is matched with the current video frame using cross-correlation between Harris corners and a RANSAC computation of a planar homography; the camera orientation is finally refined using the generated correspondences.

The sphere discretization angle is chosen depending on the desired overlap between images of adjacent areas (3/4 in our experiments). By selecting complete frames rather than using a combined image, we avoid all misalignment, nonlinearities and discontinuities issues. This method is very fast as the OpenGL graphic engine is used to render the predicted images. Finally, by keeping *non*-normalized images, we have more chance to get comparable luminosity and contrast between the rendered image and the video image, as the lighting changes are generally repeatable during a short working session, depending on the orientation of the camera.

Each time a view is added to the environment scene, Harris corners are matched between this view and the views associated with the already visited adjacent areas. The orientation of the new view is iteratively refined in order to optimally satisfy the whole set of generated correspondences. Global bundle adjustments of all views in the environment scene are regularly performed, based on the correspondences obtained each time a view is added. This operation is performed on a separate thread, so as not to slow down the tracking process. Note that these two tasks (local and global adjustments) naturally ensure loop closures of the environment scene, due to the unflattened spherical representation.

4 SCENE MODELING AND USER INTERFACE

While the environment scene is built and the camera orientation is tracked, the user describes 2D polygons by clicking their vertices in the video frames. A color bar is displayed to help the user assess the accuracy of the tracking based on the number of inliers (green: good tracking, orange: poor tracking, red: IMU only - see

figure 2-bottom right). A black square cursor is superimposed with the laser light reflected at the object surfaces. When this cursor is detected inside a user-defined polygon, a 3D point is obtained using the camera orientation and laser distance and is associated with the polygon. When at least four 3D points are known for a polygon, a RANSAC PCA is performed to compute the principal plane and the 3D coordinates of the polygon vertices in the camera coordinate system. The polygon color passes from blue to purple when a well-distributed planar point set is detected, namely when $(\lambda_1 + \lambda_2)/I > 70\%$ and $\lambda_2/I > 20\%$, where $\lambda_1 > \lambda_2 > \lambda_3$ are the PCA eigenvalues and $I = \lambda_1 + \lambda_2 + \lambda_3$.

5 EXPERIMENTAL RESULTS

Figure 3-left shows some modeling results obtained in an office, superimposed over a 3D reference model of the office. In addition to this visual assessment, the recovered dimensions of some posters are provided on the figure. All posters have a size of 59×84 cm except the one at foreground that has a size 83×61 cm. The mean relative error of the recovered dimensions is 5.5% (the mean distance between the camera center and the posters is 2.9 m). Figure 3-right shows other results obtained in a house, superimposed over architectural floor plans. These two examples prove that our method is able to manage 360-degree scanning of scenes with an accuracy sufficient for many purposes, and particularly for AR tasks.
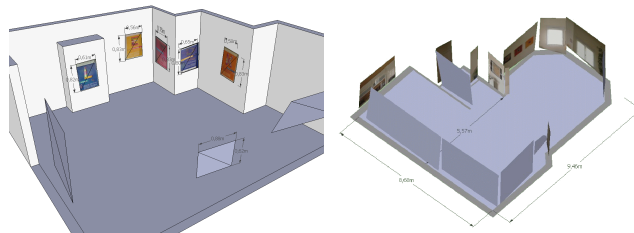

Figure 3: Experimental results. Left: modeling results obtained in an office with recovered dimensions. Right: a 3D model obtained in a house, superimposed over architectural floor plans.

6 CONCLUSION

We presented an easy-to-use tool for *in situ* modeling, based on a new accurate and efficient hybrid orientation tracker. We are now working on adding a GPS to this system, in order to be able to continue modeling a scene from other viewpoints and finally get a complete AR system where virtual objects insertion and 6-degree-of-freedom camera tracking, based on the GPS data and the reconstructed polygons, will be allowed.

REFERENCES

[1] M. Aron, G. Simon, and M.-O. Berger. Use of Inertial Sensors to Support Video Tracking. *Computer Animation and Virtual Worlds*, 18:57–68, 2007.

[2] P. Bunnun and W. W. Mayol-Cuevas. OutlinAR: an assisted interactive model building system with reduced computational effort. In *Proc. IEEE/ACM ISMAR 2008*, Cambridge, UK.

[3] S. Diverdi, J. Withert, and T. Hllerert. Envisor: Online environment map construction for mixed reality. In *Proc. IEEE VR 2008*.

[4] T. Hoang and B. Thomas. In-situ refinement techniques for outdoor geo-referenced models using mobile AR. In *Proc. IEEE ISMAR 2009*, Orlando, Florida, USA.

[5] G. Simon. In-Situ 3D Sketching Using a Video Camera as an Interaction and Tracking Device. In *Eurographics 2010*, Norrköping, Sweden.

[6] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg. Real-time panoramic mapping and tracking on mobile phones. In *IEEE VR 2010*, Boston, MA, USA.

[7] J. Wither, C. Coffin, J. Ventura, and T. Hollerer. Fast annotation and modeling with a single-point laser range finder. In *Proc. IEEE/ACM ISMAR 2008*, Cambridge, UK.