VIDEO SUPER-RESOLUTION USING LOW RANK MATRIX COMPLETION

Jin Chen, Jose Nunez-Yanez, and Alin Achim

Vision information Lab, University of Bristol, UK
e-mail: {J.Chen, J.L.Nunez-Yanez, Alin.Achim} @bristol.ac.uk

ABSTRACT

In this paper, a novel video super-resolution image reconstruction algorithm is proposed. We design a patch-based low rank matrix completion algorithm. The proposed algorithm addresses the problem of generating a high-resolution (HR) image from several low-resolution (LR) images, based on sparse representation and low-rank matrix completion. The approach represents observed LR frames in the form of sparse matrices and rearranges those frames into low dimensional constructions. Experimental results demonstrate that, high-frequency details in the super resolved images are recovered from the LR frames. The gains in terms of PSNR and SSIM are significant.

Index Terms—Low-rank Matrix Completion, Video Super-Resolution, Singular Value Thresholding

1. INTRODUCTION

Super-Resolution techniques are widely utilized in many fields, and it has become a popular research topic over the past decade. Due to hardware limitations, physically increasing the pixel density of the charge-coupled device (CCD) arrays can be very expensive [1], therefore research has favoured using SR algorithms to increase the resolution of the observed data.

Tsai and Huang proposed the first multi-frame SR method in [2], a frequency domain based approach. Later research extended the frequency domain approach to spatial domain, and proved that spatial based SR approaches can overcome some drawbacks of the frequency domain method. Earlier spatial SR methods include: IBP (iterative back projection) [3], POCS (projection onto convex sets) [4] and ML (maximum likelihood) [5]. This was later followed by the Bayesian framework for SR. The advantage of Bayesian methods is that they include an explicit prior, where the main idea is to utilize a constraint distribution which is as close as possible to the distribution of the original scene [6].

Recovering a low rank matrix from given observations is a recurring problem in dimensionality reduction and compressive sensing. In [7], a single frame super-resolution was proposed. In their work, the authors implicitly determined the order of the linear model by minimizing the rank of the targeted matrix, which also reflects the order of the region aware linear model. In order to address the multi-frame denoising problem, researchers formulated the problem of removing salt&pepper noise as a low-rank matrix completion problem [8]. The authors firstly grouped similar noise patches in the spatial and temporal domain, and minimized the nuclear norm of the formulated low-rank matrix.

In this paper, we propose a matrix completion multi-frame super resolution method. Unlike most existing video super resolution techniques which rely on a single statistical model of the system, our method is; derived with minimal assumption on the observation models. The fundamental idea of the proposed method is; to treat the LR frames pixels as the observations in a incompleted matrix (HR frames), and to fill out the missing pixels through minimizing the nuclear norm of this incomplete matrix.

The remaining of the paper is organized as follows: Section 2 briefly introduces the matrix completion idea, in particular the Singular Value Thresholding (SVT) algorithm. We formulate the matrix completion super resolution problem in Section 3, and provide details of our algorithm. Results for super-resolved video frames are presented in Section 4. Finally, we conclude the paper in Section 5.

2. MATRIX COMPLETION

2.1. Matrix Completion Fundamentals

Suppose $M \in R^{m_1 \times n_2}$ is a low rank matrix. However, the observations of the matrix $M$ is a sampled set of entries, denoted as $\hat{M}_{i,j}, (i,j) \in R^O$, where $R^O$ is a subset of the complete set of entries. From here we use $X$ to represent the matrix to be recovered and $Y$, the observations, are a linear projection of $X$.

$$R_O(X) = Y \quad (1)$$

Where $R_O(\cdot)$ denotes the sampling process. Therefore, the original matrix $M$ can be recovered by solving the following optimization problem:

$$\text{minimize rank}(X); \text{ s. t. } R_O(X) = Y \quad (2)$$

This is an NP-Hard problem. Moreover, all known algorithms for solving this are doubly exponential in the-
ory [9] [10]. Alternatively, using the convex relaxation of Eq. (2) is an acceptable alternative.

\[
\min \|X\|_* \text{ s.t. } R_O(X) = Y
\]

where \(\|\cdot\|_*\) is the nuclear norm. It was already proven in [11], that nuclear-norm minimization is the tightest convex relaxation of the NP-hard rank minimization problem.

2.2. Singular Value Thresholding algorithm

Many algorithms have been proposed to solve the nuclear-norm minimization problem as in (3). Among them, Cai proposed an efficient first order algorithm [11], when the matrix \(X\) possess a strictly low rank character. Instead of optimizing directly, the Singular Value Thresholding algorithm solves the problem in (3) as:

\[
\min \tau \|x\|_* + \frac{1}{2}\|x\|_F^2 \text{ s.t. } R_O(X) = R_O(M)
\]

The algorithm can be inductively defined:

\[
\begin{align*}
X^k &= D_\tau(Y^{k-1}); \\
Y^k &= Y^{k-1} + \delta_k R_O(M - X^k);
\end{align*}
\]

In the above equation, \(D_\tau\) denotes the shrinkage operator, which is also known as the soft-thresholding operator, and \(\delta\) is the step size.

