Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence

Low-Rank Tensor Completion with Spatio-Temporal Consistency

Hua Wang[†], Feiping Nie[‡], Heng Huang[‡*]
[†]Department of Electrical Engineering and Computer Science
Colorado School of Mines, Golden, Colorado 80401, USA
[‡]Department of Computer Science and Engineering
University of Texas at Arlington, Arlington, Texas 76019, USA
huawangcs@gmail.com, feipingnie@gmail.com, heng@uta.edu

Abstract

Video completion is a computer vision technique to recover the missing values in video sequences by filling the unknown regions with the known information. In recent research, tensor completion, a generalization of matrix completion for higher order data, emerges as a new solution to estimate the missing information in video with the assumption that the video frames are homogenous and correlated. However, each video clip often stores the heterogeneous episodes and the correlations among all video frames are not high. Thus, the regular tenor completion methods are not suitable to recover the video missing values in practical applications. To solve this problem, we propose a novel spatially-temporally consistent tensor completion method for recovering the video missing data. Instead of minimizing the average of the trace norms of all matrices unfolded along each mode of a tensor data, we introduce a new smoothness regularization along video time direction to utilize the temporal information between consecutive video frames. Meanwhile, we also minimize the trace norm of each individual video frame to employ the spatial correlations among pixels. Different to previous tensor completion approaches, our new method can keep the spatio-temporal consistency in video and do not assume the global correlation in video frames. Thus, the proposed method can be applied to the general and practical video completion applications. Our method shows promising results in all evaluations on both 3D biomedical image sequence and video benchmark data sets.

Video completion is the process of filling in missing pixels or replacing undesirable pixels in a video. The missing values in a video can be caused by many situations, *e.g.*, the natural noise in video capture equipment, the occlusion from the obstacles in environment, segmenting or removing interested objects from videos. Video completion is of great importance to many applications such as video repairing and editing, movie post-production (*e.g.*, remove unwanted objects), *etc*.

Missing information recovery in images is called inpainting, which is usually accomplished by inferring or guessing the missing information from the surrounding regions, *i.e.* the spatial information. Video completion can be considered as an extension of 2D image inpainting to 3D. Video completion uses the information from the past and the future frames to fill the pixels in the missing region, *i.e.* the spatio-temporal information, which has been getting increasing attention in recent years.

In computer vision, an important application area of artificial intelligence, there are many video completion algorithms. The most representative approaches include video inpainting, analogous to image inpainting (Bertalmio, Bertozzi, and Sapiro 2001), motion layer video completion, which splits the video sequence into different motion layers and completes each motion layer separately (Shiratori et al. 2006), space-time video completion, which is based on texture synthesis and is good but slow (Wexler, Shechtman, and Irani 2004), and video repairing, which repairs static background with motion layers and repairs moving foreground using model alignment (Jia et al. 2004).

Many video completion methods are less effective because the video is often treated as a set of independent 2D images. Although the temporal independence assumption simplifies the problem, losing temporal consistency in recovered pixels leads to the unsatisfactory performance. On the other hand, temporal information can improve the video completion results (Wexler, Shechtman, and Irani 2004; Matsushita et al. 2005), but to exploit it the computational speeds of most methods are significantly reduced. Thus, how to efficiently and effectively utilize both spatial and temporal information is a challenging problem in video completion.

In most recent work, Liu *et. al.* (Liu et al. 2013) estimated the missing data in video via tensor completion which was generalized from matrix completion methods. In these methods, the rank or rank approximation (trace norm) is used, as a powerful tool, to capture the global information. The tensor completion method (Liu et al. 2013) minimizes the trace norm of a tensor, *i.e.* the average of the trace norms of all matrices unfolded along each mode. Thus, it assumes the video frames are highly correlated in the temporal direction. If the video records homogenous episodes and all frames describe the similar information, this assumption has no problem. However, one video clip usually includes multiple different episodes and the frames from different episodes

[*]To whom all correspondence should be addressed. This work was partially supported by US NSF IIS-1117965, IIS-1302675, IIS-1344152.

could be totally different. Thus, the homogeneity assumption of the previous tensor completion method (Liu et al. 2013) restricted its application to the general and practical video missing data estimation.

