Compressed Sensing and Its Applications in Video Processing

1 Introduction

Compressive sampling is an emerging bulk of work that deals with sub-Nyquist sampling of sparse signals of interest [1]-[3]. Rather than collecting an entire Nyquist ensemble of signal samples, CS can reconstruct sparse signals from a small number of linear measurements via convex optimization [4], linear regression [5], [6], or greedy recovery algorithms [7].

A somewhat extreme example of a CS application that has attracted much interest is the “single-pixel camera” architecture [8] where a still image can be produced from significantly fewer captured measurements than the number of desired/reconstructed image pixels. Arguably, a natural highly desirable next-step development is compressive video streaming.

2 Block-based Motion-aware Decoding of CS Videos

My research aims at the decoding of purely compressed-sensed videos sequences. In my major work of Ph.D. studies, I considered block-based CS video transmission systems. In such systems, the transmitter/encoder partitions each video frame into non-overlapping blocks, and performs nothing more than compressed sensing acquisition of each block independently, without the benefits of the familiar sophisticated forms of video encoding. Therefore, the burden of quality video sequence reconstruction falls solely on the receiver side. I developed block-adaptive Karhunen-Loève (KL) bases for sparse representation [9]-[11]. In a sliding-window decoding approach, the KL basis $\Psi_{m,KL}^t$ for block $X_m^t$ is estimated from the previously reconstructed frames $\hat{F}_k, k = t-1, t-2, ..., t-n$, where $n$ is the number of reference frames. Experimental results demonstrate that the proposed decoder outperforms significantly the conventional fixed basis intra-frame [12] and inter-frame [13], as well as the K-SVD [14], decoders. Performance is improved as the number of reference frames (what we call “decoder order”) increases, with order values in the range of two to ten appearing as a good compromise between computational complexity and reconstruction quality.
3 Frame-wise Decoding of CS Videos

Later, I investigated the decoding of frame-wise compressed-sensed video via total variation (TV) minimization. TV minimization has been widely used as an image de-noising algorithm [15], [16]. Based on the principle that signals with excessive and possibly spurious detail have high TV, reducing the TV of the signal subject to it being a close match to the original signal removes unwanted detail whilst preserving important details such as edges. After 2-dimensional TV minimization algorithm was successfully used in CS image recovery, researchers started to consider the multi-frame compressed-sensing video encoding with 3-dimensional (3D) TV based reconstructions [17]. Though promising, such a system requires complex and expensive spatial-temporal light modulators that make the techniques difficult to realize in practice. In contrast, I proposed a frame-by-frame CS video encoder that performs intra-frame encoding, and inter-frame similarities are exploited at the decoder via spatial-temporal TV minimization [18]. The advantage of such a system is that it can be implemented practically with existing CS imaging system such as the single-pixel camera, while the inter-frame decoder outperforms an intra-frame decoder with 2D TV minimization.

Nevertheless, the above mentioned TV based CS video decoders have not fully exploited motion estimation (ME), a defining matter in the feasibility and success of digital video. To take motion information into account in frame-wise CS video systems, [19] proposed for the first time a motion-adaptive spatial-temporal sparse regularization method that minimizes the $\ell_1$-norm of both the 2D-DWT coefficients of the video frames and the motion-compensated frame residuals. This motion adaptation idea is later extended for accelerated dynamic MRI [20]. Although motion-compensation greatly improved video sparsity, hence the reconstruction peak signal-to-noise-ratio (PSNR), using 2D-DWT as the spatial sparse transform is inefficient. The method proposed in [21] minimizes the motion-compensated temporal TV to reconstruct dynamic MRI images. Although motion information is accounted for, this method does not exploit spatial sparsity. Recently, I have started working on “Motion-adaptive TV Minimization for CS Video Reconstruction”. In this topic, TV-based sparsity and motion-compensated sparse frame residuals are combined to promote video signal sparsity, hence the reconstruction performance is greatly improved with the same CS sampling ratio.
4 Other Topics of CS Video

I have also worked on “Rate-adaptive Compressed Sensing for Video Coding” [22], [23], “CS based Depth Map Coding” [24], and “Rate-distortion Minimization for CS Video Systems” [25].

References


