Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Video Recovery via Learning Variation and Consistency of Images

Zhouyuan Huo,[1] Shangqian Gao,[2] Weidong Cai,[3] Heng Huang[1]*
[1]Department of Computer Science and Engineering, University of Texas at Arlington, USA
[2]College of Engineering, Northeastern University, USA
[3]School of Information Technologies, University of Sydney, NSW 2006, Australia
zhouyuan.huo@mavs.uta.edu, gao.sh@husky.neu.edu, tom.cai@sydney.edu.au, heng@uta.edu

Abstract

Matrix completion algorithms have been popularly used to recover images with missing entries, and they are proved to be very effective. Recent works utilized tensor completion models in video recovery assuming that all video frames are homogeneous and correlated. However, real videos are made up of different episodes or scenes, *i.e.* heterogeneous. Therefore, a video recovery model which utilizes both video spatiotemporal consistency and variation is necessary. To solve this problem, we propose a new video recovery method Sectional Trace Norm with Variation and Consistency Constraints (STN-VCC). In our model, capped ℓ_1-norm regularization is utilized to learn the spatial-temporal consistency and variation between consecutive frames in video clips. Meanwhile, we introduce a new low-rank model to capture the low-rank structure in video frames with a better approximation of rank minimization than traditional trace norm. An efficient optimization algorithm is proposed, and we also provide a proof of convergence in the paper. We evaluate the proposed method via several video recovery tasks and experiment results show that our new method consistently outperforms other related approaches.

Introduction

Image recovery task can be treated as a matrix completion problem, and it is proved to be really effective (Buades, Coll, and Morel 2005; Ji et al. 2010; Liu et al. 2009; Wang, Nie, and Huang 2014). In this task, pixels in an image are missing because of noise or occlusion, and we need to make use of observed information to reconstruct the original image without deviating too much from it. Rank minimization regularization is proposed to solve matrix completion problems, and has already achieved great success in other fields (Bennett and Lanning 2007). However, it leads to a non-convex and NP-hard problem. To solve this problem, trace norm is introduced to approximate rank minimization, and since then many algorithms have been invented to solve matrix completion problems using trace norm regularization (Shamir and Shalev-Shwartz 2014; Cai, Candès, and Shen 2010). However, trace norm is not the best approximation for rank minimization regularization. If a non-zero singular value of a matrix varies, the trace norm value will change simultaneously, but the rank of the matrix keeps constant. So, there is a big gap between trace norm and rank minimization, and a better model to approximate rank minimization regularization is desired for learning a low-rank structure.

In a recent work (Liu et al. 2009), Liu applied low-rank model to solve image recovery problem. He assumed that video is in low-rank structure, however, videos are not homogeneous, and they may contain sequences of images from different episodes. If we simply project all the video into a low-rank subspace, it is very likely to lose important information in the original video. More recently, Wang *et al.* introduced a spatiotemporal consistency low-rank tensor completion method (Wang, Nie, and Huang 2014). They proposed to utilize the content continuity within videos by imposing a new ℓ_2-norm smoothness regularization. However, this ℓ_2-norm smoothness regularization tends to force two consecutive frames to be similar, and it can not capture spatiotemporal variation properly.

In this paper, we propose a novel video recovery model Sectional Trace Norm with Variation and Consistency Constraints (STN-VCC). Our contributions are summarized as following: firstly, we utilize capped ℓ_1-norm smoothness function to capture the spatiotemporal variation and consistency between consecutive frames in video clips; secondly, we introduce a new sectional trace norm to capture the low-rank structure, which is a better approximation of rank minimization than the traditional trace norm. Thirdly, instead of using Alternating Direction Method of Multipliers (ADMM) which was used by previous video recovery models, we use proximal gradient descent method to optimize and prove that our method is guaranteed to converge.

Video Recovery with Capped Norm Constrained Variation and Consistency

Capture Spatiotemporal Consistency and Variation in Video Clip via Capped ℓ_1-norm

In (Wang, Nie, and Huang 2014), authors proposed a ℓ_2-norm regularization $||E||_2$ to maintain spatiotemporal consistency in video clips, it is reasonable to suppose that two consecutive frames in video clips are alike. However, not

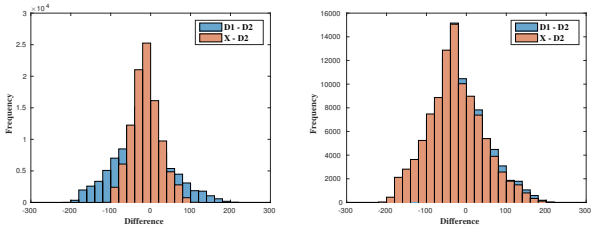
*To whom all correspondence should be addressed. This work was partially supported by the following grants: NSF-IIS 1302675, NSF-IIS 1344152, NSF-DBI 1356628, NSF-IIS 1619308, NSF-IIS 1633753, NIH R01 AG049371.