In this work, we propose a novel spatially-temporally consistent tensor completion method to utilize both spatial and temporal information of a video clip. Instead of assuming the global correlation along the temporal direction as in existing works, we introduce a new smoothness regularization to utilize the content continuity within videos due to the following observation. Although one video clip can have multiple totally different episodes, most information in every two successive frames are similar. Figure 1 illustrates the content continuity in two corrupted videos. In video with image noise or obstacle occlusion, the first and last image frames have considerably large difference, but every two consecutive frames have pretty small difference. Besides using the new smoothness regularization to keep the temporal consistency, we also minimize the trace norm of each individual video frame to make use of the spatial correlations among pixels. Because our new tensor completion method uses the content continuity and does not assume the low-rank approximation among temporal direction, it can be applied to general video completion applications. We employ the Alternating Direction Method (ADM) to optimize our new objective with the global optimal solution. Both 3D biomedical image sequence and video benchmark data sets are used to evaluate the proposed method. In all experimental results, our approach outperforms the most recent methods.

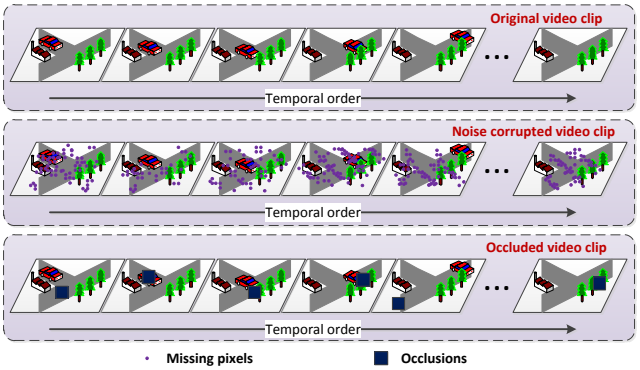


Figure 1: A video clip (top panel) naturally forms up a tensor. Missing pixel (middle panel) imputation and occlusion (bottom panel) correction can be formalized as a tensor completion problem. An important observation for video data is the content continuity, *i.e.*, the pixels at the same location in consecutive frames tends to change very small. For example, the static objects in the figure, including the road, trees, house, do not change in all the video frames. Although the dynamic object, *i.e.*, the car, has significant location change between the starting frame and the ending frame, its location changes between the consecutive frames are considerably small. As a result, continuity over the third mode of a 3-mode tensor convey very important information to recover corrupted entries, which is utilized in the proposed method to improve the tensor completion performance.

Related Work

The matrix completion problem is to find a matrix with low-rank or low norm based on the observed entries, and has been actively studied in statistical learning (Rennie and Srebro 2005; Mazumder, Hastie, and Tibshirani 2009), optimization (Candes and Recht 2008; Recht, Fazel, and Parrilo 2007; Nie, Huang, and Ding 2012; Nie et al. 2012), signal processing (Candes and Tao 2009), information retrieval (Huang, Nie, and Huang 2013; Huang et al. 2013) areas. The matrix completion problem of recovering a low-rank matrix from a subset of its entries is,

$$\begin{aligned} &\min_{\mathbf{X}\in\mathbb{R}^{n\times m}} \text{rank}(\mathbf{X}), \\ &\text{s.t.} \quad X(i,j) = D(i,j) \quad \forall\ (i,j)\in\Omega, \end{aligned} \tag{1}$$

where rank($\mathbf{X}$) denotes the rank of matrix $\mathbf{X}$, and $D(i,j) \in \mathbb{R}$ are observed entries from entries set Ω. Directly solving the problem (1) is difficult as the rank minimization problem is known as NP-hard. Recently, (M.Fazel 2002) proved the trace norm function is the convex envelope of the rank function over the unit ball of matrices, and thus the trace norm is the best convex approximation of the rank function. More recently, it has been shown in (Candes and Recht 2008; Candes and Tao 2009; Recht, Fazel, and Parrilo 2007) that, under mild conditions, the solution of problem in Eq. (1) can be found by solving the following convex problem:

$$\begin{aligned} &\min_{\mathbf{X}\in\mathbb{R}^{n\times m}} \|\mathbf{X}\|_*, \\ &\text{s.t.} \quad X(i,j) = D(i,j) \quad \forall\ (i,j)\in\Omega, \end{aligned} \tag{2}$$

where $\|\mathbf{X}\|_*$ is the trace norm of $\mathbf{X}$. Several methods (Toh and Yun 2010; Ji and Ye 2009; Liu, Sun, and Toh 2009; Ma, Goldfarb, and Chen 2009; Mazumder, Hastie, and Tibshirani 2009) have recently been published to solve this kind of trace norm minimization problem.

