Video Concealment via Matrix Completion at High Missing Rates

Minh D. Dao, Dzung T. Nguyen, Yuan Cao, Trac D. Tran
The Department of Electrical & Computer Engineering
Johns Hopkins University, Baltimore MD

***Abstract*—Video error concealment is an important technique in video communication to recover corrupted parts when erroneously transmitting compressed sequences over network. In this work, we propose a novel error concealment scheme by grouping similar patches in the temporal domain to construct a low-rank matrix and recover the missing areas by the matrix completion technique. Different from most state-of-the-art algorithms which recover the lost blocks based on at least one clean frame (I-frame), the proposed algorithm can work for the most general case when all frames are violated. When having I-frames in the receiver, the proposed algorithm also achieves much higher PSNR compared with the Boundary Matching Algorithm (BMA) adopted in H.264.**

***Index Terms*—Video error concealment, BMA, low-rank matrix, matrix completion.**

I. Introduction

Along with the dramatic explosion of internet in these days, video communication over networks is becoming more and more popular. However, the information of videos may be lost during the transmission due to the band-limited network and channel noise. Especially, with block-based coding such as H.261/263/264 or MPEG family standards, the packet errors may lead to severe degradation of the visual quality at the user. When variable length coding (VLC) is used, the error of only one bit coding can corrupt many following bits in the same synchronization and make a bunch of blocks undecodable. There are several error control techniques to tackle this effect including retransmission, forward error correction (FEC), error-resilient coding and error concealment. The first three techniques aim at lossless recovery with some penalties such as transmission delays or extra bandwidth needed. On the other hand, error concealment strives to obtain a close approximated signal. With the insensitiveness of human eye with small amount of errors, the reconstructed signal can get comparable visualization with the original video.

Most of the existing error concealment techniques recover the missing blocks by exploiting the inherent spatial correlation in the intra-frame domain or temporal correlation among adjacent frames. While maximizing the smoothness is the typical technique in the spatial domain, temporal approaches focus on using motion vector (MV) to estimate missing block as motion compensated block from neighborhood frames. The temporal approach wins the favor in many researches as there is more correlation among inter frames than in the intra domain of a single frame. In [1], the zero MVs, the collocated blocks in the reference frame, as well as the average of the MVs from spatially adjacent blocks are used as estimated MVs for the lost blocks. The popular Boundary Matching Algorithm (BMA) in [2] tries to minimize the total variation between the internal and external boundaries of the reconstructed block to find the best candidate MVs. One variation of this method is adopted in H.264. Some more sophisticated methods based on BMA have recently been proposed to better estimate the MVs. For example, the Spatio-Temporal Boundary Matching Algorithm (STBMA) in [3], [4] introduce a new distortion function to simultaneously exploit both the spatial and temporal smoothness properties to recover the lost MV.

Almost every existing temporal error concealment method makes use of clean neighborhood frames to reconstruct the missing blocks. However, when all adjacent frames are violated with error blocks, the candidate motion estimation of the missing block as well as the motion estimations of all its neighborhoods may be lack of information in some parts, hence fail to successfully recover the lost regions. This paper will introduce a more powerful algorithm that can tackle this problem. In this proposed approach, each of the missing block is divided into smaller pieces and these pieces are grouped with their spatial neighborhoods to form new patches for recovery. Relying on the partial information of these new patches, the algorithm strives to find the predicted MVs over multiple frames. Since the adjacent frames are also damaged, the found motion estimations may be incomplete with many missing elements. All the matched motion compensations should have similar underlying image structures and the completed version of these patches lie in a very low dimensional subspace. Therefore, by vectorizing the stack of motion compensations and re-arranging into columns of a matrix, such matrix becomes an incomplete version of a low-rank matrix with many missing indexes. Accordingly, the problem of error concealment is tranformed to the problem of completing a low rank matrix from its clean observation.

The rest of the paper is organized as follows. In section 2, we describe the formulation of video error concealment based on low rank matrix completion. The detail algorithm of matrix completion is also discussed in this part. Section 3 evaluates the experiments of the proposed method and compare the results with the state-of-the-art BMA method. We conclude this paper in section 4 and propose some future improvements for the algorithm.

