ABSTRACT

This paper proposes a framework to support interaction with virtual objects integrated in a real life scenario. In the proposed interaction concept, the user can reshape or re-design a real space using virtual objects using several pictures of the desired space. The images are analyzed for known features such as surfaces, floor or room orientation. Using these elements, it is possible to devise an augmented reality system where the user can add virtual objects to the scenario that react to its content. These are magnetic objects because they attach itself to elements on the scene such as floor or walls. This approach liberates the user from moving objects in three-dimensional space. Instead they just push around objects in a more natural interface. The paper studies the impact of adding these magnetism effects in the interaction with an augmented reality-editing interface.

Categories and Subject Descriptors
H.5.1 Multimedia Information Systems - Artificial, augmented, and virtual realities.

General Terms
Design, Experimentation, Algorithms.

Keywords

1. INTRODUCTION

Virtually reshaping a known space can be a troublesome task because it usually requires the creation of a 3D model of that same space in order to introduce and visualize new objects and 3D content. Another approach would be to use an image of the space, captured with any camera the user has available, use the photo as input, and add virtual objects on top. The problem here is that most users do not have the required skills to correctly place objects in a manner that is visually correct and appealing.

The main contribution of this paper is the introduction of magnetic augmented objects that are aware of the properties of the photographed scene, and react to them as if they were in the real world. These properties can be the existence of floor or walls, the orientation of the scene, the vanishing points or the existence of obstacles in the scene.

Our main contribution is the integration of high-level features obtained with computer vision algorithms in an augmented reality interface. In this work a case study was developed to showcase the use of such features. This case study presents examples of snapping and glue techniques that are based on the detected features.

The main novelty of this paper is the use of state of the art techniques from the computer vision field in an augmented reality interface. The idea is to bring computer vision and image processing applications into the realm of interfaces. It tries to use images and pictures taken by the users (using widely available cameras with no extra hardware or marks) to build an environment where they can virtually change their physical spaces.

2. RELATED WORK

The tools discussed in this work could be useful in interior design applications, architecture or augmented reality games. This work follows the steps of several augmented reality systems by using the user workspace image as input [6,16], analyzing the scene to detect features [4,14] and superimposing virtual objects [6,14,16]. It benefits from several previously developed computer vision algorithms [4,6,11,12]. The explorations and reconstruction of interiors from photos has also been previously explored [4,7,12,14].

We are experimenting with images that can be taken with a regular camera or webcam. There are other techniques that use extra equipment, such as: accelerometers [1], depth sensing cameras [5] or AR marks for calibration [2]. There are several commercial applications that explore the notion of reconstructing an interior design space [1,2]. Most of them only allow interaction with predefined models, or use markers to define where the virtual objects should stay. Tri-dimensional interpretation of a scene is also an active field of research [12]. Using depth-sensing cameras [5] can improve the results of the scene reconstruction but they are currently not as widely available as regular cameras.

Using snapping to facilitate interaction was also discussed in several projects [3, 10]. It has been extensively researched in CAD projects [10], although not in the context of automatic feature detection from images as in this paper.

3. INTERACTION CONCEPT

The main proposal is the virtual transformation of a real space by someone with no real experience with geometry and 3D manipulation. The user is encouraged to take one or several pictures of an indoor scenario, with a regular camera, a webcam or a TabletPC, as seen in figure 1. After that, the interaction takes place in the computer or tablet. Taking as an example, the context of an interior design application, the user can experiment with...
several virtual objects (i.e., furniture), testing them in different positions [8]. Moving the objects should take into consideration the affordances provided by the photographed space.

Objects should react to the floor, to object orientation (most interior house arrangements mainly consider a Manhattan world [7]) and to existing volumes. Objects should be pushed back and forth, left and right without the user having to worry with the third-dimension. The final result is then the original image, augmented with objects that are correctly positioned without the help of markers [14].

