Performance vs Quality of Experience in a remote control application based on real-time 3D video feedback

PERFORMANCE VS QUALITY OF EXPERIENCE IN A REMOTE CONTROL APPLICATION BASED ON REAL-TIME 3D VIDEO FEEDBACK

Enrico Masala, Antonio Servetti

Control and Computer Engineering Department, Politecnico di Torino — 10129 Torino, Italy
e-mail: masala@polito.it, servetti@polito.it

ABSTRACT

This paper presents a preliminary study that relates the ability in performing remote control tasks with 3D video feedback and the users' quality of experience (QoE) when users are asked to judge the video quality only. The QoE has been evaluated by means of traditional subjective quality experiments, while an objective performance indicator (alignment error) has been used in the remote control task. Results show that, unexpectedly, even a relatively poor subjective video quality still provides sufficient 3D perception to achieve satisfactory alignment performance.

Index Terms— stereoscopic video, remote control, quality of experience, low-cost hardware

1. INTRODUCTION

It is known from several researches [1, 2] that stereoscopic vision helps in remote control applications. However, despite the recent diffusion of stereoscopic devices in the consumer market, most of the studies are still based on professional grade devices. Thus, it remains unclear if it is possible to employ such cheap devices to create 3D teleoperation systems similar to the more expensive ones at a fraction of the cost. In fact, while their low price makes those devices particularly attractive, their performance is typically limited by stringent cost constraints. Clearly, the optimal quality of experience (QoE) of professional level systems cannot be obtained, however it is yet unclear if their capabilities would be sufficient to ensure at least an acceptable QoE.

We developed a system, using commercial off-the-shelf (COTS) devices, that trades off video quality with low latency to meet the delay constraints imposed by real-time applications. In fact, the limited processing capabilities of the mobile device required to work at low video quality in order to be able to process images at an acceptable frame rate.

In this work we investigate how much the QoE reduction may affect the ability to perform tasks, i.e., the performance of the operator using the remote control system with the 3D video feedback.

This work was partially funded by the Regione Piemonte through the FAROS project (POR.FESR07-13.I.1.3, project no. 147-311).

2. EXPERIMENTAL SETUP

Our setup consists of a commercially-available mobile phone (HTC Evo 3D) with a stereoscopic camera, which has been mounted on a radio-controlled (RC) toy car. A real-time video is transmitted using a wireless channel to a computer with a stereoscopic display that allows 3D experience by means of active shutter glasses.

The software has been developed ad hoc for this particular setup with the primary aim to minimize latency. Although the hardware capabilities allowed to capture images with 1280×720 resolution in stereoscopic mode, an acceptable latency could be achieved with our setup only when the resolution was equal or lower than 640×360. More technically, the transmitter part relies on the primitives provided by the Android OS, and video compression routines have been written in native code to improve performance. The receiver software runs on a desktop PC with Linux OS, equipped with the Nvidia 3D vision kit and a graphics hardware supporting OpenGL with the stereoscopic capability.

Fig. 1. Sample video scene captured by the stereoscopic video system. Rendered as a red-cyan anaglyph for printing only.

In the presented setup the 3D camera on the RC car is facing front-wards, and it is pointing to a glass fixed on the car itself as it can be seen in Fig. 1. The experiment consists in using the 3D video feedback to remote control the car so that the glass is aligned in two directions (transversal and depth) under a suspended static ball, so that the ball, if released, would fall inside the glass. Clearly, the use of 3D video is fundamental to achieve good precision in the depth direction. The cues useful to get a 3D perception have been limited as much as possible: background and floor are uniform, no shadows, etc.

Thus, the depth perception is expected to come only from the image disparity. Once the remote operator judges the alignment as satisfactory, the actual alignment error (in transversal and depth direction) is measured. The alignment error is considered as an indicator of the performance obtained in the remote control task.

We conducted two experiments, one to evaluate the video QoE and the other to evaluate the performance achieved in each task as a function of the quality of the 3D video feedback.

First, after an initial phase where the subject familiarizes with the 3D technology used in the experiment, the subject is exposed to pre-recorded videos similar, in terms of quality, to the ones that will be encountered in the experiments. Three videos at different resolutions are compared to a reference high-quality video. The subjective opinion about the quality of the videos is recorded by means of the Double Stimulus Continuous Quality Scale (DSCQS) [3].