3. SUPER RESOLUTION BY MATRIX COMPLETION

3.1. Super resolution problem formulation

We assume that the set of low resolution frames are obtained from their corresponding high resolution frame. The LR frames are denoted as \(S = \{S_k\}, k = 1, ..., K\), each \(S_k\) corresponds to a LR image frame. Let the HR image be denoted as \(G\), then we have:

\[
S = HG + \text{Noise}
\]

Where \(G\) is the Nyquist sampled HR image of the scene and \(S\) represents the observed LR images. The sparse matrix \(H\) is generally used to represent the blurring and downsampling process. In this paper, only downsampling is considered. The goal of super resolution is to recover frame \(G\) from the observed \(S\). Assume the size of LR frame is \(P \times Q\), then the size of the matrix \(S\) is \(P \times Q \times K\). In order to avoid large matrix manipulations, we are taking a patch-based approach. For every patch in the HR frame \(G\), there are \(K\) correspondent LR patches, denote as \(P_{n,k}, 1 \leq k \leq K\). The size of HR patch, which is denoted by \(P_h\), is \(Fn \times Fn\), \(F\) is the resolution factor and \(n\) is the size of LR patch, which is usually set to be 4. The goal of matrix completion super resolution is for a specific HR patch to obtain \(K\) LR correspondent patches. We should be able to construct a sparse low rank matrix with those similar LR patches and by reducing the rank of the given matrix, the missing values are estimated.

We use LR-HR to represent the LR to HR projection. In the ideal situation, if we project the LR patches onto the HR grid ‘perfectly’, and assuming all the missing pixels are perfectly filled, those \(K\) LR-HR patches will be identical to each other. In this case, if we use a matrix \(E\) to represent all concatenated LR-HR patches, each row of this matrix is a vectorized LR-HR patch. Hence we can easily tell that \(E\) is low rank and the rank is one.

\[
E_{j,k} = (e_{j,1}, e_{j,2}, ..., e_{j,K})^T
\]

where \(j\) denotes \(j_{th}\) patch of a frame. In super resolution problems the noise term in (6) usually comes from the registration process, which relates to the non-perfect motion estimation between frames.

3.2. Matrix Augmentation

In the matrix completion approach to super resolution, we should notice that the size of the matrix is unchanged. In order to construct the target matrix, where the number of column is equal to the size of the HR patch, the LR patches need to be mapped on to the HR grid (LR-HR) before the further processing. This represents a significant difference compared with matrix completion video de-noising, where the size remains the same [8].

Fig. 1 is an illustration of the patch-based matrix completion super resolution (MCSR). In Fig. 1, (a), (b) and (c) are three LR-HR interpolated patches. This is obtained by expanding a LR patch in both vertical and horizontal directions \(F\) times. In this case \(F = 2\). Afterwards, the missing pixels in the expended LR frames are filled with zeros. We would like to note the shift between those expanded patches are due to the sub-pixel motion between the neighboring frames.

In Fig. 1 (d), we can see that each row of this matrix is linearly independent on the others. Therefore, the SVT matrix completion will not be suitable for this kind of matrix. In a real situation, the rank of such matrix will be equal to the resolution factor \(F\).

In order to build a linear dependent matrix according to the matrix \(E_{j,k}\), we augment every row of \(E_{j,k}\) horizontally with \(e_{m,j,k}'\). Where \(e_{m,j,k}'\) has the same size as the \(e_{j,k}\), but instead of zero padding, cubic interpolation is used to fill out the missing pixels in the LR-HR. This extended matrix is denoted as \(E_{j,k}' = (e_{1,j,k}', e_{2,j,k}', ..., e_{m,j,k}')^T\). Accordingly, in conjunction with the optimization method in Eq. (3), the problem becomes:

\[
\min \|X\|_* \text{ s.t. } H(X) = (E_{j,k}, E_{j,k})
\]
Fig. 1. (a) (b) (c) LR-HR patches (d) Vectorized and concatenated LR-HR patches

3.3. Algorithm Implementation

For a number of problems involving image sequences, motion estimation is always necessary. Like all other multi-frame super resolution problems, when obtaining a high resolution image or video sequence from a set of low resolution images, it is essential to have a good sub-pixel motion estimate. In [8], the author address the video de-noising problem. They adopt a fast three-step hierarchical search algorithm [12] to find the similar patches between frames.

In the preliminary steps of the proposed work, we adopt the Combined Local Global (CLG) optical flow [13]. This relatively new method was shown to outperform other general optical flow motion estimation method, and it has been used in our previously work [6] successfully. Once the sub-pixel motions are obtained, LR frames can hence be mapped onto the HR grid according to the motion and the initial LR-pixel motions are obtained, LR frames can hence be mapped onto the HR grid (LR-HR) based on the obtained motions.

