Parts for the Whole: The DCT Norm for Extreme Visual Recovery

Yunhe Wang, Chang Xu, Shan You, Dacheng Tao, Fellow, IEEE and Chao Xu

Abstract—Here we study the extreme visual recovery problem, in which over 90% of pixel values in a given image are missing. Existing low rank-based algorithms are only effective for recovering data with at most 90% missing values. Thus, we exploit visual data’s smoothness property to help solve this challenging extreme visual recovery problem. Based on the Discrete Cosine Transformation (DCT), we propose a novel DCT norm that involves all pixels and produces smooth estimations in any view. Our theoretical analysis shows that the total variation (TV) norm, which only achieves local smoothness, is a special case of the proposed DCT norm. We also develop a new visual recovery algorithm by minimizing the DCT and nuclear norms to achieve a more visually pleasing estimation. Experimental results on a benchmark image dataset demonstrate that the proposed approach is superior to state-of-the-art methods in terms of peak signal-to-noise ratio and structural similarity.

Index Terms—Discrete Cosine Transform, Visual recovery, DCT norm, Low-rank minimization.

I. INTRODUCTION

VISUAL data can be corrupted due to sensory noise or interferential outliers during data acquisition. A fraction of image pixels could be missing sometimes, if the conditions deteriorate further, missing pixels will constantly increase, e.g., terribly-damaged images due to unstable online transmission, cameras covered by noises, photographs overexposed accidentally, and outdoor pictures taken behind a screen window. For satisfactory visual recognition, detection, and tracking, corrupted visual data must be recovered during pre-processing [22], [5], [4], [19], [6], [24].

There has recently been a surge in low rank-based matrix completion methods for visual recovery. Given a matrix $M \in \mathbb{R}^{n \times m}$, where $n$ and $m$ are the width and height of $M$, respectively, $M_{i,j} = 0, (i,j) \notin \Omega$ denote the observed data while the others represent missing data. In general, a low rank matrix $M$ can be discovered to approximately represent matrix $M$. The rank of $M$ is usually assumed to be lower than any of its two dimensionalities, i.e., $\text{rank}(M) \ll \min(n,m)$.

Fig. 1: A schematic of the calculation of the proposed DCT norm at two scales. The input image is first densely divided into several $p_i \times p_i$ patches that are then converted to the frequency domain. The DCT norm is returned by accumulating squared values of high-frequency coefficients whose positions are larger than $q_i$ in all the patches.

Although rank minimization provides an approach for recovering the missing observations, it is computationally intractable (NP-hard and non-convex). Under broad conditions, $\text{rank}(M) \ll \min(n,m)$, where $n$ and $m$ are the width and height of $M$, respectively. $M_{i,j} = 0, (i,j) \notin \Omega$ denote the observed data while the others represent missing data. In general, a low rank matrix $M$ can be discovered to approximately represent matrix $M$. The rank of $M$ is usually assumed to be lower than any of its two dimensionalities, i.e., $\text{rank}(M) \ll \min(n,m)$.

$\text{[5]}$ reported that the rank function minimization can be replaced by the trace norm $||X||_\tau = \sum_k \sigma_k (X)$ minimization, where $\sigma_k$ is the $k$-th maximum singular value of $X$. Soft-thresholding methods $[4]$ are often employed to solve optimization problems with this trace norm regularization. However, trace norms have been improved to better investigate the low rank constraint. For example, $[15]$ designed a truncated nuclear norm more suitable for matrix completion and $[13]$ proposed a weighted nuclear norm minimization for image denoising.

Existing rank minimizing techniques are effective and have delivered promising performance in many visual recovery problems. However, most of these algorithms have only been evaluated for data with at most 80% missing values $[19]$. They may not, therefore, be applicable to data with extremely high numbers of missing values (e.g., 95%) that occur when data collection devices fail or there is damage to the transmission medium. This extreme visual recovery problem with a very large number of missing pixel values is, therefore, very challenging due to the conflict between the small amount of...
observed data and the tremendous amount of data that needs to be recovered.