II. Error concealment algorithm using matrix completion

In this section, we describe in detail the proposed error concealment algorithm by two main steps. In the first step, a low-rank matrix with a lot of missing will be effectively constructed based on the high correlation in the temporal domain of video sequences. The second step deals with completing the lost parts in the matrix solved by minimizing the nuclear norm of the matrix with linear constrains.

A. Problem formulation and low-rank matrix construction

Consider a video sequence corrupted by missing blocks in every frame. Let f be the current frame with a number of error blocks that we want to recover. Denote $\mathcal{F}=\{f_k|k=1,2,...K\}$ as the set of all neighborhood frames that we will use to recover the frame f. The current frame f contains n erroneous blocks and each frame f_k also has n_k missing blocks which we assume to be set to zero. Suppose that we know the locations of all these missing blocks. The goal of the concealment algorithm is to recover the set of all the lost blocks $B=\{b_i\}_{i=1}^{n}$in the current frame f. We setup the problem in the most general case when the locations of missing blocks can stay randomly in the frame. As the lost blocks can border to each other, large missing objects can appear in a frame, hence some blocks does not have enough information of the neighbor for recovery. However, if we complete the missing from the left to the right and from top to bottom of each frame, and only the information of the left and upper nearby blocks are used, we can deal with the problem of adjacent missing. The detail method will be discussed in the follows.

In the frame f, consider one image block b_i of size $N\times N$ (say $N = 16$). By dividing b_i into 4 smaller pieces $b_i=\cup_{j=1}^{4}p_{ij}$, we step by step recover each of these pieces p_{ij} and then complete the block b_i. The first quarter missing piece p_{i1} is grouped with the three quarters of the upper left neighborhood to form a new patch of the same size $N\times N$. The partial information from the spatial neighborhood and all adjacent frames are exploited to fill in p_{i1}. After p_{i1} is completed, it can be considered as the known information and the same technique to complete p_{i1} can be applied to recover p_{i2}. Then p_{i3} and finally p_{i4} are recovered in succession just like what we have done with p_{i1} and p_{i2}. The four steps to complete one missing block can be illustrated as in figure 1.

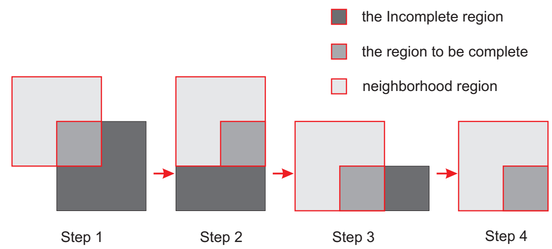

Figure 1. Illustration of completion steps

Now we will consider how one patch, called e_{ij}, with three quarters of known information and one fourth of the unknown piece p_{ij} can be completed. Setting this patch as a reference patch, we search within all neighborhood frames for the best motion estimations that are similar to e_{ij}. Only three fourth of known information in the patch e_{ij} are used for the comparison of mean square error (MSE) to find one motion estimation candidate. Due to the wide spreading of erroneous blocks in every frame, the estimation candidates are not expected to be full of clean indexes, but a lot of missing elements may be lost and only the clean ones are used for MSE comparison. One searched patch for motion estimation will not be counted if the number of missing elements within the patch exceeds 50% of the total number of indexes N^2.

Generally, assume that K motion estimations $\{e_{ijk}\}_{k=1}^{K}$ of e_{ij} are found in the temporal domain. By presenting e_{ij} as a vector $\mathbf{e}_{ij}\in\mathbb{R}^{N^2}$ and vecterizing each motion estimation e_{ijk} as $\mathbf{e}_{ijk}$, we define a $N^2\times(K+1)$ matrix E_{ij} as follows:

$$E_{ij}=\{\mathbf{e}_{ij},\,\mathbf{e}_{ij1},\,\mathbf{e}_{ij2},...,\,\mathbf{e}_{ijK}\} \quad (1)$$

The matrix E_{ij} contains a lot of missing part that we want to recover, however, it is always checked that every row of E_{ij} is not a full missing row. Otherwise, it can not be recovered by using the proposed matrix completion technique. Although there is a very high probability that in every row of E_{ij} there contains some clean elements, this property is not always true. Therefore once the algorithm detects one full missing row, we have to reduce the similarity with the reference patch of several candidate motion estimations by exchanging them with other patches but having available indexes in the row.

