VIDEO FRAME INTERPOLATION VIA WEIGHTED ROBUST PRINCIPAL COMPONENT ANALYSIS

Minh Dao [1], Yuanming Suo [1], Sang Chin [2], Trac Tran [1]

[1]Department of Electrical and Computer Engineering, The Johns Hopkins University, MD 21218
[2]Applied Physics Laboratory, The Johns Hopkins University, MD 21723

ABSTRACT

In this paper, we propose a new video frame interpolation technique by a locally-adaptive robust principal component analysis (RPCA) with weight priors. The proposed algorithm relies on two main steps: 1. the pre-processing step initializes the new frame by a simplified motion-compensated frame interpolation and assigns each pixel a confident weight based on both the difference of motion estimation and local consistency; and 2. the refinement step updates the frame by a proposed weighted robust principal component analysis (WR-PCA) algorithm. Experiments demonstrate that the proposed method outperforms the state-of-the-art algorithms, both in visual quality and PSNR performance.

***Index Terms*—** Frame interpolation, motion-compensated interpolation (MCI), robust principal component analysis (RPCA).

1. INTRODUCTION

Video frame interpolation is the technique that enhances the temporal resolution of low bit rate coded videos by inserting intermediate frames into the existing sequences. The primitive methods such as frame repetition or linear frame interpolation which do not need to rely on motion estimation (ME) provide reasonable quality for low-complexity motions but fail with fast moving objects. Motion compensated interpolation (MCI) [1] and the related techniques have been developed to take into account ME for better frame interpolation. This concept has been used by most of the state-of-the-art frame interpolation algorithms. However, the minimum error block estimation may not always provide true motions because of repetitive structures or fast changing objects; which introduces blockiness or ghost artifacts in the interpolated frames. Therefore, the common approaches for these MCI-based methods are to exhaustively constrain the true motions by forcing them to be smooth in neighborhood regions. One successful way is the use of block-matching algorithm (BMA) and mesh-based motion compensation [2] to refine motion distribution on object boundaries. Another approach is the utilization of image segmentation [3] to force all blocks in a defined salient region to share the same motion vectors (MV). A number of other approaches have been developed to tackle this problem, but only few of them have been successful in effectively reducing block artifacts and still preserving structural details in the images.

The recent emergence of matrix recovery techniques has opened a new trend in solving interpolation and completion problems. Missing or corrupted elements can be robustly recovered by structural constraints of the input data like sparsity or low-rank structure. Matrix completion [4] is a low-rank matrix recovery technique that can robustly recover low-dimensional structures from high-dimensional observations, especially for scenarios where the data is highly missing. This technique has been proved to efficiently solve video completion problems like video inpainting or error concealment by incorporating the spatial and temporal correlations of the frames to construct low-dimensional structures which are normally in the forms of low-rank matrices [5] or low-rank tensors [6]. However, these methods can only achieve pleasing reconstructions for the partial incompletion in every frame. No low-rank matrix recovery methods have been promoted to recover full-frame missing so far.

Robust principal component analysis (RPCA) [7] is another well-known low-rank matrix recovery framework that can recover the low-rank structures from severely corrupted high-dimensional data. This method has proved its robustness in solving background modeling, target tracking, or image alignment problems [7, 8] where signals normally contain abnormal objects lying on a low-dimensional background. However, as far as the knowledge of the authors, this model has not been applied in solving any video completion problem. Furthermore, the general assumption of most of the applications is the global low-rank property of the sequence which is only true for the case the background of the sequence is unchanged i.e. the camera capturing the scene is fixed. Moreover, the traditional RPCA model tends to treat all pixels equally which should not always be the case since in some situations we can get prior-knowledge of where the outliers are more certain to present.

This work has been supported in part by the National Science Foundation (NSF) under Grant CCF-1117545, the Army Research Office (ARO) under Grant 60219-MA, and the Office of Naval Research (ONR) under Grant N000141210765.