In practice, the given data $D(i,j)$, $(i,j)\in\Omega$, might contain noise. To handle this case, the constraint in Eq. (2) is relaxed to minimize the error of observed data entries as following:

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times m}} \|\mathbf{X}\|_* + \gamma \sum_{(i,j)\in\Omega} (X(i,j) - D(i,j))^2. \tag{3}$$

In recent work, several methods extended the matrix completion to tensor completion (Liu et al. 2013; Signoretto et al. ; Tomioka et al. 2011). They mainly minimize the average norm of all matrices on each mode. If the tensor data is homogenous, these methods can have good estimations on missing values. However, the video data is heterogenous, because one video often captures many episodes that are totally different and have few correlations along the temporal direction. Thus, the rank minimization along the temporal mode does not help the video completion. To solve this problem, we propose a new tensor completion method with keeping spatio-temporal consistency in video to estimate the missing values cross frames.

Tensor Completion with Spatio-Temporal Consistency

Given a tensor data $\mathcal{X} = \{X_1, X_2, ..., X_n\} \in \mathbb{R}^{r\times c\times n}$, and the observed values in each $X_i(1 \le i \le n)$ are $\{D_i(i,j)\,|(i,j) \in \Omega_i\}$. Suppose the third dimension of the tensor $\mathcal{X}$ is the temporal direction, and $\{X_1, X_2, ..., X_n\}$ are ordered along the temporal information. The task is to recover the unobserved values in in each $X_i(1 \le i \le n)$. Because most pixels' values are redundant, the video frames are low-rank, the low-rank matrix completion (Cai, Candès, and Shen 2010) can be applied to every frame. Thus, we can solve the following objective to complete each video frame individually:

$$\min_X \ \gamma \sum_{(k,h)\in\Omega} (X(k,h) - D(k,h))^2 + \mathrm{rank}(X). \quad (4)$$

It is known that the rank minimization problem is NP-hard, and the tightest convex relaxation to it is the following trace norm minimization problem:

$$\min_X \ \gamma \sum_{(k,h)\in\Omega} (X(k,h) - D(k,h))^2 + \|X\|_*. \quad (5)$$

Given a tensor data $\{X_1, X_2, ..., X_n\}$, we can perform the tensor completion by

$$\min_{X_i|_1^n} \ \gamma \sum_{i=1}^{n} \sum_{(k,h)\in\Omega_i} (X_i(k,h) - D_i(k,h))^2 + \sum_{i=1}^{n} \|X_i\|_*. \quad (6)$$

However, problem (6) do not take into account the temporal information of the video data, which, though, is crucial for video data. In the video, the successive data X_i and X_{i+1} should not change much between each other. In this paper, we explore the prior information of the smoothness in the tensor, and propose the smoothness regularization for tensor completion. Concretely, we add the following smoothness regularization into problem (6)

$$g(\mathcal{X}) = \sum_{i=1}^{n-1} \|X_{i+1} - X_i\|_F^2, \quad (7)$$

and to solve the following problem:

$$\begin{aligned} \min_{X_i|_1^n} \ & \gamma \sum_{i=1}^{n} \sum_{(k,h)\in\Omega_i} (X_i(k,h) - D_i(k,h))^2 \\ & + \sum_{i=1}^{n} \|X_i\|_* + \alpha \sum_{i=1}^{n-1} \|X_{i+1} - X_i\|_F^2. \end{aligned} \quad (8)$$

In Eq. (8), the parameter γ does not have explicit meanings. From the value of γ, we do not explicitly how much difference there is between the predicted value and the observed value. Thus tuning this parameter is inconvenient in practice. To alleviate this issue, we propose to solve the following problem:

$$\begin{aligned} & \min_{X_i|_1^n} \sum_{i=1}^{n} \|X_i\|_* + \alpha \sum_{i=1}^{n-1} \|X_{i+1} - X_i\|_F^2 \\ & s.t. \ |\ X_i(k,h) - D_i(k,h)| \le \varepsilon, \ \forall (k,h) \in \Omega_i, \ \forall i \end{aligned} \quad (9)$$

In Eq. (9), ε explicitly measures the difference between the predicted value and the observed value at the same location (pixel) between two consecutive frames, which makes the tuning of this parameter is controllable in practice. Moreover, compared to the overall error measured over the entire video frame as in the first term of Eq. (8), such more stringent error bound defined over every pixel is highly desirable for practical use, because the observed value are typically expensive to obtain and trustworthy, such that we should not allow too much deviation from each of these observed values.