As well as the low rank assumption, smoothness is another important property carried by visual data. The total variation (TV) norm [23] is an effective way to exploit smoothness by calculating the differences between neighboring pixels and has been widely used in image processing. However, minimizing local differences via the TV norm risks over-smoothing the image detail and texture [3].

We consider employing the Discrete Cosine Transformation (DCT) [11] to devise a new smoothness regularizer called the “DCT norm”. By constraining the frequency coefficients of the data obtained via DCT, the DCT norm flexibly adjusts the degree of smoothness over the data (Fig.1). To investigate the local smoothness and multi-scale properties of visual data, we extend the DCT norm to derive local and multi-scale versions, respectively. Our theoretical analysis shows that the DCT norm includes the TV norm as a special case. By combining the TV norm as a special case. By combining the low rank and smoothness properties of visual data for extreme visual recovery. Experimental results on real-world datasets demonstrate that the proposed DCT norm is highly effective and that the low rank and smoothness issues need to be integrated for successful extreme visual recovery.

II. RELATED WORKS

We first briefly introduce related works on visual completion and TV norm minimization.

Nuclear norm minimization-based matrix completion [5] is formulated as

\[ \hat{X} = \arg \min_X \|X\|_* \quad \text{s.t.} \; X_M = M, \]  

(1)

where \( \hat{X} \) is the optimal estimation of \( M \). The truncated nuclear norm minimization problem [15] is defined as:

\[ \hat{X} = \arg \min_X \|X\|_F \quad \text{s.t.} \; X_M = M, \]  

(2)

where \( \|X\|_F = \sum_{i=1}^{\min(n,m)} \sigma_i(X) \) is the sum of \( \min(n,m) - r \) minimum singular values.

It is widely known that the TV norm [23] is an efficient smoothness regularizer that accumulates all the gradients of a given image \( X \):

\[ \|X\|_{TV} = \sum_{i,j} \sqrt{|X_{i+1,j} - X_{i,j}|^2 + |X_{i,j+1} - X_{i,j}|^2}, \]  

(3)

where \( i \) and \( j \) denote the vertical and horizontal positions of \( X \), respectively. Since the TV norm accumulates gradient modules of the entire image, minimizing the TV norm can result in a smooth estimation:

\[ \hat{X} = \arg \min_X \|X\|_{TV} \quad \text{s.t.} \; X_M = M. \]  

(4)

[12] proposed to simultaneously use the TV and nuclear norms for image recovery:

\[ \hat{X} = \arg \min_X \|X\|_* + \lambda \|X\|_{TV} \quad \text{s.t.} \; X_M = M, \]  

(5)

where \( \lambda \) balances the two norms. To the best of our knowledge, Fcn[5] is the first algorithm to integrate smoothness and nuclear norm regularization.

Since the \( \| \cdot \|_{TV} \) is isotropic and not differentiable, an anisotropic version is proposed in [17] that is easier to minimize:

\[ \|X\|_{anisoTV} = \sum_{i,j} |X_{i+1,j} - X_{i,j}| + |X_{i,j+1} - X_{i,j}|. \]  

(6)

Additionally, in [14], a modified linear total variation was defined as:

\[ \|X\|_{LT} = \sum_{i,j} |X_{i+1,j} - X_{i,j}|^2 + |X_{i,j+1} - X_{i,j}|^2, \]  

(7)

which leads to a smooth low-rank matrix completion problem:

\[ \hat{X} = \arg \min_X \|X\|_* + \lambda \|X\|_{LT} \; \text{s.t.} \; X_M = M. \]  

(8)

An ADMM-like optimization scheme [2] can be adopted to solve Fcn[8]. However, the traditional TV norm only can guarantee an estimation presenting a locally smooth visualization, when in reality a natural image should be smoothed at every scale. In the next section, we propose a multi-scale DCT norm in the frequency domain.

III. THE DCT NORM

A natural image has smooth regions in the spatial domain. Furthermore, there is less high-frequency information than low-frequency information in the image’s frequency domain [18]. This section presents a novel Discrete Cosine Transformation (DCT) norm for smoothing the objective variable, which provides advantages over the TV norm due to its linear and convex properties.

