ABSTRACT

In our current work we examine the development of visuo-haptic augmented reality setups and their extension to collaborative experiences in entertainment settings. To this end, an expandable system architecture supporting multiple users is one of the most indispensable prerequisites. In addition, system stability, low latency, accurate calibration and stable overlay of the virtual objects have to be assured. In this paper we provide an overview of our framework and present our collaborative example application, an augmented reality visuo-haptic ping-pong game for two players. The users play with a virtual ball in a real environment while, by using virtual bats colocated with haptic devices, they are able to feel the impact of the simulated ball on the bat.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities; I.3.7 [Three-Dimensional Graphics and Realism]: Virtual reality

General Terms

Design, Performance

Keywords

Augmented reality, haptics, collaboration

1. INTRODUCTION

The underlying idea of augmented reality (AR) systems is the combination of real and virtual objects into one environment [2]. However, the perception of the augmented scene is usually restricted to the visual domain and no haptic cues are provided, limiting user interactions. The integration of haptic interfaces into AR systems removes this constraint and permits to not only see the virtual objects, but in addition to feel them. This can be used for numerous applications where natural, three-dimensional human computer interfaces are necessary, which allow transferring skills and knowledge from and to daily life.

Using augmented reality environments as user interfaces for games has been a topic of research for a while, as described for instance in [7, 12, 18, 5, 8]. Some research related directly to our example application of AR ping-pong have been carried out in the past. For instance, in [19] a virtual table-tennis setup with real rackets has been developed. However, the system did not include haptic feedback and only used the rackets as an interface to position a virtual racket on the screen. A similar approach was also taken in [11], which neither provided haptic feedback. In [13], a collaborative, mobile AR-based billiard game has been discussed, already incorporating tactile feedback. In most of these works, either no details on overlay accuracy, stability and latency are given, or the reported results are of lower quality. Regarding the integration of haptic feedback in general AR setups only limited work has been carried out. Some groups have experimented with adding props to provide passive haptic feedback in the AR environment, e.g.,[14, 6, 15]. However, active behavior of scene objects cannot be generated with these systems. Recent research has examined the possibility of adding devices providing active haptic feedback. A few proof-of-concept systems have been developed in this context, for instance, in [17, 10, 20, 1]. However, lag reduction, exact alignment, or error minimization is only seldomly addressed. Moreover, the haptic device is usually not colocated with the visual representation of the augmented objects, thus compromising hand-eye coordination. In addition, if simultaneous interaction with real and virtual objects should be allowed, visuo-haptic colocation in the augmented scene is indispensable. In this context, the real world represents a reference frame into which the virtual elements have to be perfectly integrated. Any errors in this process would reduce the usability of a system, and compromise user interaction and immersion.

To alleviate the currently existing shortcomings, we have developed a high precision, collaborative, colocated visuo-haptic augmented reality environment. Main components are a distributed, low latency framework for synchronization, the accurate calibration of the haptic device in world coordinates and a hybrid tracking setup. In this short paper we describe the developed framework and present an example testbed – multimodal AR ping pong. Figure 1 depicts the overall setup for one user as well as his augmented view of the scene.
2. COLLABORATIVE AR FRAMEWORK

The computational requirements for achieving stable multimodal interaction in multuser AR go beyond the capabilities of currently available hardware. Therefore, the system framework had to be distributed to several machines. The core components of our system are the physics server, a graphics and a haptics client. The first machine performs the physical simulation of the virtual scene. A graphics client carries out all tasks typical to a standard AR setup. The external input data needed on this client are position and orientation of the haptic interfaces, as well as the movement of the virtual objects in the scene. Finally, haptic interfaces can be added to the environment with the haptics clients, which communicate directly with the physics server.

In our example setup, one device is directly connected to the physics server, while a second interface is added via a haptics client. To provide high level of immersion, the visual data acquired from the simulation always has to coincide with the moment the image was taken on the graphics machine. Therefore, accurate synchronization between the physics server and a graphics client is required. To this end, the actual clock difference is determined at startup of the application. This computation is done by a comparison of the clocks on both machines, while taking into account the round trip time. After synchronizing both clocks, the graphical data is stored in a timestamped ring buffer and the graphics client can request data for the timestamp the image was taken. The data are transmitted, temporarily stored and updated when needed by the rendering process. Therefore, no latency is added to the overall delay. One round trip of the communication takes 500–900$\mu$s for small-sized packets. The data exchange between the physics server and the haptics client is collision free with a round trip time below 200$\mu$s. Figure 2 illustrates all information communication occurring in the framework.

In our testbed setup for two-player interaction in multimodal AR, the physics server is run on a dual processor PC with 2.4GHz CPUs (2GB RAM, 512kB cache). Rendering and tracking on the graphics client is done on a dual processor PC with 2.8GHz CPUs (2GB RAM, 512kB cache). To the former a PHANToM is attached, while to the latter the optical position tracking device and the cameras are connected. A second haptic device is attached to a third 1.0GHz single CPU machine (256MB RAM, 256kb cache) running a haptics client. All components are using Linux OS. As external tracker we use the infrared (IR) optical 6DoF tracking device OPTOTRAK 3020 manufactured by Northern Digital Inc. The view of one user is captured by a head-mounted Videre Design MEGA-DCS FireWire camera (640x480 pixel image at 30 fps). In addition, a user-independent top view of the scene is generated by an Allied Vision Technologies FireWire camera (800x600 at 30fps). Since only one video-see-through HMD was available, one player currently has to rely on the images displayed via the overhead camera. A second HMD setup is already manufactured for our system. Finally, to allow touching of virtual objects, two SensAble PHANToM devices are integrated into our AR setup. As previously mentioned, the haptic devices are colocated with the augmented environment, thus allowing interaction with virtual and real objects via the same tool. Therefore, high demands for tracking and calibration have to be met, which will be discussed in the following.

