Research Article

Adaptive Multihypothesis Prediction Algorithm for Distributed Compressive Video Sensing

Jinxiu Zhu, 1,2 Ning Cao, 3 and Yu Meng 1,2

1 College of Computer and Information, Hohai University, Changzhou 213022, China
2 Changzhou Key Laboratory of Sensor Networks and Environmental Perception, Changzhou 213022, China
3 College of Computer and Information, Hohai University, Nanjing 210098, China

Correspondence should be addressed to Jinxiu Zhu; zhujinxiu1972@163.com

Received 20 February 2013; Accepted 23 April 2013

Academic Editor: Liangmin Wang

Copyright © 2013 Jinxiu Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel adaptive multihypothesis (MH) prediction algorithm for distributed compressive video sensing (DCVS) is proposed in this paper. In the proposed framework, consistent block-based random measurement for each video frame is adopted at the encoder independently. Meanwhile, a mode decision algorithm is applied in CS-blocks via block-based correlation measurements at the decoder. The inter-frame MH mode is selected for the current block wherein the interframe correlation coefficient value exceeds a predetermined threshold. Otherwise, the intraframe MH mode is worthwhile to be selected. Moreover, the adaptive search window and cross-diamond search algorithms on measurement domain are also incorporated to form the dictionary for MH prediction. Both the temporal and spatial correlations in video signals are exploited to enhance CS recovery to satisfy the best linear combination of hypotheses. The simulation results show that the proposed framework can provide better reconstruction quality than the framework using original MH prediction algorithm, and for sequences with slow motion and relatively simple scene composition, the proposed method shows significant performance gains at low measurement subrate.

1. Introduction

In the wireless video sensor network, wireless video cameras are widely used to timely operate and store the video data which is waiting for the later upload to a fixed network base station. However, in such a network, the resources, for example, the computing and the storage capacity, are more limited at the encoder than the decoder. To address this issue, an advanced video coding scheme, referred to distributed video coding (DVC) [1], is presented to satisfy the low-complexity capability at the encoder. Its main idea is founded on the principle of distributed source coding (DSC) with a number of separate encoders and one joint decoder. However, it still suffers from the disaster where huge amounts of raw image data are captured at the decoder with the memory-intensive consumption, even for still image. Fortunately, with the development of the single-pixel camera architectures [2], the underlying compressed sensing (CS) theory seems to be a key approach to mitigate such a problem [3]. As an effective data compression method, CS enables to directly achieve the compressed data via a random projection on the raw image data. Recently, distributed compressive video sensing (DCVS) [4] is proposed to flexibly utilize the CS in the framework of the DVC. The compressed video data is firstly obtained at a low complexity encoder. Then the CS reconstruction is performed with the correlation exploitation among successive frames at a high-complexity decoder [5–8].

Motivated by recent progress in the distributed video coding, we propose a novel distributed compressive video sensing framework, wherein several stopping criteria are used to improve and speed up the multihypothesis prediction algorithm for the CS frame reconstruction by the dual-discrete wavelet transform (DDWT) [9] basis. The remainder of the paper is organized as follows. In Section 2, the DCVS and MH prediction are reviewed for the block-based CS. In Section 3, the DVC framework based on adaptive MH prediction is introduced. The experimental results are presented in Section 4. Finally, relevant conclusions and some remarks are given in Section 5.
2. Related Works

2.1. Distributed Compressed Video Sensing. To the best of our knowledge, the DCVS framework mainly can be divided into two categories according to the original video data reception at the encoder. At the encoder of DCVS, video frames are grouped into group of pictures (GOP) consisting of a key frame (also called K-frame) and a number of nonkey frames (also called CS-frames). In the scheme proposed in [4], the encoder obtains the original information, key frames are encoded using traditional MPEG/H.264 encoding, while for CS frames, measurements are taken via random measurement matrix. Side information is generated from the neighboring reconstructed K-frames. The disadvantage of this framework is that the complex MPEG/H.264 encoding is still required. On the other hand, in [10] where the encoder gets the measurements directly but not the original information, their approach is different from the scheme proposed in [4] in which CS measurements are applied to both key and nonkey frames. K-frames are reconstructed using GPSR [11] while stopping criteria based on side information generated from the K-frames are used during the reconstruction process of the CS-frames.