4. PROTOTYPE DESIGN

To implement and evaluate the interaction concept, a prototype was designed to capture and analyze images and provide a platform for magnetic augmented objects. The current prototype is clearly separated into three stages: (1) capturing the image, (2) analyzing the image with image processing and computer vision algorithms and (3) interacting with it with touch or mouse input.

4.1 Capturing the Scene

The current prototype supports the introduction of images through webcams or directly through an image file. Capturing the right picture of the scene is vital for the system to work. Since most of the system is based on the analysis of the image, it is important that the image has enough detail to detect features and lines. Blurred images should also be avoided. The user is instructed to take a picture holding the camera straight. It should have reference points such as edges, doors or windows. Using the webcam mode our system advises the user in real-time of whether an image is detailed enough as depicted in Figure 2.

![Figure 1. The user takes a photo of a space using a device such as a tablet with a camera. Using the photo she can add virtual augmented objects that react to elements of that image, such as floor or room orientation.](image)

4.2 Analyzing the Scene

In order to enable natural interaction between objects and scene the image must be preprocessed to detect certain properties. Here we will focus mainly in the following properties: (P1) main vanishing points, (P2) room orientation, (P3) floor and (P4) 3D depth relation between different points of the scene. Currently, certain assumptions are considered: the picture is vertically aligned, with the floor on the bottom and the scene is relatively squared following a Manhattan world (most indoor scene have this property). The current algorithm has a similar approach of previous model based 3D predicting systems [7,9].

For the detection of the main vanishing points [10] (P1), the input image is preprocessed using a sequence of operations (Figure 3). Initially the contours of the image are detected. To clean the clutter several Erode and Dilate operations are applied. Next, using a Hough line detection algorithm the main straight lines that compose the figure are detected. Using a RANSAC (Random Sample Consensuses) system several sets of lines are intersected in order to create candidate vanishing points (green dots in Fig. 3). Using clustering it is possible to define the most probable vanishing point (in red in Fig.3) [9]. This entire process is done several times with different thresholds and parameters, in order to improve the result, using a score system. The system performs well in most cases although it is not guaranteed to work in all images [7].

![Figure 3. Image analysis to detect vanishing points, room orientation and floor.](image)

Since all parallel lines have the same vanishing point, it is possible to orient (P2) objects so that their vanishing points are the same as the one of the room. This gives the feeling that objects are aligned with the room and belong there (Figure 4 and 5). Taking into consideration the camera parameters (extracted from image EXIF or previously known), it is possible to infer the floor (P3) by using the average height of a person [9].

The current prototype also explores another property, the 3D depth (P4) of certain keypoints. This property requires at least two images taken side-by-side. For each image the SIFT [16] points and descriptors are extracted. By matching the points in the two images, it is possible to calculate the displacement of each point. Using stereovision it is also possible to understand the depth of
each point, using this displacement. Using the depth and the position of each point a 3D structure can be constructed using the Delaunay triangulation algorithm as represented on the right side of Figure 6. This 3D structure can be used to support the placement of 3D objects.

4.3 Adding Virtual Objects

Using the four properties defined in the previous section, it is possible to create an interface to add virtual objects to the scene that are automatically positioned and oriented with the layout of the photographed scene. As an example, Figure 4 shows three objects that are automatically placed on the floor and oriented towards the detected vanishing point. The user only has to push the objects on the floor, back and forth (Z axis) and left and right (X axis). The rotation and movement in the Y axis of the objects are optional, and used only if the user does not want to apply the detected properties.

