Mobile Virtual Interior Stylization from Scale Estimation

Shintaro Murakami* Tomoyuki Mukasa† Tony Tung†
The University of Tokyo* Rakuten, Inc.†

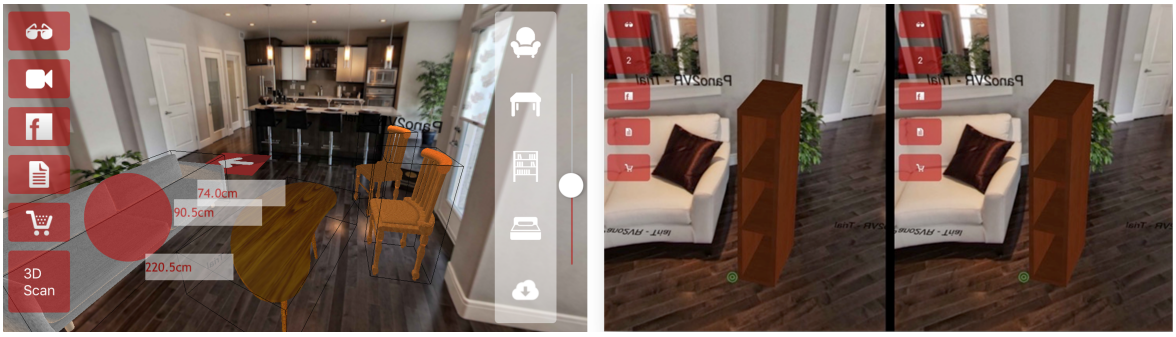

Figure 1: *Mobile AR/VR for interior design. (Left) Virtual objects representing furniture scaled to real-world scene seen through mobile device camera. Absolute scale is obtained by estimation of device elevation with respect to floor level, given object dimensions. (Right) Navigation in immersive virtual environments (e.g., fully stylized apartment) using stereo rendering and VR headset.*

Abstract

We present a new feature for AR/VR applications for consumer mobile devices equipped with video camera (e.g., smartphone). Direct or indirect scale estimation of scene or objects is necessary for realistic rendering of virtual objects in real-world environment. Standard approaches usually rely on 3D vision with sensor fusion (e.g., visual SLAM), or pattern recognition (e.g., using AR markers, reference object learning), and suffer from various limitations. Here, we argue that combining inertial measurements and visual cues, the problem reduces to a 1D parameter estimation representing distance from device to floor. In particular, we discuss robust solutions to solve absolute scale estimation problem for indoor environments.

Keywords: AR, VR, mobile CV, 3D vision, 360 degree imaging

Concepts: •**Computing methodologies** → **Virtual reality;**

1 Introduction

Mobile applications for augmented reality (AR), mixed reality (MR) and virtual reality (VR) have become ubiquitous in the past decade thanks to the rapid evolution of mobile device technology. Image sensors and IMU for mobile (i.e., Inertial Measurement Unit, consisting of gyroscope and accelerometer) have enough precision to track users' head motion and provide compelling experiences in immersive virtual reality. Consumer products like Oculus Rift, Gear VR, Vive, Cardboard and Hololens are paving the future of the Internet as new communication tools. While applications are numerous (e.g., games, education, tourism, etc.), the creation of realistic immersive environments in 3D still requires professional resource. Nevertheless, simpler alternatives based on AR and/or 360 degree images (e.g., Google Street View, or using 360 degree camera) are also commonly employed.

In visual effects society, 3D vision techniques have been commonly employed to insert virtual 3D objects in videos of real-world environment (e.g., Matchmoving [Dobbert 2012], Boujou [Boujou v5.0.2]). Similarly, robot vision community has been extensively relying on SfM and visual SLAM to capture observed 3D environments [Engel et al. 2014].This requires to recover 3D information from sequence of observed images, assuming efficient feature extraction and tracking. While these approaches have been successfully applied to reconstruct outdoor scenes and textured objects, it is still challenging to perform dense feature extraction in indoor scenes. In this case, contours and lines can also be exploited [Hofer et al. 2016]. Furthermore, object recognition can be applied to retrieve scene scale [Wu et al. 2015].On the other hand, mobile AR/VR applications are becoming ubiquitous. In its simplest form, AR technology has been used to display virtual objects where a known marker (i.e., flat pattern) is detected. This has the advantage of directly providing objects and floor orientation. However, markers are necessary, only one location or direction is observed at a time with limited field of view, and performance depends on marker detection (i.e., lighting condition). In addition, virtual objects are usually not displayed at real-world dimension.

Here, we propose a markerless approach, that allows users to create and navigate in mixed reality scenes, with natural rendering using real and virtual objects. Hence, authored scenes can be explored in virtual reality using a VR head mounted display. A hand-free head-based navigation system, using the system inertial sensors, allows users to simply navigate from scene to scene. For example, a user can use the VR App to upload 360 images and place virtual furniture in each room of his house, and then explore or showcase his complete interior design in immersive virtual reality (see Fig.). An essential key feature is that all virtual 3D objects have known dimensions and are scaled to real-world dimensions in observed scenes. Absolute scale estimation is obtained by estimating the mobile device height with respect to the floor. In what follows we review the different solutions to achieve robust absolute scale estimation in indoor environments where feature points are sparse.

