A New Multihypothesis-Based Compressed Video Sensing Reconstruction System

Shuai Zheng, Jian Chen, Member, IEEE, Xiao-Ping Zhang, Senior Member, IEEE, and Yonghong Kuo

Abstract—Multihypothesis-based compressed video sensing scheme attracts wide attention in the research of resource-constrained video application scenarios. However, high-accuracy weight prediction of hypotheses is always challenging especially for the high-motion sequences. To solve this problem, this paper proposes a novel multihypothesis-based distributed compressed video sensing (NMH-DCVS) system. The new multihypothesis system contains two components: hypotheses acquisition, and weight prediction. First, to acquire more high-quality hypotheses, a new hypotheses acquisition scheme is proposed by constructing the search window based on the temporal, and spatial correlation of the image blocks, respectively. The optimal matching block can be quickly determined. Second, to improve the accuracy of the multihypothesis weight prediction, a new residual transforming preprocessing-based weight prediction algorithm is proposed by transforming the original hypothesis set to residual hypothesis set. The influence of the quality fluctuation of the hypotheses on prediction accuracy is effectively suppressed. Moreover, the improved hypotheses further improve the sparsity of the residual hypothesis set, leading to the additional improvement of the accuracy of the proposed residual-based weight prediction algorithm. Experiment results show that compared with the state-of-the-art methods reported in the literature, the proposed new multihypothesis system significantly improves the decoding performance both in objective, and subjective quality.

Index Terms—Compressed sensing, hypotheses acquisition, residual transforming, weight prediction.

I. INTRODUCTION

With the development of wireless communication technology especially the arrival of 5G, the Internet of things (IoT), wireless multimedia sensor networks (WMSN), and mobile video terminals will be further developed significantly. More and more new video application scenarios are showing up. Video-based services such as the mobile video live show are believed to be the major driving force for future network traffic [1]–[3]. The huge video data services require high transmission bandwidth and high computing ability. However, in these new video application scenarios, the limited resources such as the battery, computing ability, and storage space are conflicted with these requirements. Realizing the efficient encoding and transmission of video data in resource-constrained scenarios has become a significant challenge [4]–[6].

In traditional video encoding standards, complex computation is performed at the encoder, such as motion estimation, motion compensation, and coding mode prediction processing [7]–[10]. It causes high computation and resource consumption pressure to the encoder in the resource-constrained video application scenes. To solve this problem, new imaging systems and sampling technologies are constantly proposed [11]–[14]. Compressed sensing (CS) [15]–[17] technology provides an effective way to reduce the signal sampling rate and improve the encoding efficiency. It samples and compresses a signal simultaneously at a sub-Nyquist rate by multiplying the signal with a measurement matrix [18]. CS theory points that a signal can be recovered from a small number of measurements by solving an optimization problem if the signal is sparse in its original domain or a transform domain [19]. Block-based CS sampling (BCS) [20] strategy further improves the encoding efficiency of the video encoder. CS-based signal processing has attracted wide attention in many research areas [21]–[27]. Distributed video coding (DVC) [28], [29] technology moves the complex computation from encoder to decoder by independent encoding and joint decoding. Distributed compressed video sensing (DCVS) systems [30] provide a feasible solution to further improve the encoding efficiency by combining the DVC and CS technologies. DCVS systems significantly reduce the computing pressure of the encoder for processing the massive video data [31]–[33].

Current research on DCVS mainly focuses on the study of the reconstruction scheme at the decoder. Although many reconstruction algorithms have been developed, such as the approximate message passing (AMP) algorithm [34], the greedy pursuit algorithms [35], [36], the fast gradient projection sparse reconstruction (GPSR) algorithm [37], and total variation (TV) based CS recovery algorithm [38]–[40], the recovery quality of DCVS system is unsatisfactory compared to the traditional video.
coding standards [41]–[43]. In [44]–[46], a multihypothesis-based block CS smooth projection Landweber (MH-BCS-SPL) reconstruction scheme is proposed by Fowler et al. Multihypothesis (MH) prediction is first introduced in the DCVS recovery scheme and achieves significant success in improving the CS reconstruction quality. In MH scheme, the side information (SI), a prediction of the target image block, is first obtained by a weighted sum of a group of hypotheses. The recovery of the original block is transformed into the recovery of the residual block. The recovery performance is improved due to the residual block is more sparse [45]. Due to the good performance of MH, it has been widely used in CS-based multimedia signal recovery research [47]–[50].

For the performance of MH prediction, the acquisition of hypotheses and the design of the weight prediction model are two critical factors. For hypotheses acquisition, in traditional MH schemes [44]–[47], the source of hypotheses is limited. Only the key frame is utilized as the reference frame. Hypotheses are acquired by constructing a fixed search window centered on the location of the block to be reconstructed. The quality of hypotheses fluctuates significantly due to the complex video motion states. To extend the hypotheses source, a hypotheses interpolation method is proposed in [52]. New hypotheses are obtained by high-quality hypotheses interpolation method and are utilized to replace the low-quality hypotheses in the original hypothesis set. It works well for low-motion sequences. In [53], [54], a secondary matching processing for the matching block prediction in reference frame is proposed to get more high-quality hypotheses. However, the adjacent non-key frame with better inter-frame correlation than the key frame is ignored in the above work. With the increase of the inter-frame distance between the current frame and the key frame, the quality of the hypotheses degrades significantly especially for high-motion sequences.

In [51], the adjacent non-key frames are first utilized as additional reference frames to provide more hypotheses. It makes up the impact of the reduced inter-frame correlation between the non-key frame and key reference frame to a certain extent. However, the hypotheses search window is still constructed in a fixed position. The hypotheses acquisition in each reference frame is independent. The motion information obtained from the recovery of previous blocks is discarded. Fixed hypotheses search window cannot adapt to the change of the motion states of image blocks. Therefore, on the basis of multi-reference frames, how to jointly and adaptively construct the search window in the reference frames to get high-quality hypotheses is still a challenge. Especially for high-motion sequences, on one hand, complex motion states make the accurate prediction of the optimal matching block in the reference frame difficult; on the other hand, the repeated matching block prediction will also increase the computation complexity significantly.

For MH weight prediction, the accuracy of the Tikhonov-based weight prediction model used in the traditional MH schemes degrades with the decrease of the sampling rate. To solve this problem, an elastic net-based MH prediction model (MH-wNet) is proposed in [55]. It combines the $l_1$ and $l_2$ norms of the weight vector into the MH model. The accuracy of weight prediction under a low sampling rate is improved effectively. However, with the increase of the sampling rate, the computational complexity increases significantly and the improvement of the prediction accuracy cannot keep up with traditional Tikhonov-based model. In [56] and [57], the sparsity regularization of the predicted block is incorporated into the new weight prediction model. A good improvement of the recovery quality is obtained for low-motion sequences. However, for high-motion sequences, the added sparsity regularization of the predicted block has little effect on the improvement of reconstruction quality. Moreover, the above algorithms are all based on the measurement domain. The blocks being predicted have to be divided in a non-overlapping manner under a fixed block size that is entirely decided by the encoder. It restricts the acquisition of the side information of the decoder and is not good for eliminating the block effects of the recovered frames.