3. HYBRID TRACKING

An IR marker is attached to the head-mounted camera and tracked by the OPTOTRAK system to determine the user’s head pose. The system measures and computes marker orientation and position with submillimeter accuracy at an optimal distance of about two meters. Since the camera and the marker are rigidly attached, the camera-marker transformation is fixed and can thus be determined by Hand-Eye calibration [16]. The IR tracking data are inherently noisy due to the inaccurate measurements of the LED positions, especially when the head-mounted marker moves. As a consequence, the registration between the real and the virtual world is affected, causing instability of virtual objects in the augmented images. Therefore, we correct the estimated camera pose of the IR optical tracker with a vision-based approach. The first step is to detect precalibrated visual landmarks in the recorded images. Furthermore, to ensure that a landmark is sufficiently visible, an occlusion test is carried out and occluded landmarks are discarded. Given the 2D-3D correspondence of the remaining landmarks, the position and the orientation of the camera are then refined by error minimization. An image space approach is used to estimate the camera pose as a nonlinear least squares problem. A RMS backprojection error of less than 0.7 pixels can be achieved in less than 1ms computation time. Moreover, using additional offline refinement of 3D landmark positions, the error can be further reduced to about 0.3 pixels [3].
4. HAPTIC DEVICE CALIBRATION

In order to align the virtual representation of the haptic interaction point with the correct physical location in the real world, the relationship between the haptic and the world coordinate system needs to be determined. The first step of our calibration procedure is to collect 3D point measurements in the coordinate systems of the haptic device and the optical tracker. To this end an optical marker is attached to the stylus of a PHANToM device. For the measurements, the tip position is stabilized in the middle of the haptic workspace by rendering a fixation force with the device. The marker assembly is then manually rotated following a spherical movement in space, while the marker poses are being recorded. The best fitting sphere is then determined using pivot calibration [9], giving the marker-tip transformation. This allows obtaining corresponding point measurements in both the haptic and world coordinate system. After acquiring 3D point data, the absolute orientation problem has to be solved. Due to additional errors, resulting from inaccuracies in haptic encoder initialization, a two-staged optimization process is followed [4]. The final calibration results yield an alignment error better than 1.5mm. Figure 3 demonstrates the calibration results for visuo-haptic colocation. A virtual table tennis bat is aligned with a real handle attached to the haptic device.

5. COLLABORATIVE, VISUO-HAPTIC AR PING-PONG

To demonstrate the high fidelity of the system, an AR ping-pong environment was chosen as testbed scenario for our collaborative visuo-haptic augmented reality system. This choice has been driven by the importance of fast collaborative interactions during gameplay. Real handles of table tennis rackets are attached to the haptic devices. The actual bat and the ball are, however, overlaid virtual objects. In order to integrate the real table and the net, these objects are initially calibrated in the world coordinate system, thus, allowing the virtual ball to collide with them. A simplified rigid body dynamics model is used for physical modeling of the ball. The haptic devices are used to render impact forces to the users and to control the virtual bat. The overall rendering performance of the augmented scene at a resolution of 640x480 pixels is illustrated in Table 1. Due to the multi-threaded implementation, the images are displayed at about 20Hz with an overall latency of about 60 – 100ms. The scene as shown to one of the users is depicted in Figure 4.

<table>
<thead>
<tr>
<th>Component</th>
<th>Average time [ms]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame rendering</td>
<td>15</td>
</tr>
<tr>
<td>Frame unwrapping</td>
<td>5</td>
</tr>
<tr>
<td>Corner detection + occlusion test</td>
<td>20</td>
</tr>
<tr>
<td>Pose refinement</td>
<td>0.6</td>
</tr>
<tr>
<td>Object rendering</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1: Rendering performance

6. CONCLUSION

We have presented a colocated visuo-haptic augmented reality environment for two-player ping-pong. A dedicated system providing accurate calibration, high stability, and low latency had to be developed. A problem of the current setup is the lack of binocular stereo cues. This limits the interactivity of the system since depth information has to be inferred from shadows, occlusions and motion parallax. Moreover, the workspace of the currently used haptic device is rather limited. Therefore interactions are restricted to a certain area. In addition, due to the used device only 3DOF of force feedback can be provided.

Next steps will be the extension to stereo vision for better depth perception. In addition, to reduce network traffic and to ease the synchronization a hardware trigger signal will be integrated. Also, extension to 6DOF haptic feedback and alternative haptic devices will be investigated.
Figure 4: Colocated visuo-haptic AR ping-pong for two players.

7. ACKNOWLEDGMENTS

This work has been performed within the frame of the Swiss National Center of Competence in Research on Computer Aided and Image Guided Medical Interventions (NCCR Co-Me) supported by the Swiss National Science Foundation.

8. REFERENCES