*e-mail:shin-m@ut-vision.org
†tomoyuki.mukasa@rakuten.com, tonytung.org

SIGGRAPH '16, July 24-28, 2016, Anaheim, CA,
ISBN: 978-1-4503-4371-8/16/07
DOI: http://dx.doi.org/10.1145/2945078.2945092

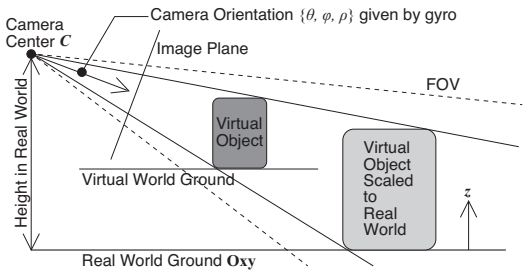

Figure 2: *Estimating the distance between the device camera* ***c*** *and the ground plane* ***Oxy*** *allows the scaling of virtual object to real-world scene.*

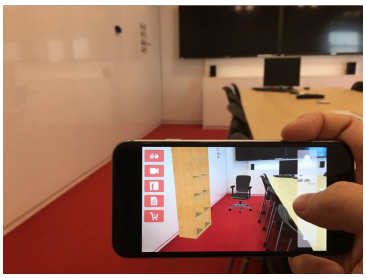

Figure 3: *AR/VR application running on iPhone 6. Sensors return vertical axis that is used to estimate floor orientation. 3D objects can then be displayed in real scene, in real-time, without markers.*

2 Scale estimation

When observing a scene through a video camera, assuming pinhole camera model, there is scale ambiguity λ. Even if synthesized 3D objects have known dimensions, orientation and scale have to be adjusted to observed scenes.

Sensors in modern consumer mobile devices (such as IMU) can nowadays provide various accurate measurements. Since device orientation $\{\theta, \phi, \rho\}$ can be obtained from gyroscope, and assuming objects lie on a flat (ground) surface **Oxy**, it is therefore straightforward to derive virtual objects' pitch and roll angles, θ and ϕ, while yaw angle ρ depends on user's preference. Hence, to place a virtual object o with known dimensions in a real-world scene observed by a (video camera) viewpoint **c**, only the height h of the mobile device to the ground has to be estimated (see Fig. 2). Virtual objects are first rotated according to $\{\theta, \phi, \rho\}$, and lie on a plane parallel to **Oxy**. Scale ambiguity means the distance between **c** and o is initially unknown. However, if the virtual ground plane of o is set at distance h from **c** along **z**, then o is scaled to observed real-world scene. Let us denote h_v, d_v, and d_r the distances between **c** and the virtual plane, the virtual object, and the virtual object placed in real-world scene. Since $\lambda = \frac{h_v}{h} = \frac{d_v}{d_r}$, if $\lambda = 1$ and $h_v = h$, then $d_v = d_r$, i.e., the virtual object is scaled to real-world scene.

Furthermore, in order to cope with indoor scenes containing few objects, and where room corners are visible, we propose the following alternatives:

1. Human statistical measurement: height of device is set with respect to average standing human (e.g., 1.45m). Parameter can be adjusted manually or automatically using user's profile, or based on the two following solutions.

2. Simple reference object: since device and object orientations can be derived using mobile IMU, floor position can be estimated using a known flat object, such as an A4 paper. Device height and object scale can then be estimated from measurements based on observation.

3. 3D structure from visual features: we combine 3D structure from motion, line extraction, and IMU sensor fusion to estimate scene scale, using EKF as in visual SLAM techniques [Kolev et al. 2014]. Ground floor approximation is obtained by fitting a plane which normal is parallel to the vertical axis given by the device gyroscope, based on the distribution of 3D structures. Assuming perspective projection, planar floor, and by pre-adjusting absolute scaling factors and field of view (based on camera parameters), we can achieve robust markerless augmented reality without requiring textured floor.

3 Applications

The proposed features have been integrated in virtual furniture simulation and interior design applications, which have become popular for e-commerce services and social platforms (see Fig. 3).

The mobile AR/VR application has been first developed for the iPhone 6. The smartphone returns accurate measurements from the 6-axis IMU at 2000 Hz, and features a 4.7-inch display with a "2x" resolution of 1334 x 740 (326 ppi), as shown in Fig. 3. This ensures a comfortable VR experience for short-term use. AR display is based on Vuforia and OpenGL. VR display is based solely on OpenGL. Computer vision based algorithms were implemented using OpenCV. 3D models come from public dataset [Trimble 2006]. Current prototype contains 20 3D models of interior objects hard-coded in the app, plus 3D objects placed in a cloud that can be easily downloaded in predefined order. We use a consumer VR HMD for immersive virtual reality experience.

4 Remarks

A version of the application integrating our algorithm is also being developed for Android smartphones. While it substantially enlarges the range of compatible devices, performances depend on the various models and are still being investigated.

To our knowledge, no similar AR/VR mobile application that combines virtual object scaling to real scenes, immersive VR, and integrated authoring tool for full interior stylization has been proposed in the literature or as commercial product.

References

BOUJOU. v5.0.2. www.vicon.com/products/software/boujou.

DOBBERT, T. 2012. *Matchmoving: The Invisible Art of Camera Tracking*. John Wiley and Sons.

ENGEL, J., SCHÖPS, T., AND CREMERS, D. 2014. LSD-SLAM: Large-scale direct monocular SLAM. In *ECCV*.

HOFER, M., MAURER, M., AND BISCHOF, H. 2016. Efficient 3d scene abstraction using line segments. *CVIU*.

KOLEV, K., TANSKANEN, P., SPECIALE, P., AND POLLEFEYS, M. 2014. Turning mobile phones into 3d scanner. In *CVPR*.

TRIMBLE. 2006. 3D warehouse, 3dwarehouse.sketchup.com.

WU, Z., SONG, S., KHOSLA, A., YU, F., ZHANG, L., TANG, X., AND XIAO, J. 2015. 3d shapenets: A deep representation for volumetric shapes. In *CVPR*.