Second, the remote control experiment is performed for different values of video quality, i.e., different image resolutions (640×352, 416×240, 320×176), while the quantization parameter is kept constant. Each remote control experiment is repeated three times.

3. RESULTS

Currently, only subjective evaluations can provide reliable 3D video quality estimates [4]. The subjective video quality corresponding to the three resolutions used in the experiments is shown in Fig. 2. Five subjects, screened for normal visual acuity, color and depth perception have been employed in the tests. The video quality gap between the lower and intermediate resolution is significant, 12 on a 100-point scale, while the gap between the intermediate and high resolution is 4 points only. Therefore, we expect a significant performance degradation in the control task when using the worst resolution.

The positioning precision achieved in the remote control experiments while using three different video resolutions is shown in Fig. 3 using three different symbols. Since the transversal plane is orthogonal to the operator point of view, the transversal error is more limited than the depth one, i.e., points density is higher near the horizontal zero axis.

As expected from the QoE experiment, the performance of the remote control task, in terms of position accuracy in the depth direction, is related to the quality of the video as

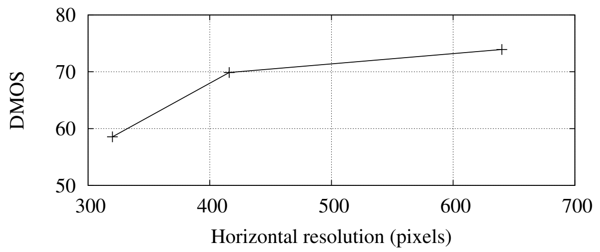

Fig. 2. Subjective video quality at three different resolutions used in the experiments.

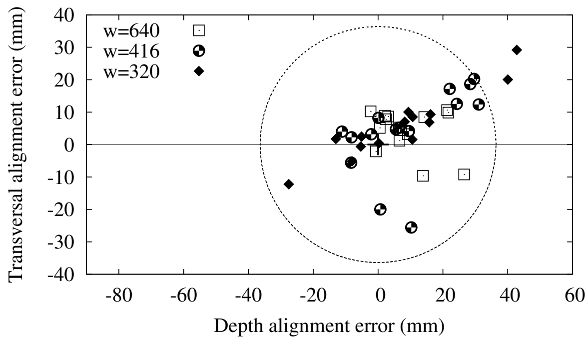

Fig. 3. Position achieved by various subjects at different horizontal video resolutions (w value). The dashed line is the limit to declare the alignment as successful for the experiment.

experienced by the user. However, quite surprisingly, results indicate that it is possible to achieve good accuracy even when the QoE of the video feedback is very limited, e.g., horizontal resolution (w) equal to 320. Although the depth error increases as the video resolution decreases, it does not cause the failure of the alignment task, i.e., the ball would fall inside the glass in nearly all cases. In other words, the poor video feedback still allows to successfully perform the remote control task. Therefore, results suggest that the processing power limitations of low-cost hardware do not significantly affect the capability of performing this kind of tasks. This fact potentially allows significant cost reductions in implementing remote control systems based on real-time 3D video feedback.

4. CONCLUSIONS

The performance of operators using a remote control system based on a real-time 3D video feedback has been quantitatively evaluated by means of a remote control experiment for different video quality levels in terms of video resolution. Although the precision in performing the tasks decreases with the video subjective quality, the overall trend suggests that even a low-quality 3D video image can convey sufficient information to successfully perform tasks where depth is important. This result provides evidence that even cheap hardware with limited capabilities may be sufficient to implement remote control systems based on 3D video feedback. Future work will be devoted to increase the number of both subjects and tests, and to investigate more complex tasks.

5. REFERENCES

[1] S. Livatino et al., "Mobile robot teleguide based on video images," in *IEEE Rob. & Aut. Mag.*, Dec. 2008, pp. 58–67.

[2] M. Ferre et al., "3D-image visualization and its performance in teleoperation," in *Proc. Intl. Conf. on Virtual Reality (ICVR)*, Beijing, China, July 2007, pp. 22–31.

[3] ITU-R Rec. BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," Jan. 2012.

[4] L. Goldmann and T. Ebrahimi, "3D quality is more than just the sum of 2D and depth," in *Proc. Hot3D workshop*, Singapore, July 2010.