all video sequences share the same content, spatiotemporal variation is also very frequent in video clips. ℓ_1-norm is a common alternative to ℓ_2-norm, and it is known for its robustness to outliers. It is intuitive to utilize ℓ_1-norm regularization to capture both spatiotemporal consistency and variation. However, one video clip is usually made up of many different episodes, and it is natural that two consecutive frames are from different scenes and backgrounds. In this case, most of the pixels in two consecutive frames are different, and even ℓ_1-norm regularization fails. In this paper, we propose to use capped ℓ_1-norm regularization $\sum_{e\in E} \min(|e|, \varepsilon)$ to capture the spatiotemporal consistency and variation in video clips. Capped ℓ_1-norm function is able to ignore the errors larger than ε, penalize the errors smaller than ε, and it is also used to handle outliers (Zhang 2009; Gong et al. 2013; Huo, Nie, and Huang 2016; Gao et al. 2015; Jiang, Nie, and Huang 2015). In this paper, it properly solves the problem of spatiotemporal consistency and variation in video clips.

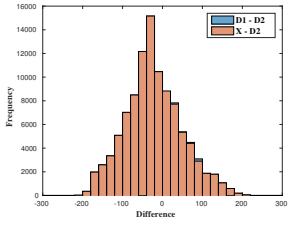
We perform an experiment to illustrate the advantage of capped ℓ_1-norm over ℓ_2-norm and ℓ_1-norm. The objective function of this task is

$$\min_X ||X - D_1||_F^2 + \lambda \; reg(X - D_2) \tag{1}$$

where $D_1, D_2 \in \mathbb{R}^{288\times 352}$ are two distinct consecutive image matrices, and $reg(E) = ||E||_F^2$, $reg(E) = ||E||_1$ or $reg(E) = \sum_{e\in E} min(|e|, \varepsilon)$ for three models. In this experiment, we set $\lambda = 0.5$ and $\varepsilon = 100$. We use stochastic gradient method to optimize and learning rate is 0.1.



(a) ℓ_2-norm (b) ℓ_1-norm



(c) capped ℓ_1-norm

Figure 1: We compare ℓ_2-norm, ℓ_1-norm and capped ℓ_1-norm regularization.

In Figure 1, as we know that ℓ_2-norm regularization penalizes all the differences, after the optimization, all differences tend to be zero, and this is not what we want if there are spatiotemporal variations in video clips. The result of ℓ_1-norm and capped ℓ_1-norm are much better than ℓ_2-norm. It is also clear that capped ℓ_1-norm outperforms ℓ_1-norm. In large difference ranges, $|Difference| \geq 100$ in the experiment, capped ℓ_1-norm does not penalize these differences for they represent different objects or episodes.

Sectional Trace Norm Based Low-Rank Model

Recently, the low-rank model and its approximation have been successfully applied to recover the images and videos with missing pixel. Most of these works used trace norm to recover the low-rank structure of the image matrix. For a low-rank matrix $X \in \mathbb{R}^{n\times m}$, $rank(X) = r$, so its k-smallest singular values should be zero, where $k = d - r$ and $d = min(n, m)$. Note that trace norm $||X||_*$ denotes the sum of singular values. **If the non-zero singular values of matrix X change, $||X||_*$ will change as well, but the rank of X keeps constant.** Thus, there is a big gap between trace norm $||X||_*$ and $rank(X)$.

In this paper, we propose a new sectional trace norm to uncover the low-rank structure of a matrix, $||X||_{str} = \sum_{i=1}^{k} \sigma_i^2(X)$. In the sectional trace norm regularization, we minimize the sum of k-smallest singular value squares of X and ignore the other larger singular values. When the non-zero singular values increase largely, they are excluded by our sectional trace norm such that the norm value keeps constant.

Alternatively, in (Hu et al. 2013), truncated trace norm was proposed to minimize the sum of k-smallest singular values, and it can also avoid the effect of large singular values, and is better than the traditional trace norm. However, minimizing the sum of k-smallest singular values is a ℓ_1 minimization problem, which leads to sparse solution, namely some k-smallest singular values will be zero, but some of them may get large values. Our sectional trace norm can solve this issue. When we minimize the sectional trace norm, the sum of square of k-smallest singular values is minimized so that all of them will be shrunk near to zero.

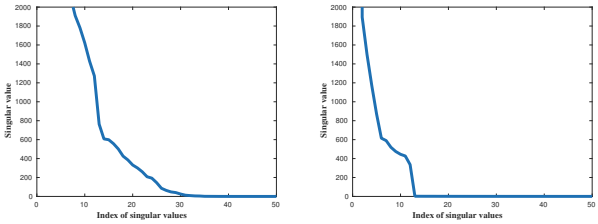
We perform a low-rank matrix approximation experiment to illustrate the advantage of sectional trace norm over truncated trace norm. The objective function of this task is:

$$\min_X ||X - D||_F^2 + \lambda \; reg(X)\,, \tag{2}$$

where $D \in \mathbb{R}^{288\times 352}$ is an image matrix, and $reg(X) = \sum_{i=1}^{k} \sigma_i(X)$ or $reg(X) = \sum_{i=1}^{k} \sigma_i^2(X)$ for two models. We compute a low-rank matrix X and $rank(X) = 12$, so $k = 276$ and $\lambda = 1000$.