However, there still exit a few disadvantages in the aforementioned methods. For example, the generation rule of side information (SI) is usually simple due to releasing the computation burden of the coders. The CS reconstruction process also cannot perform effectively with the rough prediction. As a result, the performance of the DCVS cannot achieve the best. Thereby, we focus on a novel framework where several criterions are proposed and also use adaptive search window to improve the traditional MH prediction algorithm for the better CS-frame reconstruction.

2.2. Multihypothesis Prediction for Block-Based CS of Video. To alleviate the huge computation and memory burden for both the sensing and reconstructing processes, images are usually decomposed into small blocks for the further process. An approach for block-based CS (BCS) for 2D images is proposed with the assumption of the independence among blocks in [12]. Thereby, from the perspective of the incoherence principle in CS theory, block-based measurements seem to be less efficient than frame-based measurements due to that the former sensing matrix is block diagonal and the latter one is dense. Nonetheless, by the lower request of incoherence, BCS can preserve local information that helps the decoder construct more accurate SI based on the interframe sparsity model with the sparsity-constraint block prediction. Thereby, the work in [13] explores the sparsity of small interframe difference to remove the temporal redundancy, though it is not suited for video sequences with large interframe difference and fast motion. Later, the CS reconstruction of video is presented with an alternative way [14]. An explicit prediction using motion estimation (ME) and motion compensation (MC) is utilized to overcome the above problem. Another ME/MC-based reconstruction is also proposed with the block-based measurement of a CS-frame to form a block-by-block MH motion-compensated prediction [15]. Recently, an improved strategy for incorporating MH prediction into the block-based compressed sensing with smooth projected Landweber (MH-BCS-SPL) reconstruction of video is proposed [16] which could get a more accurate prediction by finding a linear combination of all the blocks/hypothesizes in the search window.

2.3. Our Contributions. Different from current MH prediction approaches which only focus on these modified MH regularization algorithms, our approach combines both MH prediction mode and dictionary acquisition in the aforementioned DCVS schemes. Our main contribution in this paper is as follows.

(i) A novel block mode decision at the decoder is performed for blocks in CS-frame. In particular, the inter- and intramodes are adaptive used based on the correlation of CS measurements in order to obtain better reconstruction quality.

(ii) Cross-diamond search algorithm on measurement domain is adopted to build the dictionary in Inter-MH prediction mode, which provides much sparser representation for the corresponding blocks.

(iii) A practical and real-time system is designed for the DCVS via the above adaptive MH prediction.

3. Adaptive Multihypothesis Prediction Algorithm of Our DCVS

3.1. Proposed DCVS Scheme. The proposed DVCS framework based on the adaptive MH prediction (AMH_DVCS) is described in this section. As illustrated in Figure 1, at the encoder, the frames of a video sequence are divided into two categories: K-frames and CS-frames. To simplify the encode framework, in both K-frames and CS-frames the consistent block-based random measurements are adopted. Sequentially, the measurements are transmitted to the decoder with their corresponding nearby integers. Hereby, the sparse basis matrix \( \Psi \) is used as the DDWT basis. Without loss of generality, the subrate \( (S_r) \) is set higher in K-frames than in CS-frames \( (S_{CS}) \), due to the K-frames usually reconstructed with more advanced quality. In the section of the decoder, a mode decision algorithm is formed based on the block correlation of CS measurements. The blocks in CS-frame (CS-blocks) are classified into two types, that is, inter-frame MH prediction modes (Inter\_MH mode) and intra-frame MH prediction modes (Intra\_MH mode). Then the bitstream of measurements is processed with the selected decoding mode. On one hand, the K-frame is reconstructed by intra-frame reconstruction. On the other hand, the SI frame is generated by the adaptive reconstruction for the CS frame. Thus, residual frames are sparse recovered with SI frames, and the further decoded CS-frames are produced. The more detailed process is given in Section 3.2.

