INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 13, Number 2, Pages 205–219

ARBITRARY RESOLUTION VIDEO CODING USING COMPRESSIVE SENSING

HONG JIANG, CHENGBO LI, PAUL WILFORD, AND YIN ZHANG

Abstract. An arbitrary resolution video coding method based on compressive sampling is proposed. In this method, a video is coded using compressive measurements. The compressive measurements are made on videos of high resolution. The measurements may be used to reconstruct the video at the same resolution as the original video, and any subset of the measurements can be used to reconstruct video at lower resolution with a lower complexity. Video coding with arbitrary solution has important application in mobile video transmission.

Key words. Arbitrary resolution video coding, scalable video coding, compressive sampling, total variation, TV-DCT method, Kronecker product, Walsh-Hadamard transform.

1. Introduction

In a video network, a video source may be transmitted to multiple clients with different characteristics. The clients in the video network have different channel capacities, different display resolutions, and different computing resources. For example, a video source may be transmitted through the network to a high performance computer with a high resolution monitor in a residential home, and at the same time, to a mobile device with a low resolution screen and with a battery powered CPU. It is therefore desirable for a video source to be encoded in such a way that the same encoded video stream can be transmitted, and be usable by all clients, of different characteristics, in the network. In other words, we want to encode the video source once, but to transmit the same encoded video at different channel rates, and to decode it at different resolutions and with different complexities.

The traditional video coding such as MPEG2 does not provide the scalability desired for todays video network as described above. The lack of scalability exhibits itself in at least two ways. First, an encoded video is not scalable with transmission channel capacity. Because of its fixed bit rate, an encoded video stream is unusable in a channel supporting a lower bit rate, and at the same time, suboptimal in a channel with higher bit rate. This is the cause of the cliff effect encountered in video broadcast or multicast. Second, the MPEG2 video is not scalable with decoder resolution or decoding complexity. An encoded video can be decoded only at one resolution, with a fixed complexity (not considering post-processing such as resizing, or enhancement, after decoding). This creates the need for multiple encoded streams of the same video content to target decoders of different resolutions.

Efforts have been made to introduce scalability into video coding, noticeably by the scalable video coding (SVC) of H.264 [1] and the wavelet transform of Motion JPEG 2000 [2]. Both methods encode video into ordered layers, or levels, of streams, and the resolution, or quality, of the decoded video increases progressively as higher layers, or levels, are added to the decoder. Hierarchical modulation [3] may be used in conjunction with these scalable video codes to achieve more bandwidth

Received by the editors March 9, 2015.
2000 *Mathematics Subject Classification.* 68U10, 94A08.

efficiency. For example, the high priority of hierarchical modulation can be used to carry lower layers of the encoded video, and the low priority of hierarchical modulation can be used to carry the higher layers of the encoded video. These efforts have provided some alleviation to the problems such as the cliff effect in video transmission using the traditional video coding, but challenges of mobile video broadcast still remain. There has been an abundance of research activities in video coding to provide scalability in decoding resolution, see [4]-[7]. A joint video coding and transmission method was proposed in [8] to provide scalability with transmission channel capacity. These activities are in response to the fact that the scalability provided by H.264 or Motion JPEG 2000 is still not satisfactory. Specifically, the ordered layer structure does not provide scalability at a fundamental level, because a video encoded in these standards needs to be decoded at the lowest layer, and progressively built up to higher layers. The loss of a lower layer in the transmission makes the higher layers useless, even when they are received error-free. Therefore, the ordered layer structure is not scalable with the channel capacity [8].

Due to the proliferation of compressive sampling techniques [9],[19], video coding using compressive measurements is rapidly emerging [10]-[11]. Compressive video sensing offers the scalability desired in video network [12]-[13], and it is suitable for wireless transmission [14]. When the measurements of a video are made by a random (or pseudo-random) matrix, the video source information is distributed among the measurements of equal significance, and there are no measurements that are more important than others. The reconstruction of video requires a certain number of measurements to be available, but it does not need the availability of a particular subset of measurements. In this sense, a lost measurement due to transmission can simply be replaced by any other measurement. Further, since a video does not have a well defined sparsity, statistically, the more measurements are used in reconstruction, the better the quality of the reconstructed video gets [15]. If the measurements of the video are transmitted in broadcast or multicast, a receiver in a channel with higher capacity can have more measurements available, and hence a reconstructed video of higher quality, than a receiver in a channel with a lower capacity. These properties illustrate that video coding using compressive sampling is inherently scalable with the channel capacity, and it avoids the cliff effect in broadcast and multicast.

In this paper, we propose a framework for video coding using compressive measurements in which an encoded video is scalable both with the channel capacity and with decoding resolution and decoding complexity. Under the framework, a high resolution video is encoded using compressive measurements. The measurements are made once on the high resolution video. Any subset of the measurements can be used to reconstruct a video of same resolution as the oringal, or a lower resolution. The implication of this is very powerful in wireless transmission. The measurements from the high resolution video are transmitted in wireless broadcast/multicast network. A client in a good channel can correctly receive enough measurements to reconstruct a video of the original resolution with acceptable quality. A client in a poor channel may only correctly receive a subset with measurements fewer than required to reconstruct an acceptable video at high resolution, but the client may still use the correctly received measurements to reconstruct a video of a lower resolution, with an acceptable quality. The ability of arbitrary resolution reconstruction makes this video coding suitable for transmission in all channels.

Furthermore, a client in the network may be a handheld device with a small display and powered by a battery. It is undesirable for such a device to reconstruct