A. The DCT for a 2D matrix

DCT is widely used in image compression and is an approximate KL transformation [11]. For an arbitrary matrix \( X \), its DCT coefficient matrix \( C \) is:

\[ C_X = T(X) = CXC^T = (C \otimes C) vec(X), \]  

(9)

where \( \otimes \) is the Kronecker product, \( vec(\cdot) \) denotes the vectorization, and \( C \) denotes the transformation matrix with the same size as \( X \):

\[ C_{ij} = \alpha(i) \cos \left( \frac{(j+0.5) \pi}{N} \right), \]  

(10)

and

\[ \alpha(i) = \begin{cases} \sqrt{1/N}, & i = 0 \\ \sqrt{2/N}, & \text{else}. \end{cases} \]  

(11)

\( C \) has the same dimensionality with \( X \), where \( C_{0,0} \) is the DC (direct current) coefficient which only consists of the overall illumination information [7], the other coefficients in \( C \) are AC (alternating current) components denote the energies of every frequency levels, i.e., the weights of the DCT blocks as showed in Fig[7]. Additionally, it is instructive to note that DCT is a linear lossless transformation and the original data can be restored by \( X = T^{-1}(C_X) = (C \otimes C)^T C_X \). DCT is also
applied to a variety of computer vision tasks, such as image denoising [9] and image representation [25].

Since Fcn(9) involves all elements in X we can access the overall structural information of the entire matrix. The DCT and its gradient can be quickly calculated using linear transformations such that the proposed DCT norm-based optimization problems can be efficiently and easily solved.

B. The DCT norm for smoothing

The global smooth DCT norm. Neighboring pixels in a natural image are generally significantly correlated. On the other hand, the abnormal signal (e.g., noise, missing values, etc.) can be seen as a set of external data subject to an i.i.d. distribution. Hence, the frequency distributions of natural images and the abnormal signals are distinct (Fig.1). The observation, we design a DCT norm in the frequency domain:

\[ ||X||_{DCT} = ||S_q \ast (CXC^T)||^2_F \]

(12)

where \( \ast \) denotes the Hadamard product, \( || \cdot ||_F \) denotes the Frobenius norm, and \( S_q \) is a selection mask with parameter \( c \) denoting the cut-off position:

\[ S_q = \begin{cases} 
S_{ij} = 1, & i \leq q & j \leq q \\
S_{ij} = 0, & \text{otherwise.} 
\end{cases} \]

(13)

The smoothing-oriented visual recovery problem can thus be formulated as:

\[ \hat{X} = \arg \min_X ||X||^q_{DCT}, \text{ s.t. } X_{\Omega} = M_{\Omega}. \]

(14)

Although we can obtain a smooth estimation by solving the above problem, the optimal solution of Fcn(14) will have some deformations as shown in Fig.4. This is due to some remaining frequency coefficients with positions lower than \( q \) still needing to be discarded since they were difficult to estimate. A local smoothness regularization can eliminate these deformations. Therefore, we expand the global smooth DCT norm to a more comprehensive model that can also represent the local smoothness of the given image.

The locally smooth DCT norm. Inspired by non-local denoising methods [13, 13], we divide the corrupted observation into several small patches or the locally smooth DCT norm:

\[ ||X||^{p,q}_{DCT} = \sum_l ||S_{p,q} \ast (C_{p}^{r}(C_{p}^{r})^T)||^2_F \]

(15)

where \( p \) is the scale parameter, \( ||X||_{DCT} \) will be equivalent to Fcn(12) given \( p = N \), \( S_{p,q} \) and \( C_p \) denote the mask and the transformation matrix generated according to \( p \), respectively, \( x_p^{(l)} \) is a \( p \times p \) matrix denoting the \( (l-th \) patch extracted from \( X \), \( X_p = \text{vec}(x_p^{(1)}), \text{vec}(x_p^{(2)}), \ldots, \text{vec}(x_p^{(p)}) \) stacks \( x_p^{(l)} \) into a matrix, and \( S_{p,q} = \text{vec}(S_{p,q}), \ldots, \text{vec}(S_{p,q}) \), which has the same dimensionality to that of \( X \). Patches extracted at every pixel are overlapping. For an \( N \times M \) image, there are \( (N - p + 1) \times (M - p + 1) \) patches with size \( p \times p \).