3.2. Adaptive Multihypothesis Prediction for CS-Frame Reconstruction

3.2.1. Decoding Mode Decision Based on the Correlation of CS Measurements. In our work, the adaptive decoding mode
is adopted to exploit the temporal and spatial correlation of video sequences. For convenience, some attributions of parameters in the DCVS are first discussed. Generally, CS measurements can be modeled as random Gaussian sources. And the dependence between two random variable quantities is indicated by Pearson’s correlation coefficient [17]. The work in [18] explores that the frames in various video sequences have the high correlation among CS measurements with the corresponding value even above 0.9. Therefore, we define the correlation coefficient (CC) function of CS measurements as

\[ r(y_1, y_2) = \frac{\sum_{i=1}^{N} (y_1(i) - \bar{y}_1) \cdot (y_2(i) - \bar{y}_2)}{\sqrt{\sum_{i=1}^{N} (y_1(i) - \bar{y}_1)^2} \cdot \sqrt{\sum_{i=1}^{N} (y_2(i) - \bar{y}_2)^2}}, \]

where \( y_1 \) and \( y_2 \) are the different measurement vectors of block, \( N \) is the length of a measurement vector. Then, the mode decision procedure is described as follows.

**Step 1.** Calculate the inter-frame CC function \( r(y_{BS}^K, y_{BR}^K) \), where \( y_{BS}^K \) and \( y_{BR}^K \) are the measurement vectors of current coding CS-block and nearby key frames, respectively.

**Step 2.** Given a predetermined threshold \( T_1 \), the Inter_MH mode is selected for the current block, if \( r(y_{BS}^K, y_{BR}^K) > T_1 \), and the Intra_MH mode is done otherwise. That is,

\[
B_{\text{mode}} = \begin{cases} 
\text{Inter}_\text{MH} & r\left(y_{BS}^K, y_{BR}^K\right) > T_1, \\
\text{Intra}_\text{MH} & r\left(y_{BS}^K, y_{BR}^K\right) \leq T_1.
\end{cases}
\]

**3.2.2. Multi hypothesis Prediction Algorithm.** MH prediction has a major distinctive advantage wherein multiple predictions are used to yield a composite prediction, compared with the traditional simple single-hypothesis (SH) prediction. For this reason, various improved MH prediction methods are developed and widely incorporated in the recent video coding to enhance the video-coding quality. These methods impose specific structures on the hypotheses to form the more strict ultimate prediction with the target of the additional motion-vector rate constraint by multiple predictions of a block. Moreover, because the MH predictions are all performed at the decoder, without the corresponding rate burden, the more complex but better forms of MH prediction can be considered. The major work of ME/MC in residual reconstruction is to create an MH predictive block with the windows, whose distance is as close as possible to original block, in given reference frames. The optimal sparse coefficient \( \hat{w}_{i,j} \) can be cast as

\[ \hat{w}_{i,j} = \arg \min_{w} \| x_{i,j} - H_{i,j}w \|_2^2, \]

where \( x_{i,j} \) is the \( i \)-th block for the \( t \)-th frame and \( H_{i,j} \) is a matrix whose columns are the rasterizations of the possible blocks within the search space of the reference frames within the search window. And the MH predictive block \( \tilde{x}_{i,j} \) is constructed as

\[ \tilde{x}_{i,j} = H_{i,j} \hat{w}_{i,j}. \]

However, in the encoder, we can only receive the measurements \( y_{i,j} = \Phi x_{i,j} \). Thereby, in the CS framework, the estimated sparse coefficient \( \hat{w}_{i,j} \) can be obtained as

\[ \hat{w}_{i,j} = \arg \min_{w} \| \Phi (x_{i,j} - H_{i,j}w) \|_2^2 = \arg \min_{w} \| y_{i,j} - \Phi H_{i,j}w \|_2^2. \]

The most common method to solve the least-squares (LSQ) problem is Tikhonov regularization [19] which imposes a \( l_2 \) penalty on the norm of \( \hat{w}_{i,j} \) as

\[ \hat{w}_{i,j} = \arg \min_{w} \| y_{i,j} - \Phi H_{i,j}w \|_2^2 + \lambda^2 \| \Gamma w \|_2^2, \]
International Journal of Distributed Sensor Networks

The Inter_MH mode assumes a (vectorized) block in a CS-frame can be represented weightily by a linear combination of (vectorized) temporal blocks within the window centered in the position of the best matching block in preceding and following key frames.