Optimization Algorithm

Despite its nice property for practical use as analyzed above, the constraint $|\ X_i(k,h) - D_i(k,h)| \le \varepsilon$ is equivalent to the quadratic constraint $(X_i(k,h) - D_i(k,h))^2 \le \varepsilon^2$, which, however, usually makes the problem difficult or inefficient to be optimized. In this paper, we use the Alternating Direction Method (ADM) to solve this problem, in which the quadratic constraints are surprisingly very easy to be handled. We note that although ADM method has been developed for decades as an optimization framework, how to use it to solve a specific problem is not mathematically trivial. As an important theoretical contribution of our work, in this section we elegantly and rigorously derived the solution to our new objective in Eq. (9) under the ADM framework and achieved an efficient solution algorithm as summarized in Algorithm 3.

Augmented Lagrangian Method (ALM) and Alternating Direction Method (ADM)

Consider the following constrained optimization problem:

$$\min_{h(X)=0} f(X). \quad (10)$$

The Augmented Lagrangian Method (ALM) to solve the problem (10) is described in Algorithm 1. Under wild condition, the Algorithm 1 was proved to converge Q-linearly to the optimal solution (Bertsekas 1996).

Set $1 < \rho < 2$. Initialize $\mu > 0$, Λ ;
while *not converge* **do**
1. Update X by $\min_X f(X) + \frac{\mu}{2}\left\|h(X) + \frac{1}{\mu}\Lambda\right\|_F^2$;
2. Update Λ by $\Lambda = \Lambda + \mu h(X)$;
3. Update μ by $\mu = \rho\mu$;
end

Algorithm 1: Algorithm to solve the problem (10).

Consider the following optimization problem:

$$\min_{h(X,Y)=0} f(X,Y). \quad (11)$$

Alternating Direction Method (ADM) (Gabay and Mercier 1969) to solve the problem (11) is described in Algorithm 2.

Algorithm to Solve the Problem (9)

Problem (9) can be equivalently rewritten as

$$\begin{aligned} &\min_{X_i|_1^n, Z_i|_1^n} \sum_{i=1}^{n} \|Z_i\|_* + \alpha \sum_{i=1}^{n-1} \|X_{i+1} - X_i\|_F^2 \\ &s.t.\ |X_i(k,h) - D_i(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_i,\ \forall i \\ &\quad Z_i = X_i \end{aligned} \tag{12}$$

According to Algorithm 2, we need to solve the following problem:

$$\begin{aligned} &\min_{X_i|_1^n, Z_i|_1^n} \sum_{i=1}^{n} \|Z_i\|_* + \alpha \sum_{i=1}^{n-1} \|X_{i+1} - X_i\|_F^2 \\ &\quad + \tfrac{\mu}{2} \sum_{i=1}^{n} \left\| Z_i - X_i + \tfrac{1}{\mu}\Lambda_i \right\|_F^2 \\ &s.t.\ |X_i(k,h) - D_i(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_i,\ \forall i \end{aligned} \tag{13}$$

Set $1 < \rho < 2$. Initialize $\mu > 0$, Λ ;
while *not converge* **do**
1. Update X by $\min_X f(X,Y) + \frac{\mu}{2} \left\| h(X,Y) + \frac{1}{\mu}\Lambda \right\|_F^2$;
2. Update Y by $\min_Y f(X,Y) + \frac{\mu}{2} \left\| h(X,Y) + \frac{1}{\mu}\Lambda \right\|_F^2$;
3. Update Λ by $\Lambda = \Lambda + \mu h(X,Y)$;
4. Update μ by $\mu = \rho\mu$;

end

Algorithm 2: Algorithm to solve the problem (11).

When fix $X_i|_1^n$, the problem (13) is reduced to the following problem:

$$\min_{Z_i|_1^n} \frac{1}{2} \sum_{i=1}^{n} \|Z_i - M_i\|_F^2 + \frac{1}{\mu} \sum_{i=1}^{n} \|Z_i\|_* \tag{14}$$

where $M_i = X_i - \frac{1}{\mu}\Lambda_i$. Note that $Z_i|_1^n$ in problem (14) is decoupled, we only need to solve the following simplified problem for each Z_i:

$$\min_{Z_i} \frac{1}{2} \|Z_i - M_i\|_F^2 + \frac{1}{\mu} \|Z_i\|_* \tag{15}$$

The optimal solution to problem (15) can be obtained by soft thresholding method (Cai, Candès, and Shen 2010). Specifically, suppose the singular vector decomposition of M_i is $M_i = U\Sigma V^T$, and denote the i-th diagonal element of Σ_+ by σ_i, then the optimal solution Z_i^* is $Z_i^* = U\Sigma_+ V^T$, where the i-th diagonal element of Σ_+ is $\max(0, \sigma_i - \frac{1}{\mu})$.