To overcome the limitation of the prediction algorithms based on only the measurement domain, a two-stage MH prediction scheme is proposed in [58], [59]. It executes the MH prediction procedure in the measurement domain and pixel domain, respectively. The decoding delay is increased significantly with a little improvement of the reconstruction quality in the two-stage scheme. For the above MH weight prediction schemes, they are all based on a high-quality hypothesis set. Once the hypotheses degrade, the performance of the above sparsity regularization based models will also degrade. Therefore, the degradation of the recovery quality always occurs for high-motion sequences. The prediction accuracy cannot be guaranteed for high-motion sequences when hypotheses quality fluctuates. It is a challenge to improve the accuracy of MH weight prediction effectively for sequences with complex motion states.

In this paper, to improve the accuracy of MH prediction and mitigate the quality degradation for high-motion sequences, we propose a new MH-based DCVS (NMH-DCVS) system. New hypotheses acquisition method and hypotheses weight prediction algorithm are proposed in the new MH prediction scheme. First, to effectively get high-quality hypotheses from high-motion sequences, a temporal and spatial correlation-based multihypothesis (TS-MH) acquisition scheme is proposed. The optimal matching block and search window in the reference frames can be quickly and adaptively determined in the newly proposed MH acquisition method. The hypotheses quality for high-motion sequences is improved effectively. Second, for weight prediction, to improve the weight prediction accuracy of high-motion sequences, a new weight prediction algorithm based on residual transforming of hypotheses is proposed. We propose a preprocessing program of residual transforming for the obtained hypothesis set. The signal sparsity is improved effectively in the weight prediction. The degradation of the prediction accuracy caused by the fluctuation of hypotheses quality is mitigated. Moreover, with the improvement of the hypotheses quality in acquisition stage, the residual block is more compressible in the subsequent residual transforming processing. Correspondingly, a further improvement of the whole MH prediction scheme and the final recovery quality is obtained. In the experiments, the accuracy of the weight prediction is improved.
steadily. A good improvement both in objective and subjective quality for all kinds of sequences is obtained.

The rest of this paper is organized as follows. Section II first gives the problem in current MH-based compressed sensing video schemes. Then, the proposed new MH acquisition and weight prediction system and its specific implementation details are illustrated in Section III. The experimental results are given in Section IV. Finally, Section V makes a conclusion for this paper.

II. PROBLEM FORMULATION

Fig. 1 shows the block diagram of the common MH-based DCVS system. We will study the MH processing procedure represented by the modules with blue shadow. The hypotheses acquisition and the weight prediction of the final hypothesis set are the key to MH processing. The mathematical notations used in the following sections are summarized in Table I.

At the encoder, video sequences are processed in the unit of group of pictures (GOP). The frames are divided into key frame and non-key frames and sampled by block-based CS sampling. Let \( x_{\text{cur}} \in \mathbb{R}^{N \times 1} \) represent a vectorized image block to processed. The measurements \( y_{\text{cur}} \in \mathbb{R}^{M \times 1} \) obtained by BCS with the measurement matrix \( \Phi \in \mathbb{R}^{M \times N} \) is shown as below

\[
y_{\text{cur}} = \Phi x_{\text{cur}},
\]

At the decoder, key frames are first decoded. Then the MH system is performed to provide side information for the recovery of non-key frames. The reference frame is first selected. After getting the optimal matching block \( x_{\text{ref,opt}} \) in the reference frame of current block \( x_{\text{cur}} \), the search window is constructed in the reference frame centered on \( x_{\text{ref,opt}} \).

For the hypotheses acquisition scheme in traditional MH systems, the search window is always constructed in the key frame centered on the block that has the same position as the current block. It ignores the impact of complex inter-frame motion. With the increase of the inter-frame distance between the non-key frame and the key frame, the traditional search window cannot cover the optimal hypotheses acquisition range especially for high-motion sequences. The optimal matching block is usually located outside the search window due to the complex motion of the target block. For current multi-reference frames-based MH acquisition methods, the search window is constructed in different reference frames, independently. Repeated matching block prediction processing increases the computational complexity.

Table I: Mathematical Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_{\text{cur}} )</td>
<td>the current image block to be processed</td>
</tr>
<tr>
<td>( \Phi )</td>
<td>the measurement matrix</td>
</tr>
<tr>
<td>( y_{\text{cur}} )</td>
<td>the measurements vector of ( x_{\text{cur}} )</td>
</tr>
<tr>
<td>( N )</td>
<td>the number of pixels for each block</td>
</tr>
<tr>
<td>( M )</td>
<td>the number of measurements after BCS sampling</td>
</tr>
<tr>
<td>( h_i )</td>
<td>the ( i )th hypothesis for current block</td>
</tr>
<tr>
<td>( H )</td>
<td>the hypothesis set for current block</td>
</tr>
<tr>
<td>( w_{\text{cur}}^{\text{ref}} )</td>
<td>the predicted weight vector for ( x_{\text{cur}} )</td>
</tr>
<tr>
<td>( x_{\text{ref}} )</td>
<td>the reference block of ( x_{\text{cur}} )</td>
</tr>
<tr>
<td>( x_{\text{cur}}^* )</td>
<td>the block in reference frame with same position as ( x_{\text{cur}} )</td>
</tr>
<tr>
<td>( x_{\text{ref}}^* )</td>
<td>the block in reference frame with same position as ( x_{\text{ref}} )</td>
</tr>
<tr>
<td>( F_{\text{key}} )</td>
<td>the key frame in current GOP</td>
</tr>
<tr>
<td>( F_{\text{cur}(i)} )</td>
<td>the ( i )th non-key frame in current GOP</td>
</tr>
<tr>
<td>( F_{\text{cur}(i-1)} )</td>
<td>the ( (i-1) )th non-key frame in current GOP</td>
</tr>
<tr>
<td>( W_{\text{ref}} )</td>
<td>the search window of ( x_{\text{ref}} )</td>
</tr>
<tr>
<td>( h_{\text{opt}} )</td>
<td>the best hypothesis of ( x_{\text{ref}} ) in key frame</td>
</tr>
<tr>
<td>( M_{\text{ref}} )</td>
<td>the motion vector of ( x_{\text{ref}} )</td>
</tr>
<tr>
<td>( M_{\text{cur}} )</td>
<td>the motion vector of ( x_{\text{cur}} )</td>
</tr>
<tr>
<td>( D_{\text{cur}} )</td>
<td>the inter-frame distance between ( x_{\text{ref}} ) and key frame</td>
</tr>
<tr>
<td>( H_{\text{sort}} )</td>
<td>the hypotheses set after sorting</td>
</tr>
<tr>
<td>( H_{\text{opt}} )</td>
<td>the hypothesis set after residual transformation</td>
</tr>
<tr>
<td>( h_{\text{opt}} )</td>
<td>the best hypothesis of current block ( x_{\text{cur}} )</td>
</tr>
<tr>
<td>( h_{\text{opt}} )</td>
<td>the ( i )th residual hypothesis</td>
</tr>
<tr>
<td>( x_{\text{opt}} )</td>
<td>the predicted matching block in the key reference frame</td>
</tr>
<tr>
<td>( x_{\text{up}} )</td>
<td>the upper adjacent block of current block</td>
</tr>
<tr>
<td>( x_{\text{left}} )</td>
<td>the left adjacent block of current block</td>
</tr>
<tr>
<td>( x_{\text{ref},k} )</td>
<td>the ( k )th candidate matching block</td>
</tr>
<tr>
<td>( S_{\text{ref}} )</td>
<td>the candidate motion vector set</td>
</tr>
<tr>
<td>( S_{\text{cur}} )</td>
<td>the candidate reference blocks set</td>
</tr>
<tr>
<td>( x_{\text{ref},\text{opt}} )</td>
<td>the final predicted matching block</td>
</tr>
<tr>
<td>( x_{\text{res}} )</td>
<td>the residual block of ( x_{\text{cur}} )</td>
</tr>
<tr>
<td>( y_{\text{res}} )</td>
<td>the residual measurements of ( x_{\text{cur}} )</td>
</tr>
<tr>
<td>( w_{\text{opt}} )</td>
<td>the optimal block of the residual hypothesis set</td>
</tr>
<tr>
<td>( \Gamma_{\text{cur}} )</td>
<td>the diagonal regularization matrix</td>
</tr>
<tr>
<td>( x_{\text{cur}}^* )</td>
<td>the final predicted side information of block ( x_{\text{cur}} )</td>
</tr>
</tbody>
</table>