In Figure 2a, although there is a big gap between 12th and 13th singular values, some singular values after 12th are still large, and this is not what we expect. However, in Figure 2b, it is obvious that the singular values after 12th singular value are nearly zero. Thus our sectional trace norm is a better approximation of rank function than existing models, and is more suitable to be used for low-rank matrix completion tasks.

To sum up, our new video recovery model which uses sectional trace norm with consistency and variation constraints



(a) (b)

Figure 2: We compare two different rank minimization approximations. (a) minimizing the sum of k-smallest singular values. (b) minimizing the sum of k-smallest singular value squares.

can be represented as:

$$\begin{aligned} \min_{\mathcal{X}_l|_1^s} \quad & \sum_{l=1}^{s-1}\sum_{p,q}\min(|\mathcal{X}_{pq,l+1}-\mathcal{X}_{pq,l}|,\varepsilon) \\ & +\lambda\sum_{l=1}^{s}(\sum_{i=1}^{k}\sigma_i^2(\mathcal{X}_l)) \\ \text{s.t.} \quad & |\mathcal{X}_{pq,l}-\mathcal{D}_{pq,l}|\le\gamma,\ \forall(p,q)\in\Omega_l,\ \forall l \end{aligned} \tag{3}$$

where $\mathcal{X},\mathcal{D}\in\mathbb{R}^{n\times m\times s}$, and there are s frames in a video, and each frame is a matrix of $n\times m$. $\mathcal{X}_{pq,l}$ denotes the predicted value in frame l at position (p,q), $\mathcal{D}$ denotes the original video tensor, Ω_l denotes the observed pixel in frame i. ε is the threshold value and γ is the error upper bound.

Optimization Algorithm

In previous papers, they use ADMM to optimize the objective function. ADMM makes it easy to handle constraints in the objective function, however it is hard to prove convergence. In this paper, we use proximal gradient method to optimize our model, and in the next section, we will prove the convergence of our method.

Define $\mathcal{X}_l = U\Sigma V^T$, where Σ is diagonal singular matrix in ascending order. Let $\mathcal{F}_l = U_{1:k}$, where $U_{1:k}$ means k smallest singular vectors:

$$\mathrm{Tr}\left(\mathcal{F}_l^T\mathcal{X}_l\mathcal{X}_l^T\mathcal{F}_l\right)=\sum_{i=1}^{k}\sigma_i^2(\mathcal{X}_l) \tag{4}$$

First, we find out that we only need calculate $\mathcal{F}_l\mathcal{F}_l^T$, instead of computing matrix $\mathcal{F}_l$, in the optimization process. We introduce an efficient way to compute $\mathcal{F}_l\mathcal{F}_l^T$ as a whole term. Suppose singular value decomposition of $\mathcal{X}_l\mathcal{X}_l^T = U\Sigma U^T$, where U is the eigenvector matrix and Σ is the diagonal matrix in ascending order. Denote $U=[U_1,U_2]$, where $U_1\in\mathbb{R}^{n\times k}$, and $U_2\in\mathbb{R}^{n\times(n-k)}$, then $\mathcal{F}_l = U_1$. It is easy to see that $UU^T = U_1U_1^T + U_2U_2^T = I$. We have:

$$\mathcal{F}_l\mathcal{F}_l^T = I - U_2U_2^T \tag{5}$$

where $U_2\in\mathbb{R}^{n\times(n-k)}$ is $n-k$ largest singular vectors. We suppose that $\mathcal{X}_l\mathcal{X}_l^T$ is a low-rank matrix, so $n-k$ is a small value. Truncated SVD method can be applied in this procedure, and it is much more efficient to compute $\mathcal{F}_l\mathcal{F}_l^T$ in this way. Each $\mathcal{F}_l$ is updated independently according to function (5) for each $\mathcal{X}_l$.

To handle capped ℓ_1-norm term in our objective function (3), we define tensor $\mathcal{S}$ as:

$$\mathcal{S}_{pq,l}=\begin{cases}\frac{1}{2|\mathcal{X}_{pq,l+1}-\mathcal{X}_{pq,l}|} & if \quad |\mathcal{X}_{pq,l+1}-\mathcal{X}_{pq,l}|<\varepsilon\\ 0 & otherwise\end{cases} \tag{6}$$

Then, as per function (5) and (6), our objective function is transformed to:

$$\begin{aligned} \min_{\mathcal{X}_l|_1^s} \quad & \frac{1}{2}\sum_{l=1}^{s-1}\sum_{p,q}\mathcal{S}_{pq,l}(\mathcal{X}_{pq,l+1}-\mathcal{X}_{pq,l})^2 \\ & +\frac{1}{2}\lambda\sum_{l=1}^{s}\mathrm{Tr}\left(\mathcal{F}_l^T\mathcal{X}_l\mathcal{X}_l^T\mathcal{F}_l\right) \\ \text{s.t.} \quad & |\mathcal{X}_{pq,l}-\mathcal{D}_{pq,l}|\le\gamma,\ \forall(p,q)\in\Omega_l,\ \forall l \end{aligned} \tag{7}$$