If there is clean frames every a certain number of frames (say 10 or 15), we will not have to worry about the case of full missing row. The motion estimation coming from the clean frame is put into the last column of the matrix E_{ij}. This column is always full, hence every row will

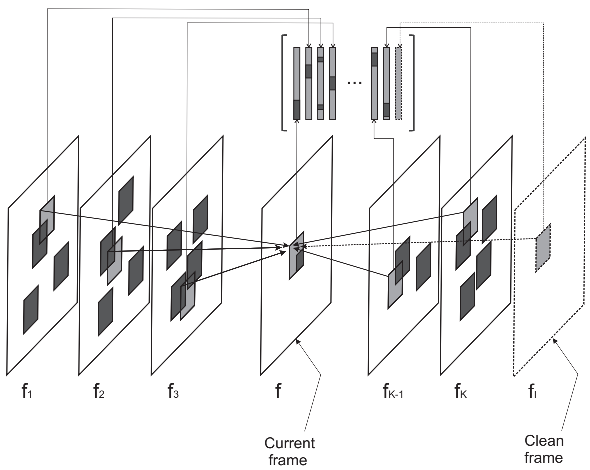

Figure 2. Low-rank matrix construction

contain at least one available element. In the worst scenario case when every other indexes in one row are missed, the last column play as the leading character to complete all remaining part in the slab. The construction of matrix E_{ij} is presented in figure 2.

B. Low-rank matrix completion

The matrix E_{ij} contains a lot of missing parts that we want to recover, and we can write it as the following form:

$$E_{ij} = F_{ij} + N_{ij} \tag{2}$$

where F_{ij} is the matrix from the clean image f and N_{if} is the missing version which can be considered as noisy form.

Intuitively, if all the frame is free of missing, the underlying structures of all columns of F_{ij} will be very similar and the approximate rank of matrix F_{ij} should be very low. In the ideal case, the singular value decomposition (SVD) shows the dominance of only one eigen vector and by simply running the SVD on E_{ij} we can easily find a good low-rank estimation of F_{ij}. However, when missing elements are presented, the SVD technique may fail due to the noise sensation of the approach. Therefore, we propose a more efficient way to robustly estimate F_{ij} relied on the fact that under mild conditions, a complete matrix can be exactly recovered from a small amount of elements. This method is called matrix completion technique by the pioneers Emmanuel J. Candès and Benjamin Recht [5].

It has been proved that if A is an $m \times n$ incoherent matrix that has $rank(r) \ll R = max(m, n)$, then A can be recovered with very high probability from a uniform sampling of its P entries, where $P \geq O(R^{1.2} rlog(R))$. In general the total number of missing elements in our constructed matrix E_{ij} is not that much. Since the approximation of E_{ij} lie in a very low dimensional subspace, and the number of missing blocks is small compared with the amount of the clean blocks in every frame, the number of samplings in our case always exceed the demand of matrix completion recovery.

Once we know that matrix completion can be done, the next question is how to develop one method. There are a variety of different matrix completion algorithms available. Candès et al. presented a method called Singular Value Thresholding (SVT) [6], which attempts to complete the matrix by solving the optimization problem to find a matrix X that minimizes the nuclear norm. The minimization is subject to the condition that the entries of X be equal to those entries of the matrix A to be completed for which we know the value. In [7], Keshavan, Montanari, and Oh offered an algorithm based on trimming the incomplete matrix to remove those values that do not help reveal much about the unknown entries and then adjusting the trimmed matrix to minimize the error at the entries whose values are known via a gradient descent procedure. Some other algorithms working on PCA method have recently been developed as well. In our work, we elected to use the fixed point iteration algorithm [9], since it minimize the approximate low-rank which is exactly our demand.

In general, the matrix rank minimization is NP-hard due to the combinational nature of the operator $rank(\bullet)$. Similar to the cardinality function $\|x\|_0$ in the compressive sensing problem, we can replace $rank(A)$ by its approximate convex envelope to get a convex and more computationally amenable optimization problem. Then the problem of minimizing the $rank$ yields the nuclear norm minimization problem:

$$\begin{aligned} & min \, \|A\| \\ & \quad s.t. \mathcal{F}(A) = b \end{aligned} \tag{3}$$

where the given vector b is the observation with small number of coefficients, and $\mathcal{F}$ is the one-to-one linear operator on the set of sparse matrix.

