SMARTPHONE-BASED 3D REAL-TIME VISION SYSTEM FOR TELEOPERATION
Demo Paper

Antonio Servetti, Enrico Masala

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

ABSTRACT

We present a small form factor 3D vision system that can easily be mounted in any system for teleoperation and presents a low-latency suitable to perform interactive control. The proposed system is based on a commercial 3D smartphone that integrates a stereoscopic camera and a wireless connection. The smartphone has been customized to acquire and transmit stereoscopic video in real-time by means of a special purpose software that runs on the Android operating system. We believe that the choice of a compact solution based on an open source framework and commercial off-the-shelf hardware, will promote the widespread adoption of this architecture by interested developers in multiple scenarios.

Index Terms— 3D video, teleoperation, real-time multimedia, wireless communications, Android

1. INTRODUCTION

3D vision systems have been shown to improve teleoperation tasks in several researches [1]. While earlier works used special purpose hardware, recent studies often employ commercial off-the-shelf solutions that have been made available on the market. However, recent technological advances were mainly focused on the video acquisition part of the system and did not consider the transmission part. Here, standard notebooks or computers were used in the vast majority of the implementations [2, 3]. A computer is needed because 3D cameras by themselves either lack the capability to transmit the video in real-time, or they provide it in a format that requires further processing before transmission.

In this work we present a solution that overcomes this limitation and enables the possibility to realize a small form factor low-cost 3D vision system that can easily be mounted in any system for teleoperation and that presents a low-latency suitable to perform interactive control. In our system the 3D camera, the processing unit and the networking hardware are integrated in a single compact device, i.e., a commercial 3D smartphone. On this device, on top of the Android operating system (OS), we developed a custom software to acquire and transmit stereoscopic video in real-time as a 3D visual feedback for remote control applications.

One of the key advantages of the framework is that it can be easily improved and customized because it is based on open platforms both for the sender and for the receiver (the latter being based on the Linux OS). Potential applications range from teleoperation, car driving, surveillance, domotics, etc. Since the system has been developed for teleoperation, the coding and transmission software has been implemented with particular attention to (i) reducing the transmission latency to fit the limits required for interactive control, (ii) reducing the impact of transmission errors on the video quality to guarantee an optimal video feedback for the operator.

2. SYSTEM DESCRIPTION

A block diagram of the prototype is illustrated in Fig. 2. It is composed of two main parts: (A) a mobile node, equipped with the mobile 3D remote-vision device, that transmits the stereoscopic video to (B) a receiver node for 3D visualization.

The following hardware components, shown in Fig. 2, have been selected for the implementation. *(1) 3D smartphone*. We used an HTC Evo 3D with a 5 megapixel stereoscopic autofocus camera, an 1.2 GHz dual core Qualcomm Snapdragon (ARMv7) processor, 1 GB of RAM memory and WiFi b/g/n support. The OS used for the software development was Android 2.3 Gingerbread with Linux kernel version 2.6.35.13 release htc-kernel@amd18-2. *(2) Wireless access point*. We used a Linksys WRT54GL WiFi router with 802.11
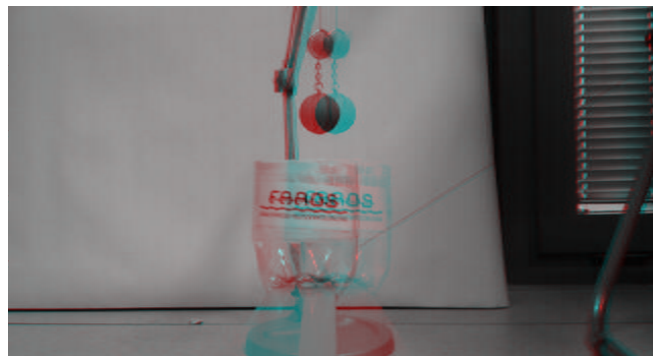