All of the above problems require an adaptive and high-efficient hypotheses acquisition scheme. The adjacent non-key frames should be utilized as the supplement of the key reference frame. It will help to resist the impact of the increased inter-frame distance between the key frame and the non-key frame on the final hypotheses quality, especially for the non-key frames in the middle of the GOP. On the basis of multi-reference frames-based hypotheses acquisition schemes, considering the temporal and spatial correlation of the image blocks, design a joint and adaptive hypotheses acquisition scheme to acquire the high-quality hypotheses in a simple way is another requirement for the new hypotheses acquisition scheme.

After building the search window, by block sliding pixel by pixel, the hypothesis \( h_i \) is collected from the search window. The hypothesis set \( H = \{ h_1, h_2, \cdots, h_L \} \) is obtained. To remove the low-quality hypotheses, the Euclidean distance \( d(x_{\text{cur}}, h_i) \) between the current block \( x_{\text{cur}} \) and the hypothesis \( h_i \) is calculated as below

\[
d(x_{\text{cur}}, h_i) = \| y_{\text{cur}} - \Phi h_i \|_2,
\]
information of the current block $x_{cur}$ can be obtained by computing the weighted sum of the hypotheses in $H$. Therefore, the weight prediction of the hypotheses, i.e., finding the optimal weight vector $w_{cur}^{mh}$, is the key problem. Commonly, the MH weight optimization problem can be solved by finding a weight vector $w$ to minimize the following problem

$$w_{cur}^{mh} = \arg \min_w \| y_{cur} - \Phi H w \|_2^2,$$  \hspace{1cm} (3)

Then the final obtained side information is applied to perform the final residual reconstruction of the block. Based on the problem (3), many new prediction methods have been proposed to improve the accuracy of weight prediction by introducing different sparse representation regularization terms. However, the performance of the sparsity penalty terms is highly related to the quality of the hypothesis set. The accuracy of the weight prediction decreases with the degradation of hypotheses quality. Therefore, current methods are highly sensitive to the quality fluctuation of the hypotheses. The accuracy of the weight prediction degrades significantly for the high-motion sequences. It causes a huge gap in reconstruction quality between high-motion sequences and low-motion sequences.

To avoid the impact of the fluctuated hypotheses on the prediction accuracy, the sparsity of the original signal to be predicted is the key. By transforming the original signal prediction to another sparser signal prediction, a stable improvement for different sequences can be obtained. Therefore, how to improve the sparsity of the original signal or how to realize the signal transforming in weight prediction procedure is the key to improve the performance of the weight prediction model for high-motion sequences. Besides, by jointly considering the hypotheses acquisition and the subsequent weight prediction procedure, designing the whole MH system to improve the MH prediction performance for high-motion sequences is also an urgent problem.

III. THE PROPOSED MH PREDICTION SYSTEM

Fig. 2 shows the proposed new MH prediction system. In this proposal, we focus on the recovery of non-key frames. Key frames are recovered using the secondary recovery method [53]. Both the key frame and previously recovered non-key frames are used as reference frames for the recovery of the current non-key frame. The proposed new MH prediction system consists of three parts: the optimal matching block prediction, the search window construction and hypotheses acquisition, and the final MH weight prediction based on residual transforming.

First, in the step of optimal matching block prediction. Two situations are considered. When there is other non-key frame between the current frame and the reference frame. Both the spatial and temporal correlation-based prediction methods are applied. The temporal correlation-based method is used to get the matching block in the key reference frame. The spatial correlation-based method is used to predict the matching block in the non-key reference frame. When there is no adjacent non-key frame between the reference frame and current frame, the adjacent blocks in the current frame are selected to provide the information of motion state i.e., the spatial correlation-based prediction method is performed. Second, centered on the optimal matching block, the search window is constructed and hypotheses are collected. The hypothesis set optimization processing is performed to remove low-quality hypotheses in the obtained hypothesis set. At last, for the final residual-based MH weight prediction, a hypothesis basis is first decided. The residual transform preprocessing is performed by transforming the original hypothesis set to the residual hypothesis set. The elastic net-based weight prediction model is utilized to perform the weight prediction of the residual hypothesis set.

A detailed description of the proposed new MH prediction system is given below. The proposed new hypotheses acquisition scheme is described in the first two subsections, where subsection III.A shows the temporal correlation-based MH acquisition method and subsection III.B shows the spatial correlation-based MH acquisition method. In subsection III.C, the proposed new residual-based MH weight prediction algorithm is given in detail.

A. The Temporal Correlation-Based MH Acquisition Scheme

Fig. 3 shows the proposed temporal correlation based hypotheses acquisition procedure. $F_{key}$ is the key reference frame; $F_{cs(i−1)}$ and $F_{cs(i)}$ are the $(i−1)^{th}$ and $i^{th}$ non-key frames, respectively; $x_{cur}$ is the current block to be reconstructed in current frame $F_{cs(i)}$; $x_{ref}$ is its corresponding optimal reference block in the previous non-key frame; In the key reference frame, $x'_{cur}$ and $x'_{ref}$ are the corresponding blocks that have the same positions as $x_{cur}$ and $x_{ref}$, respectively; $W_{ref}$ is the corresponding search window of $x_{ref}$. It is also determined by the proposed TS-MH hypotheses acquisition scheme in the previous recovery of $x_{ref}$.