Each $\mathcal{X}_l$ is also solved independently and alternatively, and for any $1<l<s$, this subproblem is,

$$\begin{aligned} \min_{\mathcal{X}_l} \quad & \frac{1}{2}\sum_{p,q}\mathcal{S}_{pq,l}(\mathcal{X}_{pq,l+1}-\mathcal{X}_{pq,l})^2 \\ & +\frac{1}{2}\sum_{p,q}\mathcal{S}_{pq,l-1}(\mathcal{X}_{pq,l}-\mathcal{X}_{pq,l-1})^2+\frac{1}{2}\lambda\,\mathrm{Tr}\left(\mathcal{F}_l^T\mathcal{X}_l\mathcal{X}_l^T\mathcal{F}_l\right) \\ \text{s.t.} \quad & |\mathcal{X}_{pq,l}-\mathcal{D}_{pq,l}|\le\gamma,\quad \forall(p,q)\in\Omega_l \end{aligned} \tag{8}$$

We use proximal gradient method to solve the subproblem above. Its gradient with respective to $\mathcal{X}_l$ is

$$\Delta = \mathcal{S}_l\circ(\mathcal{X}_l-\mathcal{X}_{l+1})+\mathcal{S}_{l-1}\circ(\mathcal{X}_l-\mathcal{X}_{l-1})+\lambda\mathcal{F}_l\mathcal{F}_l^T\mathcal{X}_l \tag{9}$$

$$\mathcal{X}_l = \mathcal{X}_l - \alpha\Delta \tag{10}$$

where α is the step size.

According to the constraint $|\mathcal{X}_{pq,l}-\mathcal{D}_{pq,l}|\le\gamma$. We need to project updated $\mathcal{X}_l$ to this constraint domain. The optimal solution to this projection can be obtained by

$$P(\mathcal{X}_{pq,l})=\begin{cases}\mathcal{X}_{pq,l} & \mathcal{D}_{pq,l}-\gamma\le\mathcal{X}_{pq,l}\le\mathcal{D}_{pq,l}+\gamma\\ \mathcal{D}_{pq,l}+\gamma & \mathcal{X}_{pq,l}>\mathcal{D}_{pq,l}+\gamma\\ \mathcal{D}_{pq,l}-\gamma & \mathcal{X}_{pq,l}<\mathcal{D}_{pq,l}-\gamma\end{cases} \tag{11}$$

As for $l=1$ or $l=s$, we just need to ignore one of the two loss functions in (8), and follow the proximal gradient method steps to update $\mathcal{X}_l$.

Our optimization method is summarized in Algorithm. 1.

Algorithm 1 Algorithm to solve problem (3).

Input: $\mathcal{D},\Omega\in\mathbb{R}^{n\times m\times s}$, α
Output: $\mathcal{X}\in\mathbb{R}^{n\times m\times s}$
while *not converge* **do**
  Update $\mathcal{F}$ via (5).
  for $l=1$ to s **do**
    1. Update $\mathcal{S}_l$ via (6) if $l<s$.
    2. Update $\tilde{\mathcal{X}}_l$ via (9), (10).
    3. Project to $\mathcal{X}_l$ via (11)
  end for
end while

Convergence Analysis

Using the algorithm above, we can solve our original non-smooth and non-convex objective function (3). In this section, we prove the convergence of our optimization algorithm, and that a local solution can be obtained in the end.

Theorem 1 *Through Algorithm. 1, the objective function (3) will converge, or the values of objective function (3) are non-increasing monotonically.*

In order to prove Theorem 1, at first, we need the following Lemmas.

Lemma 2 *According to (Theobald 1975), any two hermitian matrices $A, B \in R^{n\times n}$ satisfy the inequality ($\sigma_i(A)$, $\sigma_i(B)$ are singular values sorted in the same order)*

$$\sum_{i=1}^{n}\sigma_i(A)\,\sigma_{n-i+1}(B) \le \mathrm{Tr}\left(A^TB\right) \le \sum_{i=1}^{n}\sigma_i(A)\,\sigma_i(B) \tag{12}$$

Lemma 3 *Let $X = U\Sigma V^T$, σ_i are singular values of X in ascending order, $\sum_{i=1}^{k}\sigma_i^2(X)$ is the sum of k smallest singular value squares of X. Similarly $\hat{X} = \hat{U}\hat{\Sigma}\hat{V}^T$, $\hat{\sigma}_i$ are singular values of $\hat{X}$ in ascending order, $\sum_{i=1}^{k}\hat{\sigma}_i^2(\hat{X})$ is the sum of k smallest singular value squares of $\hat{X}$. So it is true that:*

$$\begin{aligned}&\sum_{i=1}^{k}\hat{\sigma}_i^2(\hat{X}) - \mathrm{Tr}\left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i^T\hat{X}\hat{X}^T\right)\\ \le &\sum_{i=1}^{k}\sigma_i^2(X) - \mathrm{Tr}\left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i^TXX^T\right)\end{aligned} \tag{13}$$