When fix $Z_i|_1^n$, the problem (13) is reduced to the following problem:

$$\begin{aligned} &\min_{X_i|_1^n} \alpha \sum_{i=1}^{n-1} \|X_{i+1} - X_i\|_F^2 + \tfrac{\mu}{2} \sum_{i=1}^{n} \|X_i - N_i\|_F^2 \\ &s.t.\ |X_i(k,h) - D_i(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_i,\ \forall i \end{aligned} \tag{16}$$

where $N_i = Z_i + \frac{1}{\mu}\Lambda_i$. We optimize one variable of the problem (16) when fix the other n variables.

For X_1, the problem (16) becomes

$$\begin{aligned} &\min_{X_1} \alpha \|X_1 - X_2\|_F^2 + \tfrac{\mu}{2} \|X_1 - N_1\|_F^2 \\ &s.t.\ |X_1(k,h) - D_1(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_1 \end{aligned} \tag{17}$$

which can be written as

$$\begin{aligned} &\min_{X_1} \left\| X_1 - \tfrac{2}{2\alpha+\mu} \left(\alpha X_2 + \tfrac{\mu}{2} N_1 \right) \right\|_F^2 \\ &s.t.\ |X_1(k,h) - D_1(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_1 \end{aligned} \tag{18}$$

For X_n, the problem (16) becomes

$$\begin{aligned} &\min_{X_n} \alpha \|X_n - X_{n-1}\|_F^2 + \tfrac{\mu}{2} \|X_n - N_n\|_F^2 \\ &s.t.\ |X_n(k,h) - D_n(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_n \end{aligned} \tag{19}$$

which can be written as

$$\begin{aligned} &\min_{X_n} \left\| X_n - \tfrac{2}{2\alpha+\mu} \left(\alpha X_{n-1} + \tfrac{\mu}{2} N_n \right) \right\|_F^2 \\ &s.t.\ |X_n(k,h) - D_n(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_n \end{aligned} \tag{20}$$

For $X_i (1 < i < n)$, the problem (16) becomes

$$\begin{aligned} &\min_{X_i} \alpha \|X_i - X_{i+1}\|_F^2 + \alpha \|X_i - X_{i-1}\|_F^2 + \tfrac{\mu}{2} \|X_i - N_i\|_F^2 \\ &s.t.\ |X_i(k,h) - D_i(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_i \end{aligned} \tag{21}$$

which can be written as

$$\begin{aligned} &\min_{X_i} \left\| X_i - \tfrac{2}{4\alpha+\mu} \left(\alpha X_{i+1} + \alpha X_{i-1} + \tfrac{\mu}{2} N_i \right) \right\|_F^2 \\ &s.t.\ |X_i(k,h) - D_i(k,h)| \le \varepsilon,\ \forall (k,h) \in \Omega_i \end{aligned} \tag{22}$$

We can see that the three problems (18), (20) and (22) can be reduced to solving the following problem

$$\min_{|x-a| \le \varepsilon} (x - d)^2 \tag{23}$$

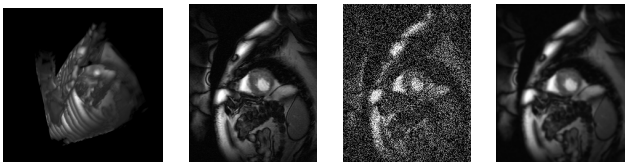
The optimal solution to this problem can be obtained by

$$\begin{cases} a - \varepsilon \le d \le a + \varepsilon & x^* = d \\ d > a + \varepsilon & x^* = a + \varepsilon \\ d < a - \varepsilon & x^* = a - \varepsilon \end{cases} \tag{24}$$

Set $1 < \rho < 2$. Initialize $\mu > 0$, $\Lambda_i (1 \le i \le n)$;
while *not converge* **do**
1. Update Z by solving problem (14) ;
2. Update X_1 by solving problem (18) ;
3. Update $X_i (1 < i < n)$ by solving problem (22) ;
4. Update X_n by solving problem (20) ;
5. Update $\Lambda_i (1 \le i \le n)$ by $\Lambda_i = \Lambda_i + \mu(Z_i - X_i)$;
6. Update μ by $\mu = \rho\mu$;

end

Algorithm 3: Algorithm to solve the problem (9).



(a) (b) (c) (d)

Figure 2: Missing value imputation for MRI data: a 3D object comprising a set of slices. (a) 3D visualization. (b) An original slice. (c) Slice with randomly removed pixels. (d) Recovered slice.

Experimental Results

In this section, we evaluate the proposed method and compare it with several related methods, where we apply our method to real world data including both 3D MRI images and video data.