The resulting optimization function w.r.t. the local DCT norm is:

\[ \hat{X} = \arg \min_X ||X||^{p,q}_{DCT}, \text{ s.t. } X_{\Omega} = M_{\Omega}. \]

(16)

The multi-scale DCT norm. Both the locally smooth DCT norm and the global smooth DCT norm have pros and cons and can be further refined. Inspired by research on the local descriptor [20, 8], detecting and describing a key point based on multiple scales can capture more geometric information. Hence, we propose to integrate local and global smoothness by integrating DCT norms from multiple image scales:

\[ \hat{X} = \arg \min_X \sum_{i=1}^s ||X||^{p_i,q_i}_{DCT}, \text{ s.t. } X_{\Omega} = M_{\Omega}. \]

(17)

where \( s \) indicates the number of scales. If \( s = 1 \) and \( p = N \), Fcn(17) will focus on global smoothness. Given \( s = 1 \) and \( p = 2 \), Fcn(17) will be reduced to the local DCT norm minimization problem in Fcn(15).

C. Relationship to the TV Norm

The TV norm is an efficient smoothing tool (see Fcn(3)). The following theorem suggests that the TV norm can be regarded as a special case of DCT norm, thereby demonstrating the superiority of the proposed DCT norm.

Theorem 1. Given \( X \in \mathbb{R}^{n \times m} \), the optimal solution of Fcn(4) with the linear TV norm is exactly that of the DNM problem in Fcn(16) with \( p = 2 \) and \( q = 1 \).

Proof. For any \( 2 \times 2 \) patch \( x = [x_{0,0}, x_{1,0}, x_{0,1}, x_{1,1}]^T \) in \( X \), its DCT coefficient matrix is \( C = T(x) \) and the local DCT norm is:

\[ ||x||^{2,1}_{DCT} = C_{0,1}^2 + C_{1,0}^2 + C_{1,1}^2, \]

(18)

where

\[ C_{0,1} = \frac{1}{2}((x_{0,0} - x_{0,1}) + (x_{1,0} - x_{1,1})), \]
\[ C_{1,0} = \frac{1}{2}((x_{0,0} - x_{1,0}) + (x_{0,1} - x_{1,1})), \]
\[ C_{1,1} = \frac{1}{2}((x_{0,0} + x_{1,1}) - (x_{0,1} + x_{1,0})), \]

and the TV norm of \( X \) is calculated as:

\[ ||x||_{TV} = (x_{0,0} - x_{0,1})^2 + (x_{0,0} - x_{1,0})^2 + (x_{0,1} - x_{1,0})^2. \]

(20)

Let \( f(x) = 4||x||^{2,1}_{DCT} \), \( e_1 = [1,1,-1,-1]^T, e_2 = [1,-1,1,-1]^T, e_3 = [1,-1,-1,-1]^T \). We have

\[ f(x) = (e_1^T x)^2 + (e_2^T x)^2 + (e_3^T x)^2 = x^T e_1 e_1^T x + x^T e_2 e_2^T x + x^T e_3 e_3^T x = \sum_{i=1}^3 x^T e_i e_i^T x, \]

(21)

Similarly, let \( g(x) = ||x||_{TV} \), \( d_1 = [1,0,-1,0]^T, d_2 = [1,-1,0,0]^T, d_3 = [0,1,0,-1]^T, d_4 = [0,0,1,-1]^T \). We have

\[ g(x) = \sum_{i=1}^4 x^T d_i d_i^T x, \]

(22)
Consider two minimizations $\min f(x)$ and $\min g(x)$. Their optimal solutions satisfy:

$$
\nabla_x f(x) = 2 \sum_{i=1}^{3} e_i e_i^T x = E x = 0
$$

and

$$
\nabla_x g(x) = 2 \sum_{i=1}^{4} d_i d_i^T x = D x = 0,
$$

where $E = 2 \sum_{i=1}^{3} e_i e_i^T$, $D = 2 \sum_{i=1}^{4} d_i d_i^T$ and their null spaces are equivalent $N(E) = N(D)$.