Fig. 1. Stereoscopic view captured from the 3D smartphone during a test. Rendered as a red-cyan anaglyph for printing only.
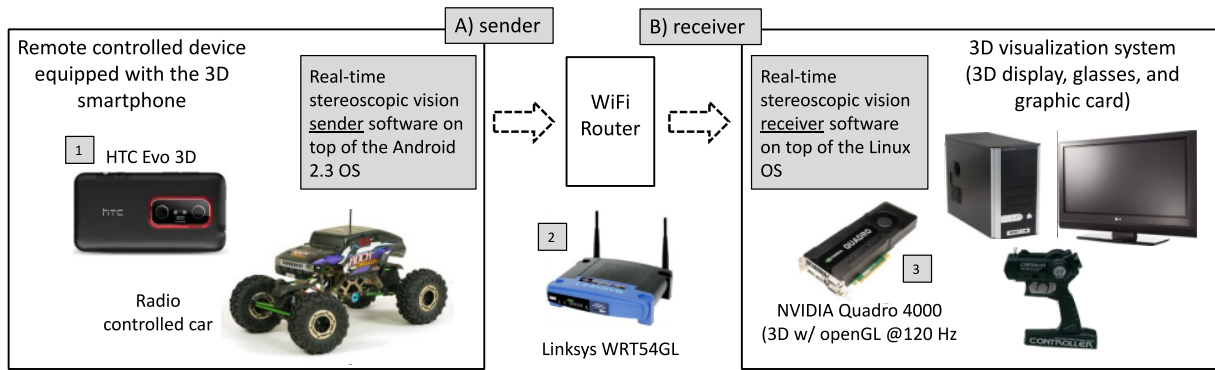
Fig. 2. Block diagram of the system components with the details of the hardware used.

b/g support. However, any standard WiFi router can be used.
(3) 3D vision system. Our choice to develop the system for the Linux OS limited us to consider only graphics cards with hardware OpenGL 3D support, so we used a Nvidia Quadro 4000 mounted on a PC with an Intel Core i7 64 bit processor. However, if the receiver software is ported to the Windows OS, low-end graphics devices with software 3D support can be used to significantly reduce the cost of the overall system.

The real enabler of the system is the custom software developed on the smartphone for low-latency robust 3D video transmission over WiFi. It consists of a sender and a receiver application. *The sender* has been implemented on top of the Android platform and includes native code routines to optimize stereoscopic video compression that may benefit from the dual core processor architecture. Video frames are captured from the 3D video camera and rescaled to different resolutions for transmission (e.g., 1280x720, 640x352, 416x240 pixels). A sample frame is shown in Fig. 1. The raw compressed data are then subdivided in independently decodable chunks, each one designed to fit into a single IP packet to achieve better resilience to communication errors. Chunk content is encoded by means of a JPEG-like algorithm, specifically tuned to limit unnecessary header overhead, and sent using the RTP protocol. *The receiver* has been implemented on top of the Linux OS, it ensures decoding of partially corrupted frames and real-time rendering featuring a multi-threaded architecture which decouples the packet reception and decoding from data visualization so that the interaction with the device rendering engine does not affect the time required for processing the input data.

3. EXPERIMENTAL RESULTS

Informal subjective tests have been realized to test the suitability of the system for remote control operations by mounting it on a radio controlled car. In these tests the operators were asked to perform simple teleoperation tasks such as driving the car in presence of obstacles or aligning the car with respect to a given marker. The latter scenario is depicted in Fig. 3. Both scenarios revealed the usefulness of the stereoscopic video feedback system and demonstrated the usability

of the system for such type of tasks.

Among the performance indicators of the system we may cite: (i) an overall end-to-end delay (over an infrastructure wireless connection) lower than 200 ms; (ii) frame rate equal to 20-22 fps at a resolution of 416x240 pixels and 14-18 fps at 640x352 pixels (the fps drops below a usable level if the full resolution of the camera of 1280x720 pixels is used); (iii) a network bitrate around 1148 kb/s for the case of 12 fps and 640x352 pixels (that was the preferred setup for the operators used in the tests).

4. REFERENCES

[1] A. Woods, T. Docherty, and R. Koch, "Field trials of stereoscopic video with an underwater remotely operated vehicle," in *Proc. of SPIE*, Apr. 1994, vol. 2177, pp. 203–210.

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

[3] 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.
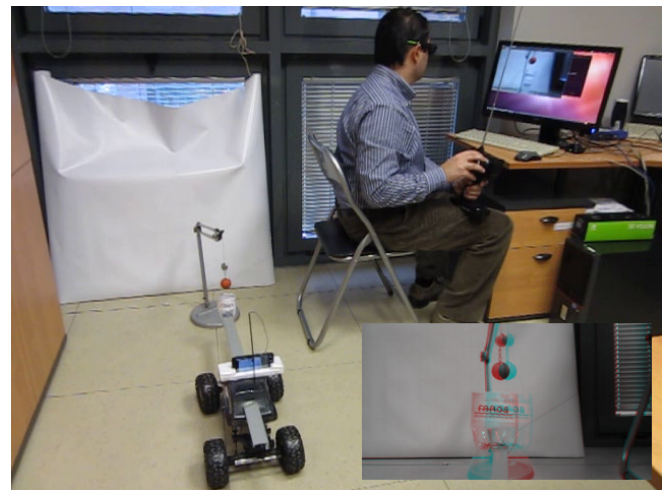
Fig. 3. The 3D vision system tested in a teleoperation task where the operator is asked to position the glass (attached to the car) exactly under the suspended ball.