Proof: Because $X = U\Sigma V^T$, it is obvious that

$$\mathrm{Tr}\left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i^TXX^T\right) = \sum_{i=1}^{k}\sigma_i^2(X) \tag{14}$$

Via Lemma 2, we know

$$\begin{aligned}\mathrm{Tr}\left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i^T\hat{X}\hat{X}^T\right) &= \mathrm{Tr}\left(U_{1:k}U_{1:k}^T\hat{U}\hat{\Sigma}^2\hat{U}^T\right)\\ &\ge \sum_{i=1}^{k}\hat{\sigma}_i^2(\hat{X})\end{aligned} \tag{15}$$

Above all, we have,

$$\begin{aligned}0 &= \sum_{i=1}^{k}\sigma_i^2(X) - \mathrm{Tr}\left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i^TXX^T\right)\\ &\ge \sum_{i=1}^{k}\hat{\sigma}_i^2(\hat{X}) - \mathrm{Tr}\left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i^T\hat{X}\hat{X}^T\right)\end{aligned} \tag{16}$$

Lemma 4 *If*

$$d = \begin{cases}\frac{1}{2|e|} & \textit{if} \quad |e| < \varepsilon\\ 0 & \textit{otherwise}\end{cases} \tag{17}$$

Then,

$$\min\{|\hat{e}|,\varepsilon\} - d\hat{e}^2 \le \min\{|e|,\varepsilon\} - de^2 \tag{18}$$

Proof: As we all know

$$\begin{aligned}|e| - 2|\hat{e}| + |e|^{-1}|\hat{e}|^2 &= |e|^{-1}\left(|e|^2 - 2|e||\hat{e}| + |\hat{e}|^2\right)\\ &= |e|^{-1}\left(|e| - |\hat{e}|\right)^2 \ge 0\end{aligned} \tag{19}$$

so,

$$|\hat{e}| - \frac{|\hat{e}|^2}{2|e|} \le \frac{|e|}{2} \tag{20}$$

If $|e| < \varepsilon$, $d = \frac{1}{2|e|}$, it is clear that $\min\{|\hat{e}|,\varepsilon\} \le |\hat{e}|$, and

$$\begin{aligned}\min\{|\hat{e}|,\varepsilon\} - \frac{\hat{e}^2}{2|e|} &\le |\hat{e}| - \frac{\hat{e}^2}{2|e|}\\ &\le \frac{|e|}{2} = \min\{|e|,\varepsilon\} - \frac{e^2}{2|e|}\end{aligned} \tag{21}$$

On the other hand, if $|e| \ge \varepsilon$, $d = 0$ the following inequality always holds,

$$\min\{|\hat{e}|,\varepsilon\} \le \min\{|e|,\varepsilon\} \tag{22}$$

Lemma 5 *Function value of (8) is non-increasing through proximal gradient descent method.*

Proof: Firstly, function in (8) is convex, and its gradient Δ is Lipschitz continuous. Secondly, projection in (8) is convex and closed. According to (Beck and Teboulle 2010), if step size is small enough, proximal gradient descent method is able to improve the objective function value at each step.

Now, we are able to prove Theorem 1 by applying these Lemmas above.

Proof: From Lemma 3 and definition of $\mathcal{F}_l$, the inequality holds:

$$\begin{aligned}&\sum_{l=1}^{s}\sum_{i=1}^{k}\hat{\sigma}_i^2(\hat{\mathcal{X}}_l) - \sum_{l=1}^{s}\mathrm{Tr}\left(\mathcal{F}_l^T\hat{\mathcal{X}}_l\hat{\mathcal{X}}_l^T\mathcal{F}_l\right)\\ \le &\sum_{l=1}^{s}\sum_{i=1}^{k}\sigma_i^2(\mathcal{X}_l) - \sum_{l=1}^{s}\mathrm{Tr}\left(\mathcal{F}_l^T\mathcal{X}_l\mathcal{X}_l^T\mathcal{F}_l\right)\end{aligned} \tag{23}$$

Applying Lemma 4 to each element $\mathcal{X}_{pq,l}$ and $\hat{\mathcal{X}}_{pq,l}$, we have:

$$\begin{aligned}&\sum_{l=1}^{s-1}\sum_{p,q}\min(|\hat{\mathcal{X}}_{pq,l+1} - \hat{\mathcal{X}}_{pq,l}|,\varepsilon) - \sum_{l=1}^{s-1}\sum_{p,q}\mathcal{S}_{pq,l}(\hat{\mathcal{X}}_{pq,l+1} - \hat{\mathcal{X}}_{pq,l})^2\\ \le &\sum_{l=1}^{s-1}\sum_{p,q}\min(|\mathcal{X}_{pq,l+1} - \mathcal{X}_{pq,l}|,\varepsilon) - \sum_{l=1}^{s-1}\sum_{p,q}\mathcal{S}_{pq,l}(\mathcal{X}_{pq,l+1} - \mathcal{X}_{pq,l})^2\end{aligned} \tag{24}$$

