Focused Technology Advances Series

Scalable Video Coding for Telepresence in the Performing Arts

Ivan V. Bajić and Xiaonan Ma, Simon Fraser University, Canada
ibajic@ensc.sfu.ca, xma2@sfu.ca

Introduction

With the development of broadband networking technology, the focus in the new media and performing arts community is shifting towards distributed telepresence performance (also known as telematic performance) – a scenario where two or more groups of performers, located in geographically distant sites, are able to perform interactively for the audience located at either site, or in an independent venue. For a number of years, universities and research centers in North America have been linked with gigabit optical networks whose high bandwidth makes such performances feasible. Yet, we are not seeing many of such events available to the wider audience. Part of the reason is that the traditional performance venues, theaters and music halls, are not linked to the high-bandwidth research networks and therefore cannot accommodate the high bandwidth requirement associated with these events.

Compression, especially video compression, appears to be the answer to some of these challenges. However, careful examination of the requirements of a telepresence performance reveals that there is more to it than simply using an off-the-shelf video codec. First, video coding is found to be detrimental to interactive communication since it introduces additional delay and makes the resulting compressed bit stream more sensitive to packet losses in the network [1]. In fact, high-end videoconferencing solutions try to stay away from video compression in order to improve user experience [2]. Commercial products, such as LifeSize, opt for hardware compression which may be suitable for corporate videoconferencing, but lacks the flexibility and accessibility needed for artistic performance.

In order to answer some of these challenges, we have developed a real-time software-only video codec for the most popular software environment used by media artists – Max/MSP/Jitter [3]. The codec is based on the well-known SPIHT algorithm [4], and feeds the video stream directly into the Max/MSP/Jitter environment, enabling easy on-stage video manipulation. Both the codec and its software environment are briefly described below.

Max/MSP/Jitter

Max/MSP/Jitter (or Max, for short) is a graphical programming environment for multimedia [3], widely used by media artists in new media productions. This environment supports real-time manipulation of video, 3-D graphics, audio, and other data in a unified processing architecture, enabling seamless on-stage performance. It is available on Mac OS and Windows platforms.

In Max, programs (called "patches") are created by connecting objects which perform specific functions. These objects can be written in C++, Open GL, Java, or simply be composed of other sub-patches. Data flow is controlled by the chords connecting these objects. All data are abstracted as multidimensional arrays (matrices). Data exchange between objects is synchronized by the internal clock which can be set at millisecond precision. If a particular chord carries a matrix or higher-dimensional array, it will appear as thick and green, otherwise it will be thin and black. Detailed information on Max can be found at [3].

Scalable Video Codec for Max/MSP/Jitter

We have developed a scalable video codec for Max based on SPIHT [4]. In the current version, only intra-frame coding is supported. While this means that compression efficiency is not as high as it could be with inter-frame coding, the codec has several advantages which are very important in a live performance. First, its CPU usage is low, allowing other video and audio manipulation tasks to run on the same machine. Second, it provides the level of robustness against dropped frames which cannot easily be achieved with inter-frame coding.

The codec has been used in t_2, a telematic dance performance premiered in Vancouver in July 2009. The trailer for the performance can be found at [5]. The performance involved live video transmission between two sites in downtown Vancouver via a conventional residential Internet connection at roughly 500 kbps. In addition, live video from a moving car, transmitted via a 3G mobile Internet connection at roughly 100 kbps, was a part of the performance. This appears to be one of the first (if not the first) use of mobile wireless video in a live telematic dance performance.

Figs. 1 and 2 below show simplified sending and receiving Max patches, respectively, utilizing our codec. In this example, the machine with IP address 142.58.88.162 is sending live video to another machine with IP address 199.60.10.135 on port 8000. The encoder appears as the object **mcl.jit.spihtaritenc** in the sending patch (Fig. 1), while the decoder is **mcl.jit.spihtaritdec** in the receiving patch (Fig. 2). Captured and decoded frames are also shown for reference in their respective patches.

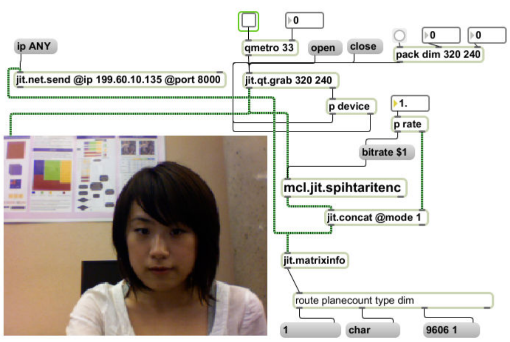

Figure 1: Max patch sending video to IP address 199.60.10.135

In addition to the simple live video unicast, the codec has also been demonstrated in the live scalable video multicast. In this scenario, live video was encoded and sent to the transcoding peer, whose task was to adjust the bitrate for several heterogeneous receivers. The transcoding was easy due to the progressive nature of SPIHT bitstreams - it amounted to simply truncating the bitstream of each frame to the points appropriate for receivers' bandwidths. Further illustration and several demo clips about our collaborative work on telepresence in the performing arts are available at [6]. External codec objects for Windows and Mac OS versions of Max, as well as sample Max patches that show how to use them, can be obtained from the authors.

Figure 2: Max patch receiving video from IP address 142.58.88.162

Acknowledgement

This work was supported in part by the NSERC/CCA New Media Initiative Grant STPGP 350740. We would like to thank Prof. John Crawford of UC Irvine for introducing us to the technical aspects of telematic dance, and Dr. Henry Daniel of SFU for choreographing telematic dance performances that utilize our codecs.

References

[1] J. R. Cooperstock, "Interacting in Shared Reality," *11th Intl. Conf. on Human-Computer Interaction*, Las Vegas, NV, Jul. 2005. www.cim.mcgill.ca/sre/publications/

[2] McGill Ultra-Videoconferencing System website: ultravideo.mcgill.edu

[3] *Max/MSP/Jitter version 5*, Cycling'74 www.cycling74.com/products/max5

[4] A. Said and W. A. Pearlman, "A new fast and efficient image codec based upon set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.* vol. 6, pp. 243–250, Jun. 1996.

[5] t_2 trailer: barry.is2.net/T2/t2trailer.mov

[6] X. Ma, Scalable Video Coding for Dance Telepresence website: www.sfu.ca/~xma2/

Ivan V. Bajić is an Assistant Professor in the School of Engineering Science at Simon Fraser University, Burnaby, BC. His research interests revolve around signal processing and its applications in image and video coding, multimedia communications, and computational biology. He has authored about a dozen, and co-authored another three dozen publications in these fields. He was recently the Telepresence Architect for t_2, a telematic dance performance premiered in Vancouver in July 2009.

Xiaonan Ma received the B.A.Sc. degree in Electrical Engineering from the University of Ottawa in 2007. She is currently pursuing the M.A.Sc. degree in the School of Engineering Science at Simon Fraser University, Burnaby, BC. Her research interests include video coding and multimedia networking for new media and performing arts. She was recently the Video Coding Engineer for t_2, a telematic dance performance premiered in Vancouver in July 2009.