Under the above observations, we propose a novel algorithm which can incorporate the RPCA technique into video frame interpolation. The system involves a preprocessing step to provide some reference information to guide the RPCA model in refining the new frame. This preprocessing scheme not only generates an initialized frame by a simplified bidirectional ME but also produces a weighting matrix describing the confidence of all introduced pixels obtained by this step. In the second step, a weighted robust principal component analysis (WRPCA) is proposed to refine every pixel in the intermediated frames. The pixels corresponding to true MEs are more likely to be preserved while the pixels resulted from incorrect MEs are automatically detected and updated.

The remaining of the paper is organized as follows. In section 2, we give a brief overview of RPCA technique and introduce the proposed WRPCA as well as the detailed algorithm. The next section formulates the frame interpolation via WRPCA in a two-stage problem. Section 4 evaluates the experimental results and we conclude this paper in section 5.

2. RPCA ALGORITHM WITH WEIGHT PRIORS

Robust principle component analysis (RPCA) is a highly applicable low-rank matrix recovery problem recently introduced by Candes et al [7] where the goal is to accurately recover an underlying low-rank matrix from its sparse but grossly corrupted entries. Mathematically, let $\mathbf{A}$ be a low-rank data matrix. It frequently happens that we are not able to observe $\mathbf{A}$ directly; instead we observe its corrupted version $\mathbf{D} = \mathbf{A} + \mathbf{E}$. The matrix $\mathbf{E}$ captures outliers, assumed to be sparse but can have arbitrarily large magnitudes. To separate $\mathbf{A}$ and $\mathbf{E}$, one would like to find the simplest model that fits the low-rank observations [7]:

$$\underset{\mathbf{A},\mathbf{E}}{Min}\, Rank(\mathbf{A}) + \lambda \left\|\mathbf{E}\right\|_0 \quad s.t.\, \mathbf{D} = \mathbf{A} + \mathbf{E} \tag{1}$$

where the l_0-norm $\left\|\mathbf{E}\right\|_0$ is defined as the number of nonzero entries in $\mathbf{E}$ and λ is a positive weighting parameter.

The above rank and l_0-norm minimization problem is an NP-hard problem. Under some mild conditions, however, the l_0-norm can be efficiently solved by recasting it as a convex l_1-based linear programming problem and the intractable rank-minimization can be relaxed to the convex problem of nuclear norm minimization:

$$\underset{\mathbf{A},\mathbf{E}}{Min}\, \left\|\mathbf{A}\right\|_* + \lambda \left\|\mathbf{E}\right\|_1 \quad s.t.\, \mathbf{D} = \mathbf{A} + \mathbf{E} \tag{2}$$

where the l_1-norm is defined as $\left\|\mathbf{E}\right\|_1 = \sum_i |e_i|$ with e_i's being the entries of $\mathbf{E}$ and the nuclear norm $\left\|\mathbf{A}\right\|_*$ is the sum of all singular values of the matrix $\mathbf{A}$.

The model (2) treats the sparse noise distribution at all locations equally. However, when we have some prior knowledge of where the noise $\mathbf{E}$ is more likely to appear, we can provide better representations of the measurements by introducing a weighting matrix into the formula. The WRPCA algorithm is modeled as the following minimization:

Algorithm 1 WRPCA algorithm

Inputs: Data input $\mathbf{D}$, weighting matrix $\mathbf{W_E}$

1. $\mathbf{Y}_0 = 0$, $\mathbf{E}_0 = 0$, $\mu_0 > 0$, $\rho > 1$
2. **While** not converged, **do**
3. $(\mathbf{U}, \mathbf{S}, \mathbf{V}) = svd(\mathbf{D} - \mathbf{E}_k + \mu_k^{-1}\mathbf{Y}_k)$;
4. $\mathbf{A}_{k+1} = \mathbf{U}S_{\mu_k^{-1}}[\mathbf{S}]\mathbf{V}^T$;
5. $\mathbf{E}_{k+1} = S_{\mu_k^{-1}\mathbf{W_E}}(\mathbf{D} - \mathbf{A}_{k+1} + \mu_k^{-1}\mathbf{Y}_k)$
6. $\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_k(\mathbf{D} - \mathbf{A}_{k+1} - \mathbf{E}_{k+1})$
7. $\mu_{k+1} = \rho\mu_k$
8. $k = k + 1$
9. **end while**

where $S_\epsilon(\mathbf{X})$ is the soft-thresholding operator defined for each element seperately: $S_\epsilon(x) = max(|x| - \epsilon, 0)\, sgn(x)$.