![Diagram of Inter_MH mode](image)

**Figure 2:** The Inter_MH mode assumes a (vectorized) block in a CS-frame can be represented weightily by a linear combination of (vectorized) temporal blocks within the window centered in the position of the best matching block in preceding and following key frames.

\[ y_{t,j} = \Phi \cdot \tilde{x}_{t,j} = \Phi \cdot \Phi^T \begin{bmatrix} H_{t,j} \end{bmatrix} W \]

where \( \Gamma \) is known as the Tikhonov matrix, and \( \lambda \) is a scale factor. For each block, \( \tilde{w}_{t,j} \) can be calculated directly by the Tikhonov solution as

\[ \tilde{w}_{t,j} = \left( \left( \Phi H_{t,j} \right)^T \left( \Phi H_{t,j} \right) + \lambda^2 \Gamma^T \Gamma \right)^{-1} \left( \Phi H_{t,j} \right)^T y_{t,j}. \]  

By taking (7) into (4), the prediction \( \tilde{x}_{t,j} \) can be finally achieved as

\[ \tilde{x}_{t,j} = H_{t,j} \tilde{w}_{t,j}. \]  

Finally, all the predicted blocks \( \tilde{x}_{t,j} \) are put together providing the SI frame.

3.2.3. Inter_MH Mode Principles. The Inter_MH mode is shown in Figure 2. Block-based measurements, along with preceding and following key frames, are used to generate MH block prediction. As for Inter_MH mode, we use a dictionary of temporal blocks in the adjacent key frames as the matrix \( H_{t,j} \). Different from the original MH-BCS-SPL reconstruction [16] with the fixed search window size, \( W = 6 \), and the window centered in the position of current CS-block, our proposed method uses the cross-diamond search algorithm (CDS) in [20] for the fast block motion estimation to calculate the maximum correlation coefficient (MCC) of CS measurements. In the CDS algorithm, an initial cross-shaped pattern (CSP), as shown in Figure 3(a), is proposed as the initial step, and then a large diamond-shaped pattern (LDSP) and small diamond-shaped pattern (SDSP) are

![Search patterns](image)

**Figure 3:** Search patterns used in the CDS algorithm. (a) CSP. (b) LDSP and SDSP.
Figure 5: The process of sparse reconstruction with decoder SI.

Figure 6: Relationship between the search window size $W$ and the correlation of original CS-frame and its SI predicted by MH-BCS-SPL algorithm on the 50th frame of four video sequences.

... used, as depicted in Figure 3(b). The best matching block is obtained by computing MCC between the current CS block and the blocks within a search window while the window size increases from 2 to 14. The process of the window match is performed until the incremental MCC is not larger than a threshold, for example, 0.005. Then the block at the current search window is viewed as the best matching block. The search window is a square region which centered on the position of the best matching block. Because the best matching block has the maximum correlation with the current block in CS frame, the generated prediction is sufficiently accurate for small search window size, so we set the window size $W$ to be 3 for lower computation.

In the Inter_MH mode, it assumes the atoms in the dictionary are composed by a set of linear combinations of temporal blocks which are also called MH predictions
within the window centered on the position of the best matching block in the adjacent key frames available at the decoder. Given its corresponding compressed measurements, our method can predict the block optimally to generate the SI for a CS block by using (7) and (8).