Furthermore, it is easy to demonstrate the null spaces of $E$ and $D$ are identical for any $N \times M$ image by formatting its $E$ and $D$ according to Fcn. 23. Hence, the optimal solution of DNM is equal to that of the TV norm minimization.

According to Theorem 1, we conclude that $\|x\|_{DCT}^2$ is equivalent to the TV norm. Moreover, we suggest that the proposed approach in Fcn. 17 has benefits over that of the TV norm since there is flexible to control over the degree of smoothness by operating at multiple scales.

![Observations](image1)

![TV norm](image2)

![DCT norm](image3)

**Fig. 2:** A comparison of the DCT norm and TV norm in a removal problem. Left to right show corrupted observations and reconstructed images using the TV norm and DCT norm, respectively.

The TV norm is noted for its strong capacity for image inpainting, e.g., text removal [15]. This is a challenging task, because the pixels covered by the text are not randomly distributed. Hence, the optimization function with a regularization that encourages global smoothness might be expected to perform well. Since color images have three channels (i.e., red, green, and blue), for fair comparison with the TV norm we simply deal with each channel separately and combine them to obtain the final result. As illustrated in Fig. 2, the visual recovery results obtained with the proposed DCT norm are superior to those of the TV norm. The results show that the proposed DCT norm outperforms the TV norm because it takes all pixels into consideration. Here, we use a two-scale DCT norm with $p_1 = 2$, $p_2 = 512$ (image sizes), $q_1 = 1$, and $q_2 = 256$. Obviously, the peak signal-to-noise ratio (PSNR) values of the proposed DCT norm are about $2dB$ higher than that of the TV norm.

**IV. DCT NORM FOR VISUAL RECOVERY**

Image completion aims to restore an original image $X$ from its corrupted observation $M$ with an observed region $\Omega$. Recovering missing values in a matrix with limited observed information has recently attracted considerable interest [19], [6], [15].

This problem is commonly addressed with inpainting [15] or denoising [13], methods, especially when the missing ratio is not too high ($\leq 80\%$). The non-local means [et al.2005] and its variations [13], [10] are the current state-of-the-art techniques and exploit the self-similarity characteristic of images. However, when the observation is very corrupt, e.g., the missing ratio is higher than $90\%$ (see Fig. 4), non-local based algorithms are less useful since they cannot find similar patches in the given image.

Base on the proposed multi-scale DCT norm and existing rank minimization techniques, we establish an efficient optimization problem that contains the various factors mentioned above:

$$
\hat{X} = \arg \min_{X} \|X\|_F + \sum_{i} \lambda_i \|X\|_{DCT}^{x_i} + \gamma \left\| P_{\Omega} (X) - P_{\Omega} (M) \right\|_F^2,
$$

where $\lambda_i$ denotes the weighting parameter for the DCT norm $\|X\|_{DCT}^{x_i}$ of $X$ in the $i$-th scale. $\gamma > 0$ is a relaxation factor that converts the original problem into an unconstrained minimization [15]. The relaxed problem can be solved using a gradient-based algorithm [16].

Fcn. 24 is naturally designed for the recovery of gray images. However, it is easy to extend Fcn. 24 to handle color images:

$$
\hat{X} = \sum_{c} \arg \min_{X} \|X(c)\|_F + \sum_{i} \lambda_i \|X(c)\|_{DCT}^{x_i} + \gamma \left\| P_{\Omega} (X(c)) - P_{\Omega} (M) \right\|_F^2
$$

where $X$ is an RGB image and $X(c)$ denotes the $c$-th channel of $X$, i.e., $X(1)$, $X(2)$, $X(3)$ corresponding to the red, green, and blue channels, respectively. An additional regularization with constant weight $\alpha$ is employed in Fcn. 25 to encourage DCT coefficients in any two channels of $X$ to be similar:

$$
\|X(c)\|_{freq} = \frac{1}{4} \sum_{i \neq c} \| S_i * C (X(c) - X(\theta)) C^T \|_F^2,
$$

where $S_i^\theta_{ij} = 0$ if $i = 0$ and $j = 0, S_i^\theta_{ij} = 1$ otherwise is a mask to remove the DC coefficient because the illumination information between channels is different. Although the value of $\|X(c)\|_{freq}$ is equal to that calculated in the spatial domain after subtracting the average value of each channel, the calculation in the frequency domain will assign greater weights to low-frequency coefficients. This is important for constructing...
images of better visual quality. The gradient of the DCT norm is:

\[
G \left( \|X_{(c)}\|_{DCT} \right) = \frac{\partial \sum_i \lambda_i \|X_{(c)}\|_{DCT}^{p_i,q_i}}{\partial X_{(c)}} = \sum_i \lambda_i \mathcal{I} \left( (C_{p_i} \otimes C_{p_i})^T [S_{p_i,q_i} \ast (C_{p_i} \otimes C_{p_i} X_{(c)})] \right),
\]

where \( \mathcal{I}(\cdot) \) is an operation that recovers the stacking matrix \( X_{p_i} \) into the original image and \( \mathcal{I}(X_{p_i}) = X_{(c)} \). The gradient of \( \|X_{(c)}\|_{freq} \) is

\[
G \left( \|X_{(c)}\|_{freq} \right) = \frac{\partial \|X_{(c)}\|_{freq}}{\partial X_{(c)}} = \sum_{i,p} C^T \left( S^c \ast C \left( X_{(c)} - X_{(i)} \right) C^T \right) C.
\]

An iterative TV norm [21] has also been proposed, and its optimal estimation is superior to that of directly solving the TV norm of the input observation. In this paper, we borrow this iterative strategy [10], [13] and extend the DCT norm to an iterative method:

\[
M^{(k+1)} = M^{(k)} + \delta P_\Omega \left( M - X^{(k)} \right),
\]

where \( k \) is the iteration number and \( \delta \) is a relaxation parameter (often set to 0.1). We describe the image completion method by exploiting the proposed multi-scale DCT norm as showed in Alg[1]

**Algorithm 1** DCT norm minimization for visual recovery.

**Input:** The corrupted observation \( M^0 = M, \hat{X}^0 = M \) and the corresponding parameters;

1. repeat
2. \( X^k \leftarrow M^k; \)
3. repeat
4. for each scale \( p_i \) do
5. Divide \( X^k \) into several \( p_i \times p_i \) patches;
6. Accumulate the gradient of \( X^k \) in the \( p_i \) scale;
7. end for
8. Calculate the gradient, except the nuclear norm;
9. Obtain the low-rank approximation;
10. until convergence
11. \( M^{(k+1)} \leftarrow M^{(k)} + \delta P_\Omega \left( M - X^{(k)} \right); \)
12. until \( \|M^{(k+1)} - X^{(k)}\|_F \leq \epsilon \)

**Output:** The estimated image \( \hat{X} \);

| \( \phi \) | SVT | TNNR | LTVNN | SAIST | DNM |
|---|---|---|---|---|---|
| 90% | 19.299 | 19.947 | 26.134 | 25.928 | **28.865** |
| 95% | 16.984 | 14.125 | 23.408 | 23.682 | **26.182** |
| 98% | 15.149 | 7.543 | 20.289 | 21.301 | **23.377** |
| 99% | 14.313 | 5.458 | 18.245 | 19.605 | **21.704** |

| \( \phi \) | SVT | TNNR | LTVNN | SAIST | DNM |
|---|---|---|---|---|---|
| 90% | 0.7170 | 0.6970 | 0.9125 | 0.8876 | **0.9472** |
| 95% | 0.6300 | 0.4281 | 0.8647 | 0.8470 | **0.9161** |
| 98% | 0.5134 | 0.1063 | 0.7917 | 0.7989 | **0.8712** |
| 99% | 0.4291 | 0.0544 | 0.7275 | 0.7624 | **0.8391** |

**V. Experiments**

**Experimental setup.** Experiments were carried on eight images [11] widely used for evaluating the performance of image restoration algorithms (Fig[3]). This is a benchmark dataset widely used for evaluating the visual recovery performance, with a variety of scenes.

**Fig. 3:** The test images.

Observations were generated by randomly sampling a small proportion (ranging from 1% to 10%) of pixels from the images subject to a Gaussian distribution. We used two standard criteria to evaluate the recovery performance: PSNR and structural similarity (SSIM) [25].