Outputs: $(\mathbf{A}_k, \mathbf{E}_k)$.

$$\underset{\mathbf{A},\mathbf{E}}{Min}\, \left\|\mathbf{A}\right\|_* + \left\|\mathbf{W_E}\circ\mathbf{E}\right\|_1 \quad s.t.\, \mathbf{D} = \mathbf{A} + \mathbf{E} \tag{3}$$

where the matrix $\mathbf{W_E}$ is the corresponding weight priors for noise support distributions, and $\circ$ denotes the Hadamard (pointwise) product. If we set $\mathbf{W_E} = \lambda\mathbf{1}$ then the RPCA problem (2) is recovered.

The WRPCA model can be efficiently solved by a number of methods including iterative thresholding (IT) [9], accelerated proximal gradient (APG) and augmented Lagrange multipliers (ALM) [10] in which ALM method is preferable because of the fast convergence property. The augmented Lagrangian function of (3) is expressed as:

$$\begin{aligned} \mathcal{L}(\mathbf{A},\mathbf{E},\mathbf{Y}) = &\left\|\mathbf{A}\right\|_* + \left\|\mathbf{W_E}\circ\mathbf{E}\right\|_1 \\ &+ \langle \mathbf{D} - \mathbf{A} + \mathbf{E}, \mathbf{Y}\rangle + \frac{\mu}{2}\left\|\mathbf{D} - \mathbf{A} + \mathbf{E}\right\|_F \end{aligned} \tag{4}$$

Here $\mathbf{Y}$ is the Lagrange multiplier, μ is a positive variable, $\langle\cdot\rangle$ is the inner product and $\|.\|_F$ denotes the Frobenius norm. The minimization of the augmented Lagrangian function can be solved iteratively by fixing one variable and update the other. The detailed WRPCA algorithm is described in the algorithm 1.

3. FRAME INTERPOLATION VIA WRPCA

The proposed algorithm to upscale the temporal frequency of a sequence is the combination of two main steps. In the preprocessing step, the interpolated frame is initialized by using a simplified version of the bi-directional MCI technique. In the second step, all pixels in each block are automatically updated by the proposed WRPCA algorithm with a defined confidence matrix obtained by the initialization step.

3.1. Adapted bi-directional MCI with weight calculation

The initialized frames are approximated by the block-based bi-directional motion estimation. Without the loss of generality, we assume that all the odd frames in a sequence are known

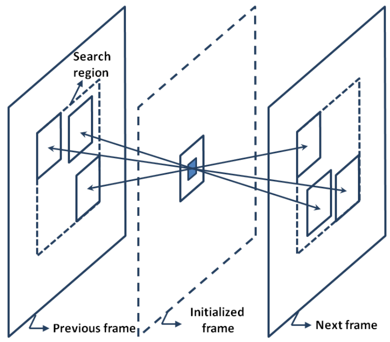

Fig. 1. Bi-directional motion compensation initiallization

and all the even frames are to be interpolated. Each even frame will then be initialized by only using the two nearby frames. Let $\mathbf{F}$ be the unknown frame that we want to interpolate, and $\mathbf{F}(p)$ be the intensity at the pixel p in that frame. Denote $\mathbf{F}_1$ and $\mathbf{F}_2$ as the previous and future frames. For the block $\mathbf{B}$ centered at each pixel p in the to-be-interpolated frame, we search for the best match motion trajectory in a specific searched region from a block $\mathbf{B}_1$ in the frame $\mathbf{F}_1$ to a block $\mathbf{B}_2$ in the frame $\mathbf{F}_2$, passing through the block $\mathbf{B}$. The initial intensity of the pixel p is then simply calculated by averaging the intensity levels of the center pixels in the blocks $\mathbf{B}_1$and $\mathbf{B}_2$. Figure 1 depicts this process of approximating the interpolated frame by motion compensation.