![Figure 4. Prototype screenshot. Image with superimposed virtual objects. The furniture is placed in a 3D world that has the same orientation as the original image. Far away objects fade into the detected vanishing point.](image)

The current prototype supports three tools that help users to place virtual objects in the scene: (T1) magnetic floor snapping, (T2) rotation snapping and (T3) direct object glue. In the first tool, shown in Figure 5, taking advantage of property floor (P3), when the user approaches the object to the floor it automatically snaps to the ground [3]. Rotation snapping, also present in Figure 5, happens when the object is aligned with the room (using P1 and P2). The third tool (T3) acts as a glue (Figure 6), the object is pinned to the spot touched by the user. It integrates the object in the scene using the 3D depth information (P4). The user places the object directly on the desired spot and the system tries to guess the scale and 3D rotation of the object using the normal of the 3D triangles and the depth of the vertex. This last tool (T3) gives a good hint on the scale of the objects although it is prone to noise. T1 does a better job on the floor and T3 is more suitable for walls and volumes. Using these three tools or attaching the objects directly to the floor, it is possible for a user with no 3D modeling and editing experience to quickly place an object in a scene.

![Figure 5. Prototype screenshot presenting the magnetic snapping system. When the user drags the object down, the floor attracts it, showing yellow guidelines (T1). When the object is rotated it snaps every 90 degrees when its orientation is aligned with the scenario (T2), showing the green guidelines.](image)

5. EVALUATION

In order to assert the effectiveness of the detected properties and proposed tools we devised a preliminary user study. The goal of the study was to understand the user difficulties in the manipulation of 3D objects and the gain achieved by introducing the proposed tools. The idea is to validate the need of a system that helps users align objects in an existing photo and to evaluate our prototype and high-level features.

The study consisted in two tasks and a questionnaire. The first task (K1) was designed to test tools T1 and T2. Users had to repeatedly place the box (Figure 5) in different points of an image with right position and scale factor (Example: “Consider a box, half the size of the bed”). Each time they had to do it with and without the guidelines provided by T1 and T2, the execution time was recorded. The second task was meant to test T3 (Figure 6). The users had to place the poster in a given place and then they had to do the same without help, using translations and rotations. The goal was to test the amount of time required to do the same operation without help of the tool. The 3D control was done using a mouse. The questionnaire helped in the characterization of each user.

The study involved 20 users, aged between 20 and 50 years old, and most of them had average technological (95%) skills but no 3D modeling skills (20%). The results show that on the first task (K1) there was a 38% performance increase when using snapping (T1 and T2), as seen in Figure 7. When positioning the objects without assistance in 3D 43% of the times users placed the objects
in geometric incorrect positions (wrong scale or rotation). With the help of the snapping system, that number decreased to 11%. Almost all users (95%) agreed that the snapping helped their task, 25% said the improvement was small, and 70% said the improvement was large. On the second task (K2) users took an average time of 12.19±3.90s to achieve the same result without help of the glue (T3). When asked which techniques they would prefer to use in an applications to virtually reshape a room with objects such as furniture, most users preferred that they would use glue (T3) initially (60%), followed by snapping (T1 and T2) (35%) and then free movement in the end. The glue tool (T3) raised concerns of lack of freedom so it must be used only in an initial approach. It should also be possible to turn of the snapping for refined adjustments.

These results show that there is a significant performance increase when using this kind of magnetic properties. They also show that users have a real difficulty in correctly placing objects in a 3D scene. Most users also show openness in using an augmented reality system that correctly interprets the scene and agree that the system improves the proposed tasks (Figure 8). Once they start using the system, they build expectations and may expect it to deal with certain situations such as occlusion, which are currently not supported.

6. CONCLUSIONS AND FUTURE WORK
The study shows that there is interest in the interaction with virtual objects following the rules and constraints of the real world. The current prototype, effectively demonstrates the detection of several properties and the use of editing tools that take advantage of these properties. It requires further integration and development to work in more portable computer devices (i.e., tablets) to enhance the experience of the user. Other topic that is being considered is the virtualization of a given space by the user for sharing and editing. Other users could then use it and make their own versions or explore it in 3D a shared space.

7. ACKNOWLEDGMENTS
Removed for blind review.

8. REFERENCES
1. Atelier Pfister.
   http://www.pfister.ch/.
2. DesignMyRoom.
9. Anonymized