Here, we assume that the optimal matching block $x_{ref}$ of $x_{cur}$ in the adjacent non-key frame is already known. The detailed computing procedure of $x_{ref}$ will be shown in the subsequent spatial correlation based hypotheses acquisition
scheme. The proposed temporal correlation based MH acquisition scheme provides a way to construct the search window of \( x_{\text{cur}} \) according to the previously recovered block \( x_{\text{ref}} \).

In the key reference frame \( F_{\text{key}} \), we first obtain the hypothesis set of \( x_{\text{ref}} \). Then, the hypotheses are sorted by the hypothesis quality. Here, the hypothesis quality is measured by the Euclidean distance between \( x_{\text{ref}} \) and its hypotheses as shown in (2). The hypothesis quality increases with the decrease of the Euclidean distance. We assume that the block \( h_B \) in \( W_{\text{ref}} \) is the best hypothesis of \( x_{\text{ref}} \) after the hypotheses sorting. The motion vector (MV) \( MV_{\text{ref}} \) of \( x_{\text{ref}} \) can be obtained by computing the coordinate difference between \( h_B \) and \( x_{\text{ref}} \) as below

\[
MV_{\text{ref}} = (h_B(j) - x_{\text{ref}}(j)),
\]

where \( j \) represents the pixel index; \( h_B(j) \) and \( x_{\text{ref}}(j) \) represent the coordinates of the \( j \)th pixel in the whole frame. According to the inter-frame correlation, we assume that the \( x_{\text{ref}} \) and \( x_{\text{cur}} \) have the same motion trends. The motion vector of \( x_{\text{cur}} \) is calculated as below

\[
MV_{\text{cur}} = \frac{D_{\text{cur}}}{D_{\text{cur}} - 1} MV_{\text{ref}},
\]

where \( D_{\text{cur}} \) is the frame distance between current frame and the key reference frame. Then, we can get one of the candidate matching blocks \( x_{\text{MV}} \) pointed by \( MV_{\text{cur}} \). Centering on block \( x_{\text{MV}} \), the search window of \( x_{\text{cur}} \) can be constructed in the reference frame.

However, for some sequences with complex motion states, when a sudden change of motion occurs between \( x_{\text{cur}} \) and \( x_{\text{ref}} \), the above strategy will cause the prediction error and the error propagation in the subsequent reconstruction. To avoid this problem, other candidate matching blocks should be acquired according to the content similarity of \( x_{\text{cur}} \) and \( x_{\text{ref}} \). Considering the best hypothesis \( h_B \) is the best matching block of \( x_{\text{ref}} \) in the reference frame. It means that \( h_B \) has high similarity with the current block \( x_{\text{cur}} \). Therefore, we utilized \( h_B \) as the other candidate matching block. The final optimal matching block \( x_{\text{opt}} \) is computed by

\[
x_{\text{opt}} = \begin{cases} 
  x_{\text{MV}} 
  & \text{d}(x_{\text{cur}}, x_{\text{MV}}) \leq \text{d}(x_{\text{cur}}, h_B) \\
  h_B 
  & \text{else}
\end{cases}
\]

where \( \text{d}(\cdot, \cdot) \) means the calculation of the Euclidean distance in the measurement domain as described in (2). Centered on the final predicted matching block \( x_{\text{opt}} \), the final search window of the current block \( x_{\text{cur}} \) is constructed by extending a fixed number of pixels. Then, hypotheses in the search window are collected and applied for the subsequent weight prediction procedure. Noted that the above hypotheses acquisition and hypotheses sorting processing of the reference block \( x_{\text{ref}} \) are all the necessary calculations in the recovery of \( x_{\text{ref}} \). It means that the computational cost of the above proposal mainly comes from the final selection of the two candidate matching block. By using the proposed temporal correlation-based hypotheses acquisition scheme, the high-quality hypotheses in the key reference frame are collected effectively and simply.

### B. The Spatial Correlation-Based MH Acquisition Scheme

The temporal correlation-based MH acquisition scheme gives the hypotheses obtaining method when there are other non-key frames between the current non-key frame and the reference frame. It does not work when the current non-key frame is next to the reference frame. Actually, this problem is equivalent to how to estimate the optimal matching block from the adjacent frame, i.e., how to get the reference block \( x_{\text{ref}} \) from the previous frame \( F_{\text{ref}(-1)} \) for current block \( x_{\text{cur}} \). For this situation, the proposed spatial correlation-based MH acquisition scheme is performed.

As shown in Fig. 4, \( x_{\text{left}} \) and \( x_{\text{up}} \) are the left block and the upper block of the current block \( x_{\text{cur}} \) in current non-key frame. To effectively show the motion of the image block, we put the above three image blocks into its adjacent reference frame \( F_{\text{ref}} \) at the corresponding positions. In the proposed spatial correlation-based search window construction scheme, the first step is to estimate the MV of the current block \( x_{\text{cur}} \). Here we consider two situations. one is that the current block is a static block. It means that there is no motion between the current frame and the reference frame for the current block \( x_{\text{cur}} \). Therefore, the candidate motion vector \( MV_{\text{nom}} = (0, 0) \). Noted that the no motion situation is always considered as one of the candidates in motion prediction. When there is a motion for block \( x_{\text{cur}} \), we assume that \( x_{\text{cur}} \) has a similar motion trend as its left block and upper block. It means that the motion vector of \( x_{\text{cur}} \) can be calculated by combing the motion vectors of \( x_{\text{left}} \) or \( x_{\text{up}} \). Or even it is equal to that of \( x_{\text{left}} \) or \( x_{\text{up}} \). In the recovery of \( x_{\text{left}} \) and \( x_{\text{up}} \), by calculating the difference of the position between the optimal hypothesis and original block, we can obtain their corresponding motion vectors \( MV_{\text{up}} \) and \( MV_{\text{left}} \). By combining the \( MV_{\text{up}} \) and \( MV_{\text{left}} \), we can calculate a combined motion vector \( MV_{\text{com}} \) for \( x_{\text{cur}} \) by

\[
MV_{\text{com}} = \frac{1}{2} (MV_{\text{up}} + MV_{\text{left}}),
\]

Considering the above situations, we get a candidate MV set as \( S_{\text{MV}} = \{ MV_{\text{left}}, MV_{\text{up}}, MV_{\text{com}}, MV_{\text{nom}} \} \). The corresponding candidate matching block set \( S_{\text{ref}} = \{ x_{\text{ref},1}, x_{\text{ref},2}, x_{\text{ref},3}, \)
The correlation between the predicted matching block in spatial correlation-based MH acquisition scheme and the original non-overlapped blocks divided at encoder.