Along with calculating all the pixels in the frame $\mathbf{F}$, we also calculate the weighting matrix $\mathbf{W}$ whose each element is considered as the confidence of the previous output. The larger a component of $\mathbf{W}$ is, the more certainty it is a correct motion vectors. Therefore, it will be less likely to be detected as a noise in the second step of correcting the interpolated frame. The weighting $\mathbf{W}(p)$ at each pixel p is the linear summation of the two terms:

$$\mathbf{W}(p) = \mathbf{W}^{(M)}(p) + \gamma \mathbf{W}^{(C)}(p) \tag{5}$$

where $\mathbf{W}^{(M)}(p)$ is calculated as one over the sum of absolute difference (SAD) of the best motion estimation going through that pixel, $\mathbf{W}^{(C)}(p)$ is decided based on the consistency of the motion vectors at the pixel with its eight surrounding neighborhoods, and $\gamma > 0$ is a parameter to balance the two terms . The matrix $\mathbf{W}$ has less weight at the location where we cannot find a well-matched motion estimation, as well as the motion vector going through its location is very different with those passing its nearby pixels.

3.2. Weighted RPCA refinement

The outputs of the initialization step is not only the rough interpolation $\mathbf{F}$, but also the weighting confident matrix $\mathbf{W}$ which will be used in the refinement WRPCA algorithm to amend the frame at every pixel. Frame $\mathbf{F}$ is now divided into blocks of the same size. For each block $\mathbf{F}^{ij}$ of size $N \times N$ (say N=16), we search for similar patches from the closed frame to construct a low-rank structure. However, the searching process is not constrained in the right previous or next frames, but can be extended to any frame in a temporal region.

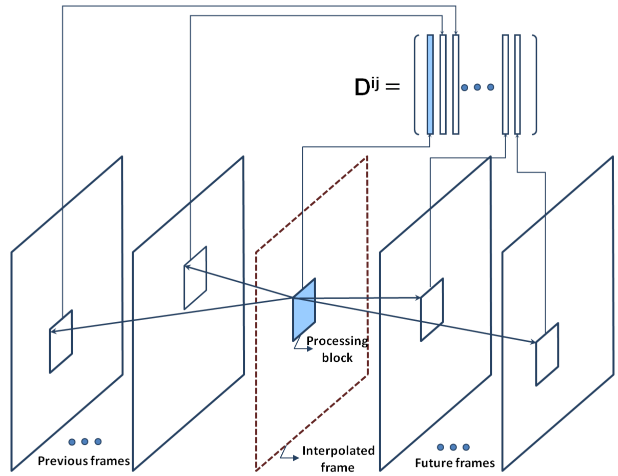

Fig. 2. Low-rank matrix construction

By presenting $\mathbf{F}^{ij}$ as a vector $\mathbf{d}^{ij} \in \mathbb{R}^{N^2}$ and vectorizing each motion estimation as $\mathbf{d}_l^{ij} \in \mathbb{R}^{N^2}$ $(l = 1, 2, ...L)$, we stack these vectors into columns of an $N^2 \times (L+1)$ low-rank matrix $\mathbf{D}^{ij}$.

$$\mathbf{D}^{ij} = \left\{ \mathbf{d}^{ij}, \mathbf{d}_1^{ij}, \mathbf{d}_2^{ij}, ..., \mathbf{d}_L^{ij} \right\} \tag{6}$$

The first column of $\mathbf{D}^{ij}$ is the block to be reconstructed in vector form and the remaining columns are the matching blocks. Intuitively, the underlying structure of all columns of $\mathbf{D}^{ij}$ should be similar thus the matrix $\mathbf{D}^{ij}$ becomes a very low-rank structure. The low-rank matrix construction is described in figure 2.