According to Lemma 5, after we minimize the function (7) by using proximal gradient descent, it is guaranteed:

$$\begin{aligned}&\sum_{l=1}^{s-1}\sum_{p,q}\mathcal{S}_{pq,l}(\hat{\mathcal{X}}_{pq,l+1} - \hat{\mathcal{X}}_{pq,l})^2 + \lambda\sum_{l=1}^{s}\mathrm{Tr}\left(\mathcal{F}_l^T\hat{\mathcal{X}}_l\hat{\mathcal{X}}_l^T\mathcal{F}_l\right)\\ \le &\sum_{l=1}^{s-1}\sum_{p,q}\mathcal{S}_{pq,l}(\mathcal{X}_{pq,l+1} - \mathcal{X}_{pq,l})^2 + \lambda\sum_{l=1}^{s}\mathrm{Tr}\left(\mathcal{F}_l^T\mathcal{X}_l\mathcal{X}_l^T\mathcal{F}_l\right)\end{aligned} \tag{25}$$

When we combine inequalities (23), (24) and (25), we can finally obtain:

$$\sum_{l=1}^{s-1}\sum_{p,q}\min(|\hat{\mathcal{X}}_{pq,l+1}-\hat{\mathcal{X}}_{pq,l}|,\varepsilon)+\lambda\sum_{l=1}^{s}\sum_{i=1}^{k}\hat{\sigma}_i^2(\hat{\mathcal{X}}_l)$$
$$\leq\sum_{l=1}^{s-1}\sum_{p,q}\min(|\mathcal{X}_{pq,l+1}-\mathcal{X}_{pq,l}|,\varepsilon)+\lambda\sum_{l=1}^{s}\sum_{i=1}^{k}\sigma_i^2(\mathcal{X}_l) \tag{26}$$

So far, it is clear that the value of our proposed objective function will not increase by using our optimization algorithm, so we prove the Theorem 1 that our optimization algorithm is non-increasing monotonically. We also know that the objective function (3) is larger than zero at least, so it is lower bounded. We can conclude that our optimization algorithm converges, and a local solution is to be obtained in the end.

Experiments

In this section, we evaluate our proposed video recovery model and compare it with other five related methods: Tucker algorithm (Tucker) (Eldén 2007), low-rank tensor completion (LRTC) (Liu et al. 2009), low-rank tensor completion with considering consistency (ℓ_2Tensor) (Wang, Nie, and Huang 2014), low-rank tensor completion with considering consistency and variation (capTensor), sectional trace norm with considering consistency ($k$msv-ℓ_2), and our sectional trace norm with considering both consistency and variation.

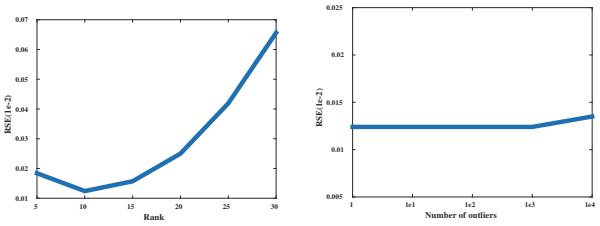
In the experiments, error bound $\gamma = 0.05\sum|\mathcal{X}|/(n\times m\times s)$. Relative square error (RSE) in (Wang, Nie, and Huang 2014) is used as performance metric criterion for comparison. R, G and B layer of each colorful video are represented as a 3D tensor and put into the model respectively and combined as final outputs.

Data Sets

- **UCF11 Dataset**: It contains 11 action categories: basketball shooting, biking, diving, golf swinging and so on. This dataset is very challenging due to large variations in camera motion, object appearance, pose and so on (Liu, Luo, and Shah 2009) .
- **YUV Video Sequences Dataset**: It includes video sequences of commonly used video test sequences in the 4:2:0 YUV format, e.g. Elephant Dream video, Highway, News and Stephan[1].
- **Hollywood Human Actions Dataset**: It contains video clips, i.e. short sequences from 32 movies: American Beauty, As Good As It Gets, Being John Malkovich, Big Fish and so on(Laptev et al. 2008).

Selection of k and ε

In this section, we evaluate the influence of parameters k-smallest singular value and ε outliers bound on recovery performance. According to the optimization analysis, instead of selecting parameter k, we take $r = n-k$, the rank of matrix, as the input. In the experiment, the value of ε is generated automatically by setting the number of outliers N_ε. We perform experiment on the Sales video clip data. In this experiment, we tune the parameters through grid search strategy, $r = \{5, 10, 15, 20, 25, 30\}$ and $N_\varepsilon = \{1, 10, 10^2, 10^3, 10^4\}$, and plot the results in Figure 3.



(a) Rank vs RSE (b) Outliers vs RSE

Figure 3: Influence of rank $r = n - k$ and the number of outliers N_ε on our method's performance.