**Parameter settings.** The proposed completion algorithm has several important parameters: \( p_i, q_i, \) and \( \lambda_i \). In the experiments, \( p_i \) was set as 2, 8, and 512 (the size of each image) to obtain a better smoothness estimation and \( q_1 = 1, q_2 = 4, q_3 = 192 \), which denote the locally, blockly, and globally smooth regularizations, respectively. The weight parameter \( \alpha \) was set to \( 10^{-3} \), \( \gamma \) was set to 0.5, and \( \lambda_1 = \lambda = 1.5 \times 10^{-2} \) empirically. For the DNM stop conditions, we set the tolerance \( \epsilon \) in Alg[1] to \( 10^{-8} \). The parameter \( \tau \) for truncating singular values was set to 192, which is much larger than that in [15]. Note that, the proposed algorithm is not sensitive to most of parameters, thus all of the parameters were set empirically which is a common setting in the context of low-rank minimization [15].

**Experimental results.** We conducted experiments using different image completion algorithms to compare the performance of the proposed DCT norm with current methods. Therefore, completion experiments were carried using the proposed DNM and state-of-the-art comparison algorithms: SVT [4], TNNR [15], LTVNN [14], and SAIST [10]. Most of the comparative methods were conducted using the source code provided by authors. SVT is reported as the baseline since it is the cornerstone of the low rank approach, and TNNR is an enhanced low rank completion algorithm that is known to produce better results. LTVNN embeds the linear TV norm into the low rank minimization. SAIST is a non-local method that improves the estimated results after taking an interpolation as its initial value.

Since TNNR will be degenerated into SVT while the parameter \( \tau \) in \( \|X\|_F \) is set to 0, we set \( \tau = 5 \) in TNNR to compare the results of TNNR and SVT. It is interesting to note that TNNR’s estimation is better than that of SVT.
Fig. 4: The completion results for various corrupted images from top to bottom and their missing pixel rates: Girl (90%), Butterfly (95%), Parthenon (98%), and Plants (99%). Corrupted observations are shown in the first column and the reconstructed images using SVT, TNNR, LTVNN, SAIST, G-DNM, L-DNM, and the proposed DNM are shown from left to right. DNM results are more visually pleasing than those of previous state-of-the-art methods.

when the missing rate $\phi$ is less than 90%, while it is inferior to SVT when $90 < \phi$. This is because TNNR retains the largest $r$ singular values, which depict the major structure of the input image, but are most likely reflect the structure of the corrupted image with a considerable missing rate; hence, TNNR is inefficient. Since the proposed DCT norm can conductively recover image structure, we set a larger threshold $r = 192$ in DNM to avoid undesired lines caused by excessive minimization of the nuclear norm. Results are showed in Tab[I] and Tab[II]. The values are the average of the eight test images, where the highest evaluation result in each case is highlighted in bold. The proposed visual recovery algorithm based on the DCT norm clearly outperforms the others.

We also performed a qualitative comparison of the different completion algorithms. Results on different scenes (Girl, Butterfly, Parthenon, Plants) and missing ratios between 90% and 99% are shown in Fig[III]. We also report the results exploiting the global and local DCT norms, marked as G-DNM and L-DNM in the figure. G-DNM produces some uncoordinated grids due excessive reduction of necessary high-frequency information, and the results of L-DNM are similar to that of the TV norm. Since images produced by DNM are smooth and natural, they have a better perceptual quality. It is obvious that the result of G-DNM has a clearer overall structure, while its local regions are not smooth enough. The result of L-DNM presents an opposite phenomenon. The multi-scale DNM achieves the best performance by combining them.

The SVT results contain several lines because this algorithm slightly undermines the image structure when shrinking all singular values; therefore, the TNNR estimation is more visually pleasing when the missing pixel rate is not too high. Although the results of LTVNN and SAIST are better than those of the previous two algorithms, LTVNN is not particularly clear and SAIST over-smooths some important structures. Specifically, SAIST post-processes after initializing the image using an interpolation method, which only completes the missing pixels by exploiting neighborhood information; thus, it lacks overall structure. Furthermore, SAIST is useless when the input corrupted image does not possess the initialization produced using interpolation methods. The estimated images of the proposed algorithm are clear, sharp, and visually pleasing because the multi-scale DCT norm makes the image and its patches at different scales smooth and natural. Specifically, it obtains an estimation in which the neighborhood pixels inside are smoothed using the locally smooth DCT norm and it can also obtain an estimation that produces a globally smooth output by exploiting the globally smooth DCT norm.