After constructing the low-rank matrix $\mathbf{D}^{ij}$, we expect that some of the elements in $\mathbf{D}^{ij}$ are outliers in contributing the low-rank structure of $\mathbf{D}^{ij}$, especially the atoms in the first column: $\mathbf{D}^{ij} = \mathbf{A}^{ij} + \mathbf{E}^{ij}$. These outliers in $\mathbf{E}^{ij}$ come from the uncorrected motion compensation process in step one and will be automatically detected and updated in the RWPCA algorithm. The other columns of $\mathbf{D}^{ij}$ are the estimations from the clean frames. Therefore, the probability that they behave as sparse noise is very low. With this analysis, we define the weighting matrix $\mathbf{W}^{ij}$ with the first column as the corresponding weightings in $\mathbf{W}$ of the elements from the first column of $\mathbf{D}^{ij}$, and the remaining columns having some small nonnegative value. Then the RWPCA algorithm for frame interpolation of block $\mathbf{F}^{ij}$ is formulated as the following optimization:

$$\underset{\mathbf{A}^{ij},\mathbf{E}^{ij}}{Min} \left\|\mathbf{A}^{ij}\right\|_* + \left\|\mathbf{W}^{ij} \circ \mathbf{E}^{ij}\right\|_1 \quad s.t.\, \mathbf{D}^{ij} = \mathbf{A}^{ij} + \mathbf{E}^{ij} \tag{7}$$

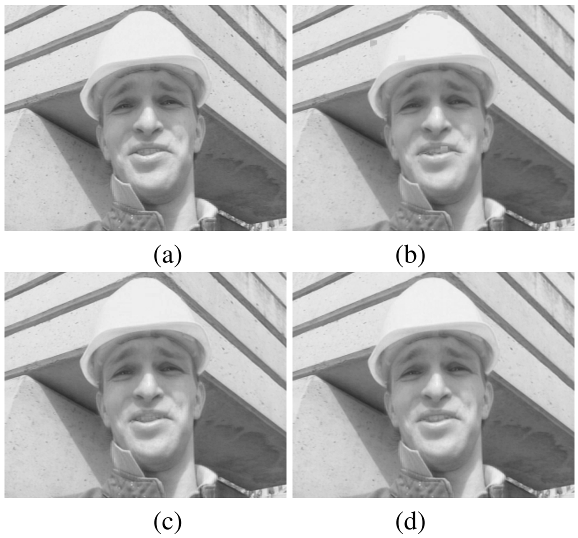

Fig. 3. Interpolation of frame #50 in Foreman.cif sequence: (a) Original frame (b) Bi-directional MCI, PSNR = 32.60 dB (c) Saliency MCI, PSNR = 34.2 dB (d) Proposed WRPCA, PSNR = 35.68 dB

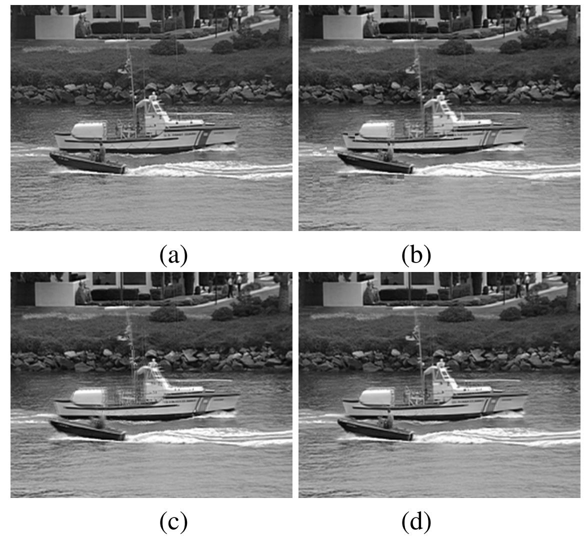

Fig. 4. Interpolation of frame #80 in Coastguard.cif sequence: (a) Original frame (b) Bi-directional MCI, PSNR = 26.24 dB (c) Saliency MCI, PSNR = 27.87 dB (d) Proposed WRPCA, PSNR = 28.86 dB

4. EXPERIMENTAL RESULTS

In this section, some experiments are conducted to evaluate the performance of the proposed frame interpolation algorithm. The method is compared with two other frame rate-up conversion techniques: bi-directional motion compensation interpolation (bi-directional MCI or BMCI) [1] and the MCI-based method using discriminant saliency and frame segmentation (saliency MCI) [3]. Three sequences in CIF format: "Foreman", "Coastguard" and "Tennis" are tested to compare both PSNR performance and visualization.