Back to our matrix completion algorithm, the task is to recover F_{ij} from the incomplete version of E_{ij}, denoted by $E_{ij}|_I$ where I is an index set of all available elements corresponding to the clean pixels and $E_{ij}|_I$ is the vector including elements in I of E_{ij} only. Then the reconstruction of F_{ij} from its incomplete observation $E_{ij}|_I$ can be formulated as the following minimization problem:

$$\begin{aligned} & min \, \|F_{ij}\|_* \\ & \quad s.t. \, \|F_{ij}|_I - E_{ij}|_I\|_F^2 \leq \tau \end{aligned} \tag{4}$$

or its Lagrangian version:

$$\min_{F_{if}} \|F_{ij}|_I - E_{ij}|_I\|_F^2 + \lambda \|F_{ij}\| \tag{5}$$

where τ is an relaxing threshold to control the accuracy of the matrix completion, $\|F_{ij}\|_*$ is the nuclear norm of the matrix F_{ij} and $\|\|_F$ denote a matrix Frobenious norm. In the unconstrained formulation (5), the parameter λ is chosen in such a way that the solution of (5) still satisfies the constrain of (4). In our implementation, the fixed point iterative algorithm [9] is used to solve (5).

III. Experiment results

The experiment will intentionally compare the performance of the proposed algorithm with the classical BMA, the standard concealment method used in H.264 reference software. The algorithm is tested on the video sequence Foreman in CIF format. Every frame is corrupted by a number of missing blocks and the loss rate is set equally to 10% in random position in every frame. The block size is set to be $N = 16$. The number of neighborhood frames used for the search of motion estimation in the case of whole-corrupted-frame is $K = 10$ in which 5 come from the previous frames and 5 from the future frames. The sequence is suffered from quantization effect by transforming the original sequence to the DCT domain and quantizing with a certain step size. The quantization step size is set to be 4, which explains why the PSNRs of all key frames at the decoder are around 47 dB.

The algorithm is tested for both cases with and without I-frame. When no I-frame is used, meaning that every frame contains missing blocks, the output sequence shows a quite good PSNR. However, due to the fact that we are not always guaranteed of having no full missing row in the constructed low-rank matrix, there are still scattered blocks that can not be correctly recovered, yielding some flicker in the completed video sequence. However, when we set I-frames in every 10 frames of the sequence, the result is much better. The output sequence shows almost no flicker and the visual quality can be compatible with the original video. Figure 3 shows the comparison of the algorithm with and without I-frame.

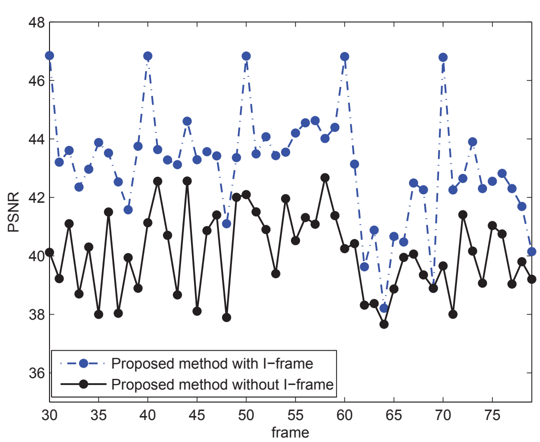

Figure 3. Comparison of the algorithm with and without I-frame

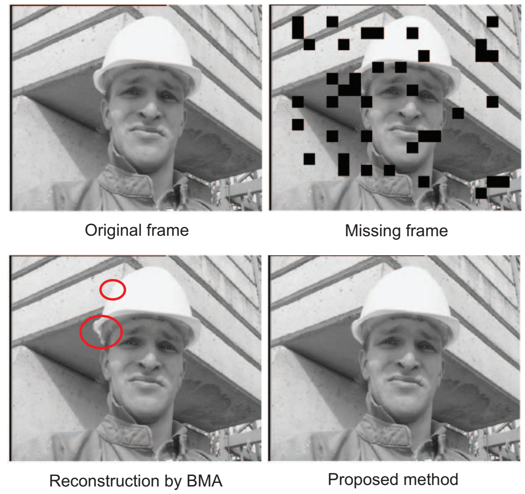

Figure 4. Example of the quality comparison between the proposed method and BMA algorithm.