1. Estimate the subpixel motion between frames, and project the LR frames onto the HR grid (LR-HR) based on the obtained motions.

2. Divide the motion compensated LR-HR frames into patches. For every targeted patch, we stack $K$ vectorized LR-HR patches into the matrix $E_{j,k}$, and discard the unreliable pixels by using the aforementioned process. Meanwhile, we obtain $E_{j,k}'$ to fill out the missing pixels in $E_{j,k}'$ by cubic interpolation. For every patch, we execute the following iterations.

3. Set $Y^0 = 0$. Iteratively do:

$$
\begin{cases}
X^{(k)} = Y^{(k)} - \tau R_\Omega (Y^{(k)} - (E_{j,k}, E_{j,k}')) \\
Y^{(k+1)} = D_\lambda (X^{(k)}),
\end{cases}
$$

where step size $\tau$ is pre-defined as $1 \leq \tau \leq 2$. The soft shrinkage operator is defined in [11]:

$$D_\lambda = U \Sigma_\lambda V^T$$

where $U$ and $V$ are two matrix with orthonormal columns, $\Sigma_\lambda = \text{diag}(\max(\rho_i - \lambda, 0))$, $\rho_i$ is the largest singular value. 4. Iteration terminates when $\|Y^{(k)} - Y^{(k-1)}\|_F \leq \epsilon$ or the iteration times reaches its maximum. The stopping criteria $\epsilon = 10^{-5}$ and the maximum iterations number of is 150.

5. When the iteration ends, we average every column of $X^k$, and acquire a size $1 \times 2(Fn)^2$ row vector, $X_p$. Then, the rearranged first $(Fn)^2$ elements of $X_p$ on the HR grid constitutes the super resolved patch. For the elements from $(Fn)^2 + 1$ to $(2Fn)^2$, if we do the same rearrangement, the matrix completion for spline patches is obtained.

Parameters of algorithm are chosen by default as; step size $\delta = 1.2$, the threshold $\lambda = (\sqrt{K} + \sqrt{2(nF)^2})/\sqrt{\delta}$, where $p$ is the ratio of the number of pixels detected as reliable pixels in $(E_{j,k}, E_{j,k}')$ [8]. In the above method, SVT is implemented on both $E_{j,k}$ and $E_{j,k}'$, therefore we obtain two MCSR results of a single patches after processing them at the same time.

4. EXPERIMENTAL RESULTS

Both the Peak Signal-to-noise ratio (PSNR) and the structural similarity (SSIM) [14] were used to assess the performance of our proposed algorithm. SSIM has been proven to be more consistent with human eye perception when compared to the PSNR and mean squared error (MSE). The SSIM is defined as:

$$SSIM(a,b) = \frac{(2\mu_a\mu_b + c_1)(2\delta_{ab} + c_2)}{\mu_a^2 + \mu_b^2 + c_1(\delta_a^2 + \delta_b^2+ c_2)}$$

where $\mu_a$ and $\mu_b$ are the mean values, $\delta_a$ and $\delta_b$ are their variances, and $\delta_{ab}$ is the covariance. The PSNR and SSIM improvement are showed in Table 1.
One real life video has been tested in simulations. The original video of size $168 \times 128$ is downsampled by a resolution factor $F = 2$. Empirically, 80 LR frames are used to obtain one HR frame ($K=80$). By default, patch size $n = 4$ and the resolution factor $F$ is equal to two. Therefore, the constructed targeted matrix $((E_{j,k}, E'_{j,k}))$ is with size $80 \times (F \times n)^2$. Because the SVT is implemented on the matrix $(E_{j,k}, E'_{j,k})$, hence both $E_{j,k}$ and $E'_{j,k}$ are processed. The super resolved frame composed from $E_{j,k}, E'_{j,k}$ contains the matrix completion results on the spline interpolations. Results of proposed MCSR are shown in Fig-2. From Table-1, we can see the proposed method has improved the quality of the video significantly, in terms of PSNR and SSIM.

### 5. CONCLUSIONS

In this paper, we presented a novel matrix completion method applied to the problem of increasing the resolution of video frames. The SVT matrix completion problem is investigated. The experimental results prove that the proposed MCSR method can significantly increase the video quality. To the best of our knowledge, no other work exist, that addresses the video super-resolution problem in a matrix completion framework.

### 6. REFERENCES


<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>Spline</th>
<th>MC Spline</th>
<th>MCSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>25.2902</td>
<td>29.1015</td>
<td>29.7816</td>
<td>30.1486</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.7011</td>
<td>0.8279</td>
<td>0.8472</td>
<td>0.8774</td>
</tr>
</tbody>
</table>

**Table 1.** PSNR and SSIM