As we can see in Figure 3a, the choice of rank value r is nontrivial. In this experiment, if $r < 10$ or $r > 10$, RSE value rises. So, in order to derive the best performance, it is of great importance to select an optimal rank value. Figure 3b indicates that when the value of outliers number ranges from 1 to 10000, the performance of our method remains stable. So, in the optimization, we do not need to spend to much time on this parameter. As for parameter λ, in the experiment, we search an optimal value from $\{10^{-2}, 10^{-1}, 1, 10, 10^2\}$

Video Recovery on Synthetic Video Clips

We mix two different video clips from UCF11 Dataset and YUV Video Sequences Dataset to simulate video consistency and variation. There are six synthetic video clips. SaleS, combination of Salesman and Suzie clip, size $144\times 176\times 3\times 50$. NewsF, combination of News and Foreman clip, size $288\times 352\times 3\times 50$. MotherW, combination of Mother And Daughter clip and Waterfall clip, size $288\times 352\times 3\times 50$. Horse, Bike, and Basketball datasets are combinations of corresponding specific action category videos. All these three videos are of size $240\times 320\times 3\times 50$.

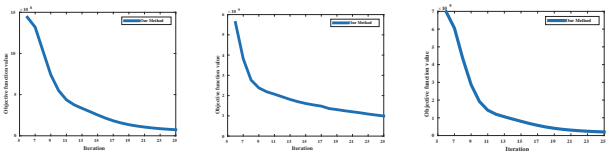
In video recovery tasks, we randomly remove 60% pixels or a small patch from each frame, and perform six compared algorithms to recover these video clips. In the experiment, $r = \{5, 10, 15, 20, 25, 30\}$ and $N_\varepsilon = 100$. Experiment results are reported in Table 1 from D1 to D6.

It is clear that these methods with capped ℓ_1-norm always show better performance than methods with ℓ_2-norm. Thus the methods considering video consistency and variation simultaneously always outperform those methods which just consider video consistency. We can also find out that sectional trace norm is better at recovering low-rank structure of image than traditional trace norm. Figure 4 shows the convergence graph of our method on all the synthetic video clips. We ignore the first 5 iterations because the value of ε is learned adaptively during this time. It is obvious that the objective function value tends to converge after 25 iterations.

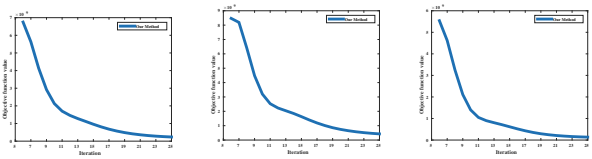
[1]http://trace.eas.asu.edu/yuv/index.html

RSE(10^{-2})						
Method	D1	D2	D3	D4	D5	D6
Tucker	2.86	2.55	2.51	4.07	4.15	1.38
LRTC	2.60	1.85	2.07	2.35	3.34	1.95
ℓ_2Tensor	2.37	1.96	1.33	2.29	3.39	1.56
capTensor	1.30	0.86	0.62	0.90	1.50	0.49
$k$msv-ℓ_2	1.38	0.87	0.85	1.12	1.54	0.79
Our method	**1.24**	**0.79**	**0.53**	**0.71**	**0.96**	**0.41**

Table 1: The RSE evaluation on SaleS, NewsF, MotherW, Horse, Bike, and Basketball video.



(a) SaleS (b) NewsF (c) MotherW



(d) Horse (e) Bike (f) Basketball

Figure 4: Convergence curves of the objective function value (3).

Video Recovery on Real Movie Clips

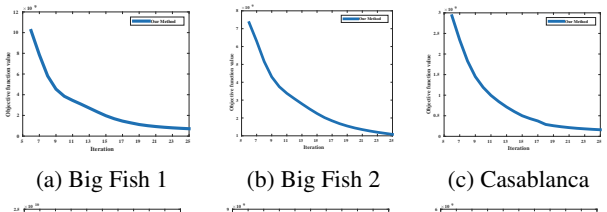
In this section, we implement the experiments on original movie clips from Hollywood Human Actions Dataset. There are six real video clips: Big Fish 1 and 2, $240\times448\times3\times100$; Casablanca, $240 \times 320 \times 3 \times 100$; Butterfly Effect, $240 \times 400 \times 3 \times 100$; LOR, $240 \times 560 \times 3 \times 100$; As Good As It Gets, $240 \times 400 \times 3 \times 100$.

RSE(10^{-2})						
Method	D7	D8	D9	D10	D11	D12
Tucker	2.81	2.99	1.63	2.99	4.18	2.54
LRTC	2.33	2.01	1.35	2.74	3.13	2.33
ℓ_2Tensor	2.13	1.12	1.01	1.68	2.81	1.58
capTensor	1.42	0.93	0.59	1.43	1.33	1.10
$k$msv-ℓ_2	1.39	1.07	0.93	1.53	1.87	1.51
Our method	**0.75**	**0.67**	**0.57**	**1.03**	**1.17**	**0.86**

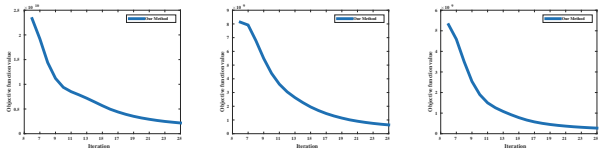
Table 2: RSE evaluation on Big Fish, Casablanca, The Butterfly Effect, LOR and As Good As It Gets.