In our experiments, we set the parameter α in Eq. (9) as 1 for simplicity. Given the input tensor $\mathcal{X}$, we empirically set the error bound $\varepsilon = 0.05 \sum_{ijk} |x_{ijk}| / (r \times c \times n)$.

Compared methods. We compare our method against one of the most closely related method, the Low-Rank Tensor Completion (LRTC) method (Liu et al. 2013), which has demonstrated the promising empirical results in several practical applications. For the parameter of LRTC method, we set $\alpha = 50$ to achieve best performance according to (Liu et al. 2013).

In addition, we report the results of the Accelerated Proximal Gradient singular value thresholding (APG) method (Toh and Yun 2010), which is a most recent matrix completion method and reported superior performance. Because matrix completion method work with each individual frame of a video clip or each slice of a 3D object, it is not able to take into account the temporal correlations among the video frames and the spatial correspondence among the 3D object slices. Following (Toh and Yun 2010), we set the parameter of APG method to optimal.

Following (Liu et al. 2013), we also compare our method against three heuristic algorithms for tensor completion as baseline, including Tucker algorithm (Eldén 2007), Parafac algorithm (Harshman 1970) and SVD algorithm. We refer readers to (Liu et al. 2013) for the details and parameter settings of these three algorithms.

Performance metric. We measure the accuracy of the computed solution $\mathcal{X}_{\text{sol}}$ of the compared algorithms by the relative squared error (RSE) (Liu et al. 2013; Toh and Yun 2010), which is defined by RSE $:= \frac{\|\mathcal{X}_{\text{sol}} - \mathcal{T}\|_F}{\|\mathcal{T}\|_F}$, where $\mathcal{T}$ is the original tensor without missing data and $\|\mathcal{X}\|_F = \sqrt{\sum_{i_1,i_2,i_3} |x_{i_1,i_2,i_3}|^2}$.

Improved missing value imputation for 3D objects

Missing value imputation for 3D objects has broad real world applications, such as medical image analysis. Typically high quality medical images require high radiation, *e.g.*, computed tomography (CT) images and Magnetic Resonance Imaging (MRI) images. However, too much radiation could hurt human health, therefore it is expected to control the amount of radiation without sacrificing too much image quality. As a result, when the quality of scanned CT or MRI images are not sufficiently good, imputation over the noisy or missing pixels could potentially improve the image quality for better diagnosis. Because of the spatial continuation of 3D objects along slices, a group of pixels at the same locations over different CT or MRI scan slices usually share common patterns, which makes our model of particular use to impute the missing values based upon not only the intra-slice correspondences but also the inter-slice ones. Therefore, we evaluate our method on MRI data.

Table 1: Performances of the compared methods measured by RSE for missing value imputation on 3D object (MRI images) data.

Method	RSE
Tucker	3.67×10^{-2}
Parafac	2.46×10^{-2}
SVD	4.51×10^{-2}
APG	2.84×10^{-2}
LRTC	2.01×10^{-2}
Our method	$\mathbf{7.95 \times 10^{-3}}$

The MRI data used in our experiments are a set of MRI slices, which form a tensor of size $208 \times 250 \times 170$. The 3D visualization of the data is shown in Figure 2(a), and an sample slice is shown in Figure 2(b). We randomly remove 70% pixels of each slice to emulate noise caused by low radiation as shown in Figure 2(c). Then we apply our method to complete the input tensor with missing entries, and the recovery result of the same slice is shown in Figure 2(d). As can be seen, the quality the recovered image is reasonably good, which provide a concrete evidence to support the usefulness of the proposed method in practical applications.

The quantitative recovery performances of the compared methods measured by RSE are listed in Table 1, which show that our method clearly outperforms the other compared methods. Moreover, although the APG method is better than two heuristic methods, it is worse than the two tensor completion method, especially when compared with the proposed method. This is consistent with our previous theoretical analysis in that our method is able to exploit the continuation over the slices of a 3D object, which thereby can capture the additional spatial information. In contrast, the APG method is a matrix completion method but not directly deals with tensor, therefore the useful information conveyed by tensor is not employed.