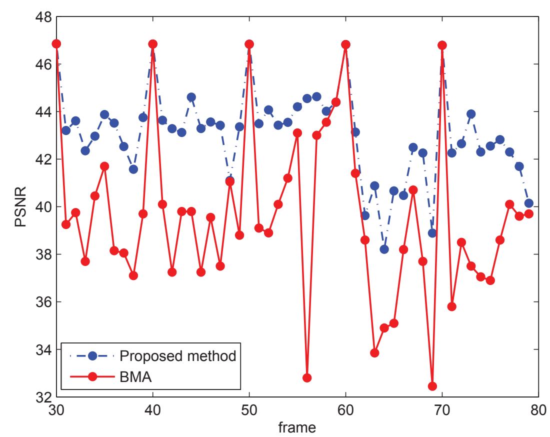

Figure 5. Comparison of the proposed algorithm with BMA algorithm with I-frame

When having no I-frame at the receiver, the BMA method does not work. This shows the potential of our method, since the existence of clean I-frame is not always the case, especially when all packets in the transportation have the same protection. Furthermore, in case I-frame is set up for both the BMA and the proposed algorithm in every 10 frames, the proposed method also achieves much better reconstruction. On average, the improvement is 3.1dB over the BMA method in the whole sequence. Figure 4 displays one example result and figure 5 shows the comparison of the proposed method with the BMA algorithm.

IV. Conclusions

In our paper, we implemented a novel algorithm for the video error concealment problem using the matrix completion technique. By grouping all the candidate motion vectors from previous and future frames, including full and corrupted patches, we construct an approximate low rank matrix with a lot of missing elements. This matrix can be completed by solving the optimization to minimize the nuclear norm with linear constrains. The algorithm has a lot of robustness properties. It not only can deal with the random sizes and random locations of missing blocks, but also can complete the a video sequence having all corrupted frames. One more power of the proposed algorithm is that we can access to recover from any frame, hence supporting for the parallel computation. We can even break the task in such a way that each computer can recover a certain number of frames. The algorithm is currently exploiting the correlation only in the temporal domain. Therefore the sequence is not well completed when the scenes change very fast. For future improvement, adding patches in the spatial domain to the dictionaries would improve the reconstruction in the texture area.

References

[1] P. Haskell and D. Messerschmitt, "*Resynchronization of motion compensated video affected by ATM cell loss*" in Proc. ICASSP-92: 1992 IEEE Int. Conf. Acoustics, Speech, Signal Process., San Francisco, CA, 1992, vol. 3, pp. 545–548.

[2] W. M. Lam, A. R. Reibman, and B. Liu, "*Recovery of lost or erroneously received motion vectors*" in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process., 1993, vol. 3, pp. 417–420.

[3] Y. Chen, O. Au, C.-W. Ho, and J. Zhou, "*Spatio-temporal boundary matching algorithm for temporal error concealment*", in Proc. IEEE Int. Symp. Circuits Syst., 2006, pp. 686–689.

[4] Y. Chen, Y. Hu, O. Au, H. Li, C. Chen, *"Video error concealment using spatio-temporal boundary matching and partial differential equation*", IEEE Trans Multimedia, 2008, 10(1): 2–15.

[5] E. J. Candes and B. Recht, *"Exact matrix completion via convex optimization"*, Found. of Comput. Math, 2008, 9:717–772.

[6] J-F Cai, E. J. Cand'es, and Z. Shen, "*A singular value thresholding algorithm for matrix completion*", arXiv: 0810.3286, 2008.

[7] Keshavan, R.H., Montanari, A., Oh, S.:, *"Matrix Completion from a Few Entries"*, arXiv:0901.3150v4, September 2009.

[8] H. Ji; C. Liu; Z. Shen; Y. Xu, *"Robust video denoising using low rank matrix completion",* Computer Vision and Pattern Recognition (CVPR), 2010.

[9] S. Ma, D. goldfarb, and Z.Wan., *"Fixed point and bregman iterative methods for matrix rank minimization*", Mathematical Programming Series A, 2009.