3.2.4. \textit{Intra\_MH Mode Principles}. The Intra\_MH mode is shown in Figure 4. Just as in the Inter\_MH mode, the block match is also used. The only difference is that the dictionary is composed by spatial blocks in CS-frame. The best matching block is obtained by computing MCC between the current CS block and the blocks within a search window in CS-frame itself with the size \( W \) increasing from 2 to 11. If the difference of MCC between the current and next sizes is not larger than 0.00001, the current size is selected as the size of window where blocks are used as the atoms of the dictionary. Similarly, given its compressed measurements, this mode enables the block to be optimally predicted to generate the SI for a CS block.

3.2.5. \textit{Sparse Reconstruction with Decoder SI}. With the aforementioned adaptive decoding modes for each CS block, all the recovered CS-blocks are put together which generates the SI frame. AMH_DVCS employs a very simple but effective algorithm to realize the sparse reconstruction with the SI frame; that is, subtract the measurement vector of the current frame from the measurement vector of the SI frame to form a new measurement vector of the prediction error. When the prediction is sufficiently precise, it can be faithfully recovered from its compressed measurements. The approximation of an input frame is then simply recovered by adding the prediction error to the SI frame. The process of sparse representation with decoder SI is summarized in Figure 5. The BCS-SPL algorithm is used in the algorithm of sparse representation with the decoder SI.
Figure 8: Performance comparison between the proposed MH-BCS-SPL and the original BCS-SPL: average reconstruction quality of the first 50 frames of the four sequences.

Figure 9: Reconstruction frame 50 of the Mother-daughter sequence with $S_{CS} = 0.3$ and $S_K = 0.7$. (a) Original image. (b) Frame reconstructed using the original MH-BCS-SPL with PSNR = 38.1821 dB. (c) Frame reconstructed using the proposed MH-BCS-SPL with PSNR = 39.9817 dB.
4. Simulation Results

The performance of our proposed framework is tested on four QCIF video sequences, that is, Foreman, Salesman, Mother-daughter, and Football, with GOP = 2 and the block size B = 16. Several subrates are also employed to evaluate the proposed method with the baseline of the original MH-BCS-SPL algorithm.

4.1. Results of the Correlation with Different Search Window Size. Figure 6 shows the relationship between the search window size $W$ and the average correlation of the blocks between original CS-frame and its SI frame predicted by the original MH-BCS-SPL algorithm for 50th frame of four video sequences. From Figure 6, we can see that at first the correlation increases with the search window size growing, which means a greater $W$ can provide a better reconstruction quality. But when the $W$ exceeds a suitable size, the correlation increases slower or decreases as the $W$ growing. We know that the larger the $W$ is, the more complex the reconstruction is. Thus, we make use of the adaptive search window size in order to trade off between the reconstruction quality and complexity.

4.2. Results for the Decoding Mode Selection. As long as the decoding mode is selected, the reconstruction quality is relatively insensitive to $T_1$. From our simulations, $T_1 = 0.999$ achieves the maximum reconstruction quality and then is used in our paper. The ratio between the Inter-MH mode and the Intra-MH mode of blocks in the first 50 frames for four sequences is provided in Figure 7. Figure 7 shows that for the sequence with little motion, most blocks select the Inter-MH mode, while most blocks select Intra-MH mode for the sequence with fast or complex motion. In the Football sequence, the motion of the first 38 frames is intense while in the rest frames the motion is little and simple, so from Figure 7(d) we can see that the curve is intersecting in the 39th frame.

4.3. Performance Comparison with Original MH-BCS-SPL Algorithm. The average PSNR performances with different subrates for the four sequences using the proposed algorithm and the original MH-BCS-SPL algorithm are shown in Figure 8. The numerical values on the x-axis denote the subrates of the CS-frames with a fixed K-frame subrate $S_K = 0.7$ while those on the y-axis represent the average reconstruction quality (PSNR in dB) of CS-frames.