$x_{ref,} \{A\}$ in the reference frame is constructed. The detailed position of the corresponding reference block $x_{ref,k}$ is calculated as below

$$(j_1, j_2)_{x_{ref,k}} = (j_1, j_2)_{x_{cur}} + S_{MV}(k),$$  

where $1 \leq k \leq 4$ is the index of the candidate matching blocks; $j_1$ and $j_2$ are the pixel position indexes in horizontal and vertical directions. The optimal matching block $x_{ref, opt}$ is calculated by

$$x_{ref, opt} = \arg \min_{x_{ref,k} \in S_{ref}} d(x_{cur}, x_{ref,k}),$$

where $d(\cdot, \cdot)$ is the calculation of the Euclidean distance. After getting the optimal matching block in the adjacent reference frame, the search window is then constructed centered on the optimal matching block.

However, it will cause some problems if we directly use the matching block $x_{ref, opt}$ predicted by the above spatial correlation-based strategy as the reference block $x_{ref}$ in Fig. 3. According to the prediction procedure shown in Fig. 4, we know that the candidate matching blocks are overlapped in this subsection. The predicted matching block may not be the original non-overlapped block divided at encoder. As shown in Fig. 5, when the predicted matching block is not one of the original non-overlapped blocks divided at encoder, it means that the predicted matching block $x_{ref, opt}$ is not reconstructed as a complete image block in the recovery of the previous non-key frame. No corresponding motion vector information for $x_{ref, opt}$ can be provided for the recovery of $x_{cur}$ in the temporal correlation-based MH acquisition scheme. To avoid this issue, the adjacent four original blocks shown in Fig. 5 covered by the predicted reference block $x_{ref, opt}$ will be selected as the latest candidate reference blocks. By calculating the similarity between these candidate reference blocks and the current block $x_{cur}$, the candidate reference block with the best similarity is selected as the final predicted reference block $x_{ref}$. Then, the subsequent search window construction and hypotheses acquisition processing are performed following the temporal correlation-based search window construction method. By using the spatial correlation-based hypotheses acquisition scheme, the high-quality hypotheses in the adjacent non-key reference frame and the key reference frame are collected effectively.

### C. Residual Transforming Preprocessing-Based Weight Prediction Algorithm

In CS reconstruction, residual-based iteration recovery algorithms provide an effective way to improve recovery quality due to the improvement of the signal sparsity caused by residual transformation [45]. Inspired by this idea, a new weight prediction algorithm based on the residual transforming of hypotheses is proposed in this subsection. We propose a preprocessing method to transform the original hypotheses to the residual hypotheses. It is the main contribution in this part. For the specific weight prediction processing, we use the elastic net model (wEnet) proposed in [35] due to its good performance at low sampling rates.

By transforming the original hypotheses to the residual hypotheses, on the one hand, the sparsity of the hypothesis is improved effectively for weight prediction model. On the other hand, residual transformation can offset the degradation of the weight prediction accuracy caused by the hypotheses fluctuation, especially for the high-motion sequences. The MH prediction system can adapt to the video sequences with different motion states better. Here, the key issue is how to transform the original hypotheses to the residual hypotheses, i.e., how to select the hypothesis basis to calculate the residual hypotheses. Meanwhile, the residual hypotheses should keep consistent with the corresponding residual of the original block to be recovered.

To realize the residual transforming of the hypotheses, we first calculate the Euclidean distance between each hypothesis and the current block after getting the final hypothesis set by the proposed TS-MH scheme. By resorting the hypotheses according to the Euclidean distance from small to large, we can construct a new hypothesis set $H_{sort} = \{h_1, h_2, \cdots, h_L\}$. The hypothesis $h_1$ with the minimum Euclidean distance is selected as the best hypothesis $h_{best}$. The best hypothesis is selected as the hypothesis basis to compute the residual hypotheses. By computing the difference between the original hypotheses and the best hypothesis, we get the residual hypothesis $h_{r,i}$ as below,

$$h_{r,i} = h_{i+1} - h_{best},$$

where $1 \leq l \leq L - 1$. The residual hypotheses are used to construct residual hypothesis set $H_r = \{h_{r,1}, h_{r,2}, \cdots, h_{r,L-1}\}$. We assume that the best hypothesis provides the most content of the final side information for current block. For the rest of the information in the current block that is not contained in the best hypothesis, they are distributed among these residual hypotheses and can be obtained by the linear combination of the residual hypotheses.

For current image block $x_{cur}$ to be recovered, the decoder receives its measurements $y_{cur}$ transmitted from the encoder. After transforming the original hypothesis set $H_{sort}$ to the residual hypothesis set $H_r$, the next step is to calculate the MH weight for each residual hypothesis in $H_r$. First, we have to make a
targeted residual block for $H_r$. To make the residual hypotheses consistent with the corresponding residual block of the original block $x_{cur}$, we compute the residual measurements $y_r$ of $x_{cur}$ as below

$$y_r = y_{cur} - \Phi h_{best}. \quad (11)$$

After the above preprocessing of the hypothesis set, we use the wEnet-based MH prediction model [55] to compute the optimal residual weight vector $w_{r,\text{cur}}^{\text{Enet}}$ due to its excellent performance under low sampling rates. It can be written as below

$$w_{r,\text{cur}}^{\text{Enet}} = \left(1 + \frac{\lambda_2}{M}\right) \arg \min_w \|y_r - \Phi H_r w\|_2^2 + \lambda_1 \|\Gamma_{r,\text{cur}} w\|_1 + \lambda_2 \|w\|_2^2, \quad (12)$$

where $M$ is the number of the measurements in $y_r$; $\lambda_1$ and $\lambda_2$ are the tuning parameters of the wEnet model. $\lambda_1 = 1000 \times CS_{\text{ratio}}$, where $CS_{\text{ratio}}$ is the sampling rate; $\lambda_2$ is usually in the range of $[0.01, 0.2]$; $\Gamma_{r,\text{cur}}$ is a diagonal regularization matrix as below

$$\Gamma_{r,\text{cur}} = \begin{pmatrix} d(x_r, h_{r,1}) & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & d(x_r, h_{r,L-1}) \end{pmatrix}, \quad (13)$$

where $d(\cdot, \cdot)$ is the computation of the Euclidean distance between the residual hypothesis and the residual block. It can be computed by (2). The final side information $\hat{x}_{cur}$ of the original block $x_{cur}$ is computed as below

$$\hat{x}_{cur} = h_{best} + H_r w_{r,\text{cur}}^{\text{Enet}}. \quad (14)$$

The finally predicted block $\hat{x}_{cur}$ and the original measurements $y_{cur}$ are sent to the final recovery model. The residual-based BCS-SPL algorithm [45] is utilized to perform the final reconstruction of the non-key frame.