Improved missing value imputation and occlusion removal for video data

Video data can be naturally described by tensors, in which the third mode is often used to capture the crucial information of the temporal order of a video clip. As a result, missing value imputation and occlusion removal for video clips

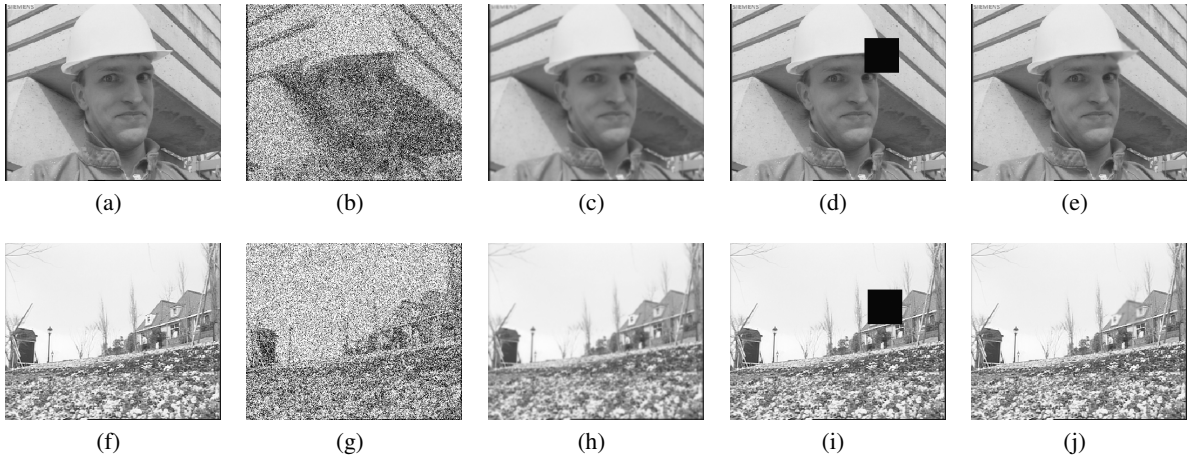


Figure 3: Missing value imputation and occlusion removal. (a—e) Foreman video data, (f—j) Flower video data. (a, f) A sample frame. (b, g) Corrupted by missing entries. (c, h) Recovered from missing entries. (d, i) Corrupted by occlusion. (e, j) Recovered from occlusion.

Table 2: Performances of the compared methods measured by RSE for missing value imputation on video data.

	Foreman	Flower
Tucker	4.52×10^{-2}	7.95×10^{-2}
Parafac	3.81×10^{-2}	7.84×10^{-2}
SVD	4.39×10^{-2}	8.02×10^{-2}
APG	3.70×10^{-2}	7.07×10^{-2}
LRTC	2.89×10^{-2}	5.11×10^{-2}
Our method	$\mathbf{1.57 \times 10^{-2}}$	$\mathbf{4.87 \times 10^{-2}}$

Table 3: Performances of the compared methods measured by RSE for occlusion removal on video data.

	Foreman	Flower
Tucker	3.88×10^{-2}	6.14×10^{-2}
Parafac	3.29×10^{-2}	6.21×10^{-2}
SVD	3.74×10^{-2}	6.53×10^{-2}
APG	3.17×10^{-2}	6.10×10^{-2}
LRTC	2.51×10^{-2}	4.96×10^{-2}
Our method	$\mathbf{1.17 \times 10^{-2}}$	$\mathbf{2.81 \times 10^{-2}}$

turn out to be tensor completion problems. Therefore, in this subsection we evaluate the proposed method in both missing value imputation and occlusion removal tasks on two benchmark video data clips: Foreman data and Flower data. These two video clips are in CIF format. The former clip has 300 frames, therefore it forms a tensor of size $352 \times 288 \times 300$; the latter clip has 250 frames, therefore it forms a tensor of size $352 \times 288 \times 250$. A sample frame of each video clip is shown in Figure 3(a) and Figure 3(f) respectively. For missing value imputation task, same as in previous experiments, we randomly remove 70% pixels of each frame of a video clip as shown in Figure 3(b) and Figure 3(g), and apply the compared methods to recover them. For the latter, for each frame, we first randomly pick a location in the frame and then remove a block of pixels of size 64×64 as shown in Figure 3(d) and Figure 3(i), and then apply the compared methods to remove the occlusions.

The results of the missing value imputation task are shown in Figure 3(c) and Figure 3(h), from which we can see that the definitions of the recovered sample frames largely remain same as the original one. The results for the occlusion removal task are shown in Figure 3(e) and Figure 3(j), in which the occluded patches are generally unnoticeable in the recovered frames. These results firmly confirm the effectiveness of the proposed method to recovery incomplete frames on video data.

Finally, we quantitatively compare our method against other related tensor (matrix) completion methods. The results measured by RSE are listed in Table 2 and Table 3. Again, our method reports the best performances. More results are shown in the accompanied videos, which are supplied as the supplementary materials of this paper.