All the even frames in each video sequence encoded at 24 frames per second (fps) are temporally eliminated to reduce the number of frames by a factor of two to 12 fps. The missing frames are then interpolated by the three methods and compared with the originals. The block size used in the experiment is fixed at 16×16 . Three nearest frames in both direction, previous and future, are used as the references in the low-rank construction step.

Sequences	BMCI	Saliency MCI	Proposed
"Foreman"	32.31	33.85	**34.52**
"Coastguard"	30.60	32.05	**32.73**
"Tennis"	26.18	28.03	**28.48**

Table 1. PSNR performance comparisons with other frame interpolation algorithms.

Figure 3 present the visual comparisons from frame 50th in "Foreman.cif" sequence. The result from bi-directional MCI shows blockiness artifacts because of incorrect MEs, while the saliency MCI and the proposed methods seem to introduce comparable visualization with the original frame. For frame 80th in "Coastguard.cif" sequence experiment, the same blockiness flickers appear in bi-directional MCI while saliency MCI brings some blurring artifacts, especially at the regions with fast object movements. The proposed WRPCA method tends to keep the sharpness detail while noticeably reduces any blockiness or image damage regions in the interpolated frame.

Table 1 summarizes the average PSNR performance of the three test sequences. Only the frames to be interpolated are used to calculate the PSNR with the primary sequences. The table shows that the proposed algorithm performs the best in all test sequences with the PSNR improvement of up to 2.3dB compared to bi-directional MCI and 0.7dB compared to the state-of-the-art saliency MCI algorithm.

5. CONCLUSIONS

In this paper we proposed a novel approach to effectively incorporate low-rank robust principal component analysis technique into video frame interpolation problem. The interpolated frame is initialized by a simple MCI-based method which calculates the intensity value at every pixel together with a weight of how confident that pixel is plausibly resulted from a true motion estimation. The frame is then polished by a locally-adaptive weighted robust principal component analysis algorithm. This new approach, while exhibits superior performance to state-of-the-art methods in both visual quality and PSNR values, also offers a potential way in solving a number of other video completion problems.

6. REFERENCES

[1] B-D Choi, J. W Han, C. S. Kim & S. J. Ko, "New frame rate up-conversion using bi-directional motion estimation," *IEEE Transactions on Consumer Electronics,* 46(3), 603-609, 2000.

[2] B-D Choi, J. W Han, C. S. Kim & S. J. Ko, "Frame rate up-conversion using perspective transform," *IEEE Transactions on Consumer Electronics*, 52(3), 975-982, 2006.

[3] N. Jacobson, Y. L. Lee, V. Mahadevan, N. Vasconcelos, & T. Q. Nguyen, "A novel approach to fruc using discriminant saliency and frame segmentation," *IEEE Transactions on Image Processing*, 19(11), 2924-2934, 2010.

[4] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization", *Foundations of Computational Mathematics*, vol. 9, pp. 717-772, 2008.

[5] M. Dao, D. Nguyen, Y. Cao, and T. D. Tran, "Video concealment via matrix completion at high missing rates," *Proc. IEEE 44th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, Oct. 2010.

[6] D. Nguyen, M. Dao, and T. D. Tran, "Error concealment via 3-mode tensor approximation," *Proc. IEEE Int. Conf. on Image Processing*, pp. 2081-2084, Brussels, Sep. 2011.

[7] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis", *Journal of the ACM*, 58(3), 2011.

[8] Y. Peng, A. Ganesh, J. Wright, and Y. Ma, "Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[9] J. F. Cai, E. J.Candes, & Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, 20(4), 1956-1982, 2010.

[10] Z. Lin, M. Chen, & Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.