The proposed residual transforming preprocessing scheme effectively improves the sparsity of the signal in MH weight prediction model. The increased signal sparsity improves the accuracy of weight prediction for all kind of sequences. Especially for high-motion videos, the degradation of weight prediction accuracy caused by the fluctuation of hypotheses is effectively mitigated by the proposed residual transforming preprocessing scheme. Moreover, as illustrated in the MH acquisition procedure, by using the proposed new hypotheses acquisition scheme, the quality of the final obtained hypothesis set is improved effectively. In the residual transforming procedure, the residual hypotheses are obtained by computing the difference between the original hypothesis and the best hypothesis. The improvement of the sparsity after residual transforming is more significant if the whole quality of the hypothesis set is better. The mutual promotion between the proposed new hypotheses acquisition scheme and the new weight prediction scheme makes the final performance improvement more significant.

IV. EXPERIMENTAL RESULTS

In this section, the experimental results are given to verify the performance of the proposed new MH-based DCVS system. The peak signal-to-noise ratio (PSNR), as the evaluation metric of the final reconstruction quality, is measured and given in this section. For the test source videos, Coastguard, Foreman, Mother-daughter and Soccer sequences, as the standard test videos, are applied in this experiment. The above four sequences cover different motion intensity. The test results based on the selected test sequences can reveal the real performance of the proposed new MH system. By the way, Coastguard, Foreman, Mother-daughter, and Soccer sequences are also widely utilized in the research of DCVS systems, including the reference systems introduced in this paper. Ensure fairness is also a consideration. For the parameter setting, the Gaussian orthogonal matrix is used as the measurement matrix $\Phi$. The size of the GOP is 8. At the encoder, the video frames are divided into non-overlapped image blocks with a fixed size of $16 \times 16$. The window size is 15 pixels. The sampling rate of key frames is 0.7. For non-key frames, the sampling rate varies from 0.1 to 0.3. Noted that the above parameter setting is consistent with the reference systems introduced in this paper. All of the experiments are performed on the platform of 64-bit Windows 7SP1, Inter(R) Core(TM) i7 CPU, 2.67 GHz, 12 G RAM. The Matlab version is R2013b.

A. The Comparison of the MH Acquisition Scheme

To verify the performance of the proposed TS-MH hypotheses acquisition scheme, we carry out the same distributed compressed video sensing scheme with the TS-MH method and without it, i.e., the traditional multihypothesis acquisition scheme [45], [55], respectively. To ensure fairness, the tested DCVS system remains unchanged except for the tested hypotheses acquisition part in this experiment. The PSNR results of the recovered non-key frames show the real effect of the proposed TS-MH method on the quality of the final acquired hypotheses. The test results of Coastguard and Foreman sequences are given in Fig. 6 as examples. The average time complexity results of the comparison systems for all of the test videos are given in Table II.
We can observe that the proposed TS-MH scheme effectively improves the quality of the recovered video sequences compared to the traditional hypotheses acquisition scheme. For the sequence of Coastguard, the average quality of the recovered non-key frames is improved by 0.55 dB on average. For the Foreman sequence, the PSNR of the recovered non-key frames is improved by 0.63 dB on average. It shows that the proposed TS-MH scheme can get more and better hypotheses from the reference frames. Due to the improvement of the hypothesis set, the final side information is predicted in higher accuracy. It means that the residual block is more sparse in the subsequent residual-based BCS-SPL reconstruction processing. Correspondingly, the final recovery quality of the video is improved. With the decrease of the sampling rate, the accuracy of MH weight prediction and the final iteration recovery processing degrade significantly. The effect of the improvement of the hypotheses quality on the gain of the final recovery quality is not obvious at very low sampling rates. Moreover, at a very low sampling rate, measurement domain based Euclidean distance is not always accurate to measure the block similarity. It will lead to the misjudgment of the optimal matching block and reduce the efficiency of the proposed TS-MH hypothesis acquisition scheme. Therefore, the performance improvement of the proposed TS-MH hypothesis acquisition scheme is not obvious at low sampling rates. However, even in the very low sampling rate, after combined with the proposed residual transforming based MH weight prediction method, the newly proposed MH system can still generate considerable incremental gain of the PSNR result as shown in Fig. 7.

Table II gives the comparison results of the computational complexity for per frame processing between the proposed TS-MH based video encoding system and the traditional multihypothesis acquisition method based video encoding system. We can observe that the proposed TS-MH hypothesis acquisition scheme only increases a little of the computational complexity compared to the traditional method. The reason is that most of the computation in the proposed TS-MH scheme are also present in the traditional MH scheme, such as the hypotheses sorting. Most of the information utilized in TS-MH scheme already exists in the MH processing. The proposed TS-MH scheme just uses such existing information in a better way. The slightly increased computation mainly comes from the decision process of the candidate reference blocks.

### B. MH Weight Prediction Algorithm

In this subsection, we compare the proposed residual-based multihypothesis weight prediction algorithm (RMH-wEnet) to the reference MH prediction models, including the Tikhonov regularization-based MH prediction model (MH-Tik) [45] and the elastic net-based MH prediction model (MH-wEnet) [55]. To ensure fairness, the above tests are performed under the same DCVS system diagram. In addition to the MH weight prediction method, the other parts of the tested system remain unchanged including the key frame recovery scheme, MH acquisition scheme, and the parameter setting. Moreover, this experiment uses the traditional hypotheses acquisition scheme utilized in [45], [55]. The new TS-MH hypotheses acquisition scheme is not utilized to avoid the influence of other factors. Table III gives the quality of the recovered videos at the decoder under different MH weight prediction algorithms when the sampling rate is 0.2. It shows the impact of the improvement of MH weight prediction accuracy on the quality of the final reconstruction at the decoder. Table IV gives the time complexity comparison results.

We can observe that the proposed residual-based MH prediction method has the best reconstruction quality. Compared to the MH-Tik model and the MH-wEnet model, the proposed RMH-wEnet algorithm can get 1.21 dB and 0.43 dB of the average PSNR gain under different test sequences. Specifically, compared to the MH-Tik weight prediction algorithm, the proposed

<table>
<thead>
<tr>
<th>Method</th>
<th>0.1</th>
<th>0.15</th>
<th>0.2</th>
<th>0.25</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS-MH</td>
<td>49.17</td>
<td>61.47</td>
<td>79.06</td>
<td>108.25</td>
<td>129.46</td>
</tr>
<tr>
<td>No TS-MH</td>
<td>47.20</td>
<td>59.37</td>
<td>76.46</td>
<td>104.17</td>
<td>124.53</td>
</tr>
</tbody>
</table>

Table II: The CPU runtime of the whole system for per frame processing under different hypotheses acquisition schemes. (Unit: S)
RMH-wEnet algorithm provides the biggest recovery quality gain for Mother-daughter sequences. The PSNR is improved by 1.58 dB. The high-motion Soccer sequence contributes the least quality improvement. The PSNR is improved by 0.82 dB in the proposal. For the Coastguard and Foreman sequences, the recovery quality is improved by 1.16 dB and 1.26 dB, respectively. For the MH-wEnet weight prediction algorithm, the least improvement of the recovery quality comes from the Coastguard sequence. The PSNR is improved by 0.36 dB in the proposed RMH-wEnet algorithm. Following by the Soccer sequence, the recovery quality is improved by 0.43 dB. For the Mother-daughter sequence, it has the best improvement of the recovery quality. 0.51 dB of the PSNR gain is obtained in the RMH-wEnet algorithm. The PSNR gain of the Foreman sequence is 0.44 dB. Compared to the high-motion sequences, the hypothesis set of low-motion sequences has higher quality. By residual transforming, the residual hypothesis set has higher sparsity for low-motion sequences than that of the high-motion sequences. Therefore, the low-motion sequences have higher quality improvement.