VI. DISCUSSION AND CONCLUSIONS

Most existing rank-minimizing techniques do not efficiently handle data with over 90% missing values. Therefore, we propose a powerful smooth regularization to overcome this problem: the DCT norm. Compared to the traditional TV norm, the proposed scheme involves all the pixel values and can guarantee estimation smoothness at different scales. Moreover, we demonstrate that the TV norm can be regarded as a special case of the DCT norm. By combining the truncated nuclear norm and the proposed scheme we establish an efficient
image completion model. Experiments show that the estimated images using the proposed multi-scale DCT norm are more visually pleasing than those produced by the previous state-of-the-art. Additionally, the proposed smooth regularization can be independently embedded into most image processing tasks, e.g., image inpainting and image denoising.

REFERENCES

[1] N. Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transform. Computers, IEEE Transactions on, 100(1):90–93, 1974.
[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning, 3(1):1–122, 2011.
[3] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In CVPR, pages 60–65, 2005.
[4] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization, 20(4):1956–1982, 2010.
[5] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. Foundations of Computational mathematics, 9(6):717–772, 2009.
[6] E. J. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. Information Theory, IEEE Transactions on, 56(5):2053–2080, 2010.
[7] W. Chen, M. J. Er, and S. Wu. Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. TSMC, Part B: Cybernetics, 36(2):458–466, 2006.
[8] H. Cheng, Z. Liu, N. Zheng, and J. Yang. A deformable local image descriptor. In CVPR, pages 1–8, 2008.
[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. TIP, 16(8):2080–2095, 2007.
[10] W. Dong, G. Shi, and X. Li. Nonlocal image restoration with bilateral variance estimation: a low-rank approach. Image Processing. IEEE Transactions on, 22(2):700–711, 2013.
[11] W. Dong, L. Zhang, and G. Shi. Centralized sparse representation for image restoration. In ICCV, pages 1259–1266, 2011.
[12] M. Golbabaee and P. Vandergheynst. Joint trace/tv norm minimization: A new efficient approach for spectral compressive imaging. In ICML, pages 457–464, 2009.
[13] X. Han, J. Wu, L. Wang, Y. Chen, L. Senhadji, and H. Shu. Linear total variation approximate regularized nuclear norm optimization for matrix completion. Abstract and Applied Analysis, 2014.
[14] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He. Fast and accurate matrix completion via truncated nuclear norm regularization. TPAMI, 35(9):2117–2130, 2013.
[15] S. Ji and J. Ye. An accelerated gradient method for trace norm minimization. In ICML, pages 457–464, 2009.
[16] X. Jin, L. Li, Z. Chen, L. Zhang, and Y. Xing. Anisotropic total variation for limited-angle ct reconstruction. In Nuclear Science Symposium Conference Record, pages 2232–2238, 2010.
[17] E. Y. Lam and J. W. Goodman. A mathematical analysis of the dct coefficient distributions for images. TIP, 9(10):1661–1666, 2000.
[18] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. TPAMI, 35(1):208–220, 2013.
[19] G. Lowe. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2):91–110, 2004.
[20] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. Multiscale Modeling & Simulation, 4(2):460–489, 2005.
[21] Y. Pi, H. Peng, S. Zhou, and Z. Zhang. A scalable approach to column-based low-rank matrix approximation. In IJCAI, pages 1600–1606, 2013.
[22] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 60(1):259–268, 1992.
[23] Z. Shen, H. Qian, T. Zhou, and S. Wang. Simple atom selection strategy for greedy matrix completion. In IJCAI, pages 1799–1805, 2015.
[24] R. Tjahyadi, W. Liu, S. An, and S. Venkatesh. Face recognition via the overlapping energy histogram. In IJCAI, pages 2891–2896, 2007.
[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. TIP, 13(4):600–612, 2004.