Conclusions

In this paper, we proposed a novel spatio-temporal consistency tensor completion method to restore the missing values in video. Different to existing tensor completion methods, we did not minimize the average of the trace norms of all matrices unfolded along each mode. Instead, we introduced a new smoothness regularization to utilize the video content continuity in temporal direction. We also minimized the trace norm of each individual video frame to employ the spatial correlations among pixels. Because our new method kept the spatio-temporal consistency in video and didn't assume the global correlation among video frames, it is suitable to the general video completion applications. Our method showed promising results in all experiments on 3D MRI image sequence and video benchmark data sets.

References

Bertalmio, M.; Bertozzi, A.; and Sapiro, G. 2001. Navier-stokes, fluid dynamics, and image and video inpainting. In *Computer Vision and Pattern Recognition*, 355–362. IEEE.

Bertsekas, D. P. 1996. *Constrained optimization and lagrange multiplier methods*. Athena Scientific.

Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4):1956–1982.

Candes, E., and Recht, B. 2008. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*.

Candes, E. J., and Tao, T. 2009. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory* 56(5):2053–2080.

Eldén, L. 2007. *Matrix methods in data mining and pattern recognition*. SIAM.

Gabay, D., and Mercier, B. 1969. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications* 2(1):17–40.

Harshman, R. 1970. Foundations of the parafac procedure: Models and conditions for an "explanatory"multimodal factor analysis. *UCLA Working Papers in Phonetics*.

Huang, J.; Nie, F.; Huang, H.; Lei, Y.; and Ding, C. 2013. Social trust prediction using rank-k matrix recovery. *23rd International Joint Conference on Artificial Intelligence (IJCAI)* 2647–2653.

Huang, J.; Nie, F.; and Huang, H. 2013. Robust discrete matrix completion. *Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13)* 424–430.

Ji, S., and Ye, Y. 2009. An accelerated gradient method for trace norm minimization. *ICML*.

Jia, J.; Wu, T.-P.; Tai, Y.-W.; and Tang, C.-K. 2004. Video repairing: Inference of foreground and background under severe occlusion. In *Computer Vision and Pattern Recognition*, 364–371. IEEE.

Liu, J.; Musialski, P.; Wonka, P.; and Ye, J. 2013. Tensor completion for estimating missing values in visual data. *IEEE Transactions onPattern Analysis and Machine Intelligence (TPAMI)* 35(1):208–220.

Liu, Y.-J.; Sun, D.; and Toh, K.-C. 2009. An implementable proximal point algorithmic framework for nuclear norm minimization. *Optimization Online*.

Ma, S.; Goldfarb, D.; and Chen, L. 2009. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*.

Matsushita, Y.; Ofek, E.; Tang, X.; and Shum, H.-Y. 2005. Full-frame video stabilization. In *Computer Vision and Pattern Recognition*, 50–57. IEEE.

Mazumder, R.; Hastie, T.; and Tibshirani, R. 2009. Spectral regularization algorithms for learning large incomplete matrices. *submitted to JMLR*.

M.Fazel. 2002. Matrix rank minimization with applications. *Ph.D. dissertation, Stanford University*.

Nie, F.; Wang, H.; Cai, X.; Huang, H.; and Ding, C. 2012. Robust matrix completion via joint schatten p-norm and lp-norm minimization. *ICDM* 566–574.

Nie, F.; Huang, H.; and Ding, C. 2012. Schatten-p Norm Minimization for Low-Rank Matrix Recovery. *Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI 2012)* 157.

Recht, B.; Fazel, M.; and Parrilo, P. A. 2007. Guaranteed minimumrank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review* 52(3):471–501.

Rennie, J., and Srebro, N. 2005. Fast maximum margin matrix factorization for collaborative prediction. *ICML*.

Shiratori, T.; Matsushita, Y.; Kang, S. B.; and Tang, X. 2006. Video completion by motion field transfer. In *Computer Vision and Pattern Recognition*, 411–418. IEEE.

Signoretto, M.; Van de Plas, R.; De Moor, B.; and Suykens, J. Tensor versus matrix completion: A comparison with application to spectral data. *IEEE Signal Processing Letters* 18(7).

Toh, K., and Yun, S. 2010. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization* 6:615–640.

Tomioka, R.; Hayashi, K.; ; and Kashima, H. 2011. On the extension of trace norm to tensors. In *NIPS Workshop on Tensors, Kernels and Machine Learning*.

Wexler, Y.; Shechtman, E.; and Irani, M. 2004. Space-time video completion. In *Computer Vision and Pattern Recognition*, 120–127. IEEE.