However, in terms of the extent of quality improvement, compared to the PSNR gain from the MH-Tik scheme to the MH-wEnet scheme, the quality gain obtained from the MH-wEnet scheme to the proposed RMH-wEnet scheme is consistent for the test sequences with different motion level. In Table III, both MH-wEnet and RMH-wEnet models use the wEnet model to do multihypothesis weight prediction. They have the same hypotheses acquisition method and the final obtained hypothesis set in the experiment. The proposed residual transforming pre-processing for hypothesis set in RMH-wEnet scheme is the only difference. Compared to the MH-wEnet model, the proposed RMH-wEnet model gets more than 0.4 dB of the PSNR gain in the final recovery quality. Even for the high-motion videos Soccer and Coastguard, 0.38 dB of the PSNR gain is also obtained in the proposed RMH-wEnet model. A consistent quality gain is obtained from the MH-wEnet scheme to the proposed RMH-wEnet scheme. It means that the proposed residual transforming preprocessing scheme can always improve the weight prediction accuracy steadily no matter how the quality of hypothesis set fluctuates. It is verified that the proposed residual transforming based weight prediction scheme can mitigate the degradation of the prediction accuracy caused by the fluctuation of hypotheses quality, as we analyzed in Section III.C. Moreover, considering the proposed new TS-MH hypotheses acquisition scheme, the proposed new MH system can get more and better hypotheses for high-motion sequences. The performance improvement of the proposed RMH-wEnet weight prediction scheme for high-motion sequences will be further amplified under the help of the proposed TS-MH hypotheses acquisition scheme. The gap between the high-motion sequences and the low-motion sequences in reconstruction quality will be reduced. It is revealed in the following overall performance comparison subsection.

Table IV gives the CPU computation time of the whole system for per frame processing under different weight prediction schemes. We can observe that the proposed RMH-wEnet based video codec systems require less CPU time than the MH-wEnet based video codec systems. There are two main reasons: First, by residual transforming, the improved sparsity increases the computation efficiency of weight prediction; Second, the proposed residual-based MH prediction scheme improves the quality of the predicted SI. More accurate SI makes the residual signal sparser in the subsequent BCS-SPL iteration recovery processing. It speeds up the iteration of BCS-SPL algorithm. Compared to the MH-Tik based video codec systems, the proposed RMH-wEnet scheme consumes more time for per frame processing especially at high sampling rates. It is caused by the wEnet weight prediction model. Nevertheless, with an acceptable increase of the complexity, a good improvement in the weight prediction accuracy is obtained in the proposed residual transforming preprocessing method.

### C. The Comparison of Overall Performance

To verify the overall performance, we compare the proposed NMH-DCVS system with several complete DCVS systems in current research. The state-of-the-art DCVS system, including the MH-BCS-SPL system [45], MS-wEnet system [55], KSR-DCVS system [53], and Up-Se-AWEN-HHP system [52] are tested as the comparison systems. The MH-BCS-SPL [45] system is the first and typical MH-based DCVS system. It is widely introduced in current research. The MS-wEnet [55] system is the representative in the research of MH weight prediction algorithm. The KSR-DCVS [53] and Up-Se-AWEN-HHP [52] systems make significant improvements in hypotheses acquisition and optimization. To ensure fairness, the above systems are performed on the same platform with the same parameter setting, such as the GOP size, search window size, and block size. The overall recovery results measured by PSNR are shown in Fig. 7. Moreover, Fig. 8 and Fig. 9 give two visual comparison results of the recovered non-key frames to show the visual quality intuitively. The feature similarity index for image
Fig. 8. The visual quality for the 76th non-key frame of Foreman (sampling rate: 0.15). (a) Original frame. (b) The selected original region. (c) MH-BCS-SPL, PSNR 29.94 dB, FSIM 0.905. (d) Up-Se-AWEN-HHP, PSNR 29.80 dB, FSIM 0.906. (e) MS-wEnet, PSNR 30.92 dB, FSIM 0.916. (f) KSR-DCVS, PSNR 30.30 dB, FSIM 0.908. (g) NMH-DCVS, PSNR 33.49 dB, FSIM 0.947. The proposed NMH-DCVS method in (g) has the best visual quality.

quality assessment (FSIM) [60] of the recovered frames shown in Fig. 8 and Fig. 9 is also measured to evaluate the visual quality objectively.

Fig. 7 shows that the proposed NMH-DCVS system has the best recovery performance compared to the reference schemes. Compared to the KSR-DCVS system, the proposed NMH-DCVS system improved the PSNR by about 1.13 dB on average. The biggest recovery gain comes from the Coastguard sequence, the PSNR is improved by about 1.31 dB. The PSNR of Soccer sequence is improved by 0.95 dB, it is the least PSNR gain. For the Up-Se-AWEN-HHP scheme, the average recovery quality is improved by about 1.19 dB in our NMH-DCVS system. The high-motion sequence Soccer and low-motion sequence Mother-daughter contribute the biggest and the least PSNR gain by 1.50 dB and 0.75 dB, respectively. Compared to the MS-wEnet scheme, the proposed NMH-DCVS system improved the average PSNR by about 1.36 dB. The recovery quality of the Foreman sequence is improved significantly by about 1.66 dB. About 1.25 dB of the PSNR gain is obtained for the other three test sequences. For the original MH based video codec system i.e., the MH-BCS-SPL system, the average value of the PSNR is improved by about 3.19 dB in the proposed NMH-DCVS system. The smallest recovery quality gain comes from the Soccer sequence, the PSNR is increased by 2.37 dB. The Coastguard sequence contributes to the biggest quality gain. The recovery quality is improved by 3.87 dB.

We can observe that the proposed NMH-DCVS system can obtain better PSNR gain for the high-motion sequences. The reason is that the proposed TS-MH hypotheses acquisition scheme can get more high-quality hypotheses for high-motion sequences than that of the other reference systems. Correspondingly, the correlation among the hypotheses is better in the hypothesis set obtained in the TS-MH scheme. In the subsequent weight prediction procedure, the residual hypotheses are sparser after transforming the original hypothesis set to the residual hypothesis set. The sparser hypothesis set makes the RMH-wEnet weight prediction procedure more accurate. It shows that the proposed TS-MH hypotheses acquisition scheme and the RMH-wEnet weight prediction scheme can promote each other to achieve the optimal performance. By the way, it also proves the conclusion given in the last subsection, i.e., the gap of the recovery quality between the high-motion and low-motion sequences is reduced effectively in the proposed whole MH prediction system.