In the experiment, we randomly mask 50% pixels or occlude a small patch from each frame, and perform six compared methods on these video clips. Parameters are set the same as in last section. Final performance are listed in the Table 2 from D7 to D12. Figure 5 shows the convergence graph of our method on all these six movie clips.

In Figure 6, we can see that the recovered images of our method are much better than others. In recovered images of other methods, the brink of occluded patch is obvious, color and pattern of the recovered patch is different from the original image.



(a) Big Fish 1 (b) Big Fish 2 (c) Casablanca



(d) Butterfly Effect (e) LOR (f) AsGoodAsItGets

Figure 5: Convergence curves of the objective function value (3).

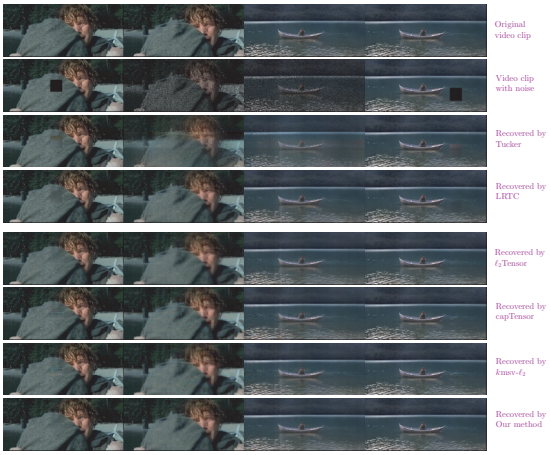


Figure 6: Video recovery results of LOR video clip.

Conclusion

In this paper, we proposed a novel video recovery model, and prove the convergence of our optimization algorithm. Capped ℓ_1-norm smoothness function is utilized as regularization to impose the video spatiotemporal consistency and variation. A new sectional trace norm is introduced to approximate rank minimization which is tighter than traditional trace norm. We evaluate our new video recovery method on several practical videos and experiment results show that our proposed method consistently outperforms other models on video recovery tasks.

References

Beck, A., and Teboulle, M. 2010. Gradient-based algorithms with applications to signal recovery problems. *Convex Optimization in Signal Processing and Communications* 42–88.

Bennett, J., and Lanning, S. 2007. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, 35.

Buades, A.; Coll, B.; and Morel, J.-M. 2005. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation* 4(2):490–530.

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.

Eldén, L. 2007. *Matrix methods in data mining and pattern recognition*, volume 4. SIAM.

Gao, H.; Nie, F.; Cai, W.; and Huang, H. 2015. Robust capped norm nonnegative matrix factorization. *24th ACM International Conference on Information and Knowledge Management (CIKM 2015)* 871–880.

Gong, P.; Zhang, C.; Lu, Z.; Huang, J. Z.; and Ye, J. 2013. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In *Machine learning: proceedings of the International Conference. International Conference on Machine Learning*, volume 28, 37. NIH Public Access.

Hu, Y.; Zhang, D.; Ye, J.; Li, X.; and He, X. 2013. Fast and accurate matrix completion via truncated nuclear norm regularization. 35(9):2117–30.

Huo, Z.; Nie, F.; and Huang, H. 2016. Robust and effective metric learning using capped trace norm. *22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2016)* 1605–1614.

Ji, H.; Liu, C.; Shen, Z.; and Xu, Y. 2010. Robust video denoising using low rank matrix completion. In *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1791–1798.

Jiang, W.; Nie, F.; and Huang, H. 2015. Robust dictionary learning with capped l1 norm. *Twenty-Fourth International Joint Conferences on Artificial Intelligence (IJCAI 2015)* 3590–3596.

Laptev, I.; Marszałek, M.; Schmid, C.; and Rozenfeld, B. 2008. Learning realistic human actions from movies. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 1–8. IEEE.

Liu, J.; Musialski, P.; Wonka, P.; and Ye, J. 2009. Tensor completion for estimating missing values in visual data. In *Proceedings / IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision*, 2114–2121.

Liu, J.; Luo, J.; and Shah, M. 2009. Recognizing realistic actions from videos in the wild. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 1996–2003. IEEE.

Shamir, O., and Shalev-Shwartz, S. 2014. Matrix completion with the trace norm: learning, bounding, and transducing. *The Journal of Machine Learning Research* 15(1):3401–3423.

Theobald, C. 1975. An inequality for the trace of the product of two symmetric matrices. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 77, 265–267. Cambridge Univ Press.

Wang, H.; Nie, F.; and Huang, H. 2014. Low-rank tensor completion with spatio-temporal consistency. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Zhang, T. 2009. Multi-stage convex relaxation for learning with sparse regularization. In *Advances in Neural Information Processing Systems*, 1929–1936.