As can be seen in Figure 8, the proposed adaptive MH-BCS-SPL provides superior reconstruction quality over the original MH-BCS-SPL across the range of tested subrates. It can also be seen that for sequences with slow motion and relatively simple scene composition, such as the Mother-daughter sequence, the proposed method shows significant performance gains, while the gap between adaptive MH-BCS-SPL and MH-BCS-SPL narrows as the subrate increases. However, for the Football sequence with fast or complex motion, the performance gains are not substantial, while the gap between adaptive MH-BCS-SPL and MH-BCS-SPL broadens as the subrate increases.

We also compare the visual quality of the reconstruction results. Figures 9 and 10 show examples for the reconstructed frame 50 of the Mother-daughter sequence and Football sequence with $S_{CS} = 0.3$ and $S_K = 0.7$. These also demonstrate better performance of our adaptive framework. Thus, the proposed method gives better subjective and objective reconstruction quality. Compared to the original MH-BCS-SPL scheme, our method explores both the temporal and spatial correlation not only between neighboring frames but also between neighboring blocks in CS-frame itself to build adaptive dictionary for more spare description and, therefore, is able to achieve higher quality.

5. Conclusion

In this paper, a new distributed compressive video sensing framework based on adaptive MH prediction is proposed to simultaneously capture and compress videos at the low-complexity encoder and efficiently reconstruct videos at the decoder. The proposed framework can estimate the inter-frame correlation between colocated blocks in neighbor frames based on CS measurements to further divide these blocks into two types. We exploit both the temporal and spatial correlation between neighbor frames and blocks in CS-frame. To enhance CS recovery, the adaptive MH
predictions are developed to find the best linear combination of hypotheses. Our simulation results demonstrate that the proposed framework can provide better reconstruction quality than the original MH-BCP-SPL algorithm. Additional considerable gain, approximately 0.5–3.6 dB, in the average PSNR can be achieved compared with the prior works.

References

[1] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed video coding,” Proceedings of the IEEE, vol. 93, no. 1, pp. 71–83, 2005.

[2] M. F. Duarte, M. A. Davenport, D. Takbar et al., “Single-pixel imaging via compressive sampling: building simpler, smaller, and less-expensive digital cameras,” IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 83–91, 2008.

[3] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” IEEE Transactions on Information Theory, vol. 52, pp. 489–509, 2006.

[4] P. N. Josep, Y. Ma, and T. Huang, “Distributed video coding using compressive sampling,” in Proceedings of the Picture Coding Symposium (PCS ’09), May 2009.

[5] L. W. Kang and C. S. Lu, “Distributed compressive video sensing,” in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’09), pp. 1169–1172, April 2009.

[6] V. Stanković, L. Stanković, and S. Cheng, “Compressive image sampling with side information,” in Proceedings of the IEEE International Conference on Image Processing (ICIP ’09), pp. 3037–3040, November 2009.

[7] H. W. Chen, L. W. Kang, and C. S. Lu, “Dynamic measurement rate allocation for distributed compressive video sensing,” in Visual Communications and Image Processing. Proceedings of the SPIE, July 2010.

[8] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, “Distributed compressed video sensing,” in Proceedings of the IEEE International Conference on Image Processing (ICIP ’09), pp. 1393–1396, November 2009.

[9] N. Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” Applied and Computational Harmonic Analysis, vol. 10, no. 3, pp. 234–253, 2001.

[10] L. W. Kang and C. S. Lu, “Distributed compressive video sensing,” in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1169–1172, Taipei, Taiwan, April 2009.

[11] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, “Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems,” IEEE Journal on Selected Topics in Signal Processing, vol. 1, no. 4, pp. 586–597, 2007.

[12] L. Gan, “Block compressed sensing of natural images,” in Proceedings of the International Conference on Digital Signal Processing, pp. 403–406, Cardiff, UK, July 2007.

[13] J. Zheng and E. L. Jacobs, “Video compressive sensing using spatial domain sparsity,” Optical Engineering, vol. 48, no. 8, pp. 1–10, 2009.

[14] H. Jung and J. C. Ye, “Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: what we can learn from video compression techniques,” International Journal of Imaging Systems and Technology, vol. 20, no. 2, pp. 81–98, 2010.
Submit your manuscripts at http://www.hindawi.com