Compared to the state-of-the-art DCVS systems, the improvement of the overall PSNR in the proposed new MH system may be not quite significant as shown in Fig. 7. However, in terms of the visual quality, an obvious improvement of the visual quality is obtained. As shown in Fig. 8 and Fig. 9, the recovered 76th non-key frame of Foreman sequence under sampling rate 0.15 and the 20th non-key frame of Soccer sequence under sampling rate 0.2 are given. To show the texture details of the recovered frame, the complete recovered frames are not shown. Only the
original frame and the selected region in the recovered frames are given in Fig. 8 and Fig. 9. Compared to the reference systems, the proposed NMH-DCVS system can preserve more texture details. In Fig. 8, the recovered contour of the face in the reference systems is blurred. In Fig. 9, the recovered tilted pole encountered the same problem in the reference systems. Especially for the MS-wEnet system, the middle part of the recovered tilted pole has been lost. In the proposed NMH-DCVS system, both the image contour and the internal texture details are better preserved and recovered. Both the PSNR and FSIM results show the superior performance of the newly proposed NMH-DCVS system.

V. CONCLUSION

In this paper, we propose a new MH system with the new hypotheses acquisition and weight prediction schemes. First, we propose a new hypotheses acquisition scheme based on temporal and spatial correlation of the image blocks. The optimal matching block and the search window in the reference frames are obtained in an adaptive and simple way in the new hypotheses acquisition scheme. The high-quality hypotheses from different reference frames are collected efficiently especially for high-motion sequences. The average PSNR gain is improved by about 0.59 dB for the final recovery quality under the effect of the new hypotheses acquisition scheme. Second, for weight prediction, we propose a new MH weight prediction algorithm based on residual transforming preprocessing of the hypothesis set. By transforming the original hypotheses weight prediction to the residual hypotheses weight prediction, the sparsity of the predicted signal is improved effectively. An effective improvement of the prediction accuracy is obtained for all kinds of test sequences. Compared to the state-of-the-art weight prediction algorithm, the proposed new residual weight prediction model improves the final recovered average PSNR efficiently.

Moreover, the improved hypothesis set in the proposed new hypotheses acquisition scheme makes the hypotheses have better correlation. It makes the residual hypotheses sparser. The improvement of the sparsity after residual transforming is more significant. A better accuracy can be obtained in the new weight prediction algorithm. The mutual promotion of the proposed hypotheses acquisition scheme and the weight prediction algorithm makes the performance of the whole MH prediction system better. For the overall performance of the proposed whole DCVS system, both the objective quality and the visual quality of the recovered sequences are improved effectively. More than 1.13 dB of the PSNR gain and 0.03 of the FSIM gain are obtained in the proposed new DCVS system compared to the recent state-of-the-art systems. Higher gain of the recovery quality is obtained for high-motion sequences than that of low-motion sequences. The gap in the recovery quality between high-motion sequences and low-motion sequences is reduced efficiently under a high-level recovery quality.

REFERENCES


Shuai Zheng received the B.S. degree in electronic information engineering from Henan University of Technology, Henan, China, in 2015, and the Ph.D. degree in military communication from Xidian University, Xi’an, China, in 2020. From October 2018 to December 2019, he was a joint Ph.D. student with the Communication and Signal Processing Applications Laboratory, Ryerson University, Toronto, Canada. His current research interests involve compressed sensing, distributed video coding, image and video processing, and multimedia communication.

Jian Chen (Member, IEEE) received the B.S. degree from Xi’an Jiaotong University, Xi’an, China, in 1989, the M.S. degree from Xian Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Beijing, China, in 1992, and the Ph.D. degree in telecommunications engineering from Xidian University, Xi’an, China, in 2005. He is a Professor with the School of Telecommunications Engineering, Xidian University. He was a visitor scholar with the University of Manchester from 2007 to 2008. His research interests include cognitive radio, non-orthogonal multiple access, multimedia communications, and wireless sensor networks.

Xiao-Ping Zhang (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1992 and 1996, respectively, and the MBA degree (Hons.) in finance, economics and entrepreneurship from the University of Chicago Booth School of Business, Chicago, IL, USA. Since Fall 2000, he has been with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada, where he is currently a Professor and the Director of the Communication and Signal Processing Applications Laboratory. He has served as the Program Director of Graduate Studies. He is cross-appointed to the Finance Department with the Ted Rogers School of Management, Ryerson University. He was a Visiting Scientist with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA, in 2015 and 2017. He is a frequent consultant for biotech companies and investment firms. He is the Co-Founder and CEO for EidoSearch, an Ontario-based company offering a content-based search and analysis engine for financial big data. His research interests include image and multimedia content analysis, machine learning, statistical signal processing, sensor networks and electronic systems, and applications in big data, finance, and marketing. Dr. Zhang is a Fellow of Canadian Academy of Engineering, a registered Professional Engineer in Ontario, Canada, and a member of Beta Gamma Sigma Honor Society. He is the general Co-Chair for the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2021. He is the General Co-Chair for 2017 GlobalSIP Symposium on Signal and Information Processing for Finance and Business, and the General Co-Chair for 2019 GlobalSIP Symposium on Signal, Information Processing and AI for Finance and Business. He is an elected Member of the ICME steering committee. He is the General Chair for the IEEE International Workshop on Multimedia Signal Processing, 2015. He is the Publicity Chair for the International Conference on Multimedia and Expo 2006, and the Program Chair for International Conference on Intelligent Computing in 2005 and 2010. He served as a Guest Editor for Multimedia Tools and Applications and the International Journal of Semantic Computing. He was a tutorial speaker at the 2011 ACM International Conference on Multimedia, the 2013 IEEE International Symposium on Circuits and Systems, the 2013 IEEE International Conference on Image Processing, the 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing, the 2017 International Joint Conference on Neural Networks and the 2019 IEEE International Symposium on Circuits and Systems. He is a Senior Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON IMAGE PROCESSING. He was an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE SIGNAL PROCESSING LETTERS. He received 2020 Sarwan Sahota Ryerson Distinguished Scholar Award, the Ryerson University highest honor for scholarly, research and creative achievements. He is awarded as IEEE Distinguished Lecturer for the term from January 2020 to December 2021 by IEEE Signal Processing Society.

Yonghong Kuo received the B.S. and M.S. degrees from Xi’an Jiaotong University, Xi’an, China, in 1989 and 1992, respectively, and the Ph.D. degree from Peking Union Medical University, Beijing, China, in 1998. She is currently a Professor with the School of Telecommunications Engineering, Xidian University, Xi’an, China. Her research interests include signal processing, cognitive radio, and wireless sensor networks.