

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/314275724>

# DCT Regularized Extreme Visual Recovery

Article in IEEE Transactions on Image Processing · March 2017

DOI: 10.1109/TIP.2017.2678798

---

CITATIONS

10

---

READS

168

5 authors, including:



Yunhe Wang

Peking University

15 PUBLICATIONS 32 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



visual recovery [View project](#)



Deep Neural Network Compression [View project](#)

# DCT Regularized Extreme Visual Recovery

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

**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 Transform (DCT), we propose a novel DCT regularizer that involves all pixels and produces smooth estimations in any view. Our theoretical analysis shows that the total variation (TV) regularizer, which only achieves local smoothness, is a special case of the proposed DCT regularizer. We also develop a new visual recovery algorithm by minimizing the DCT regularizer and nuclear norm 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 regularizer, 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 on-line 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 [36], [10], [9], [32], [11], [39].

There has recently been a surge in low rank-based matrix completion methods for visual recovery. Given a matrix  $X \in \mathcal{R}^{N \times M}$ , where  $N$  and  $M$  are the width and height of  $X$ , respectively,  $X_{i,j} = 0$ ,  $(i, j) \notin \Omega$  denote the observed data while the others represent missing data. In general, a low rank matrix  $\hat{X}$  can be discovered to approximately represent matrix  $X$ . The rank of  $\hat{X}$  is usually assumed to be lower than any of its two dimensionalities, *i.e.*,  $rank(\hat{X}) \ll \min(N, M)$  [17], [31], [34]. While, for a corrupted observation, the inside conjunct information would be damaged, means that its last several singular values must be larger than that of the original image. Thus minimizing the rank of an observation can effectively estimate a natural image.

This work was supported by NSFC 61375026 and 2015BAF15B00, and ARC Projects: ARC FT-130101457, DP-140102164, LP-150100671.

Y. Wang, S. You, and C. Xu are with the Key Laboratory of Machine Perception (Ministry of Education) and Cooprtative Medianet Innovation Center, School of EECS, Peking University, Beijing 100871, P.R. China. E-mail: wangyunhe@pku.edu.cn, youshan@pku.edu.cn, xuchao@cis.pku.edu.cn.

C. Xu and D. Tao are with the UBTech Sydney Artificial Intelligence Institute and the School of Information Technologies in the Faculty of Engineering and Information Technologies at The University of Sydney, J12 Cleveland St, Darlington NSW 2008, Australia. Email: c.xu@sydney.edu.au, dacheng.tao@sydney.edu.au.

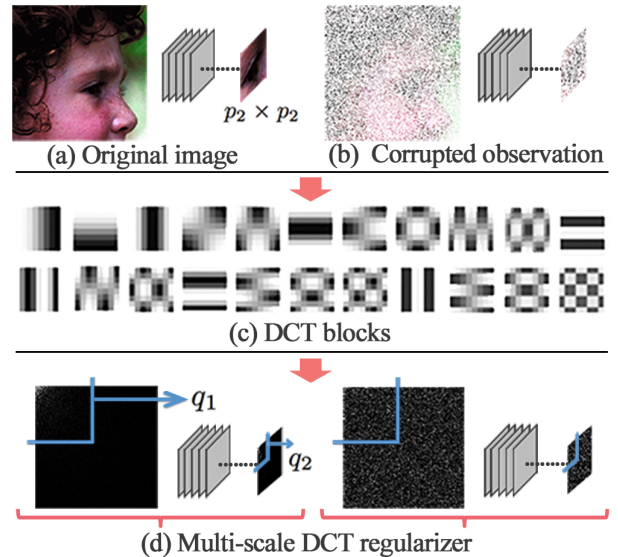


Fig. 1: A schematic of the calculation of the proposed DCT regularizer at two scales. The input image is first densely divided into several  $p_i \times p_i$  patches that are then converted to the frequency domain. The DCT regularizer is returned by accumulating squared values of high-frequency coefficients 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, [10] reported that the rank function minimization can be replaced by the trace norm  $\|X\|_* = \sum_k \sigma_k(X)$  minimization, where  $\sigma_k$  is the  $k$ -th maximum singular value of  $X$ . Soft-thresholding methods [9], [27] 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, [26] designed a truncated nuclear norm more suitable for matrix completion and [23] 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 [32]. Largest singular values usually contain the major structure and texture of the input image, thus rank minimization will maintain the noise or damaged pixels when the observation is terribly corrupted. Therefore, conventional low-rank schemes may be not 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 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 since there are many continuous areas and structures in a natural image [37], [16]. The total variation (TV) regularizer [38], [1], [2] 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 regularizer risks over-smoothing the image detail and texture [8], [5]. Additionally, minimizing the TV regularizer can only make any small patch in the estimated image smooth but neglect the global perception and structure.

We consider employing the Discrete Cosine Transform (DCT [3]) to devise a new smoothness regularizer called the ‘‘DCT regularizer’’. By constraining the high frequency coefficients of the data obtained via DCT, the DCT regularizer flexibly adjusts the degree of smoothness over the data (Fig. 1). Actually, the DCT regularizer embedded into the entire frequency domain of an image only can make this image smooth integrally. In order to investigate the local smoothness and multi-scale properties of visual data, we extend the DCT regularizer to derive local and multi-scale versions, respectively. Our theoretical analysis shows that the DCT regularizer includes the TV regularizer as a special case. By combining the classical rank minimization principle with the proposed DCT regularizer minimization (DRM), the resulting model can simultaneously exploit the low rank and smoothness properties of visual data for extreme visual recovery. Experimental results on real-world datasets demonstrate that the proposed DCT regularizer is highly effective and that the low rank and smoothness issues need to be integrated for successful extreme visual recovery.

This paper is organized as follows. Section II investigates related works. Section III and Section IV propose the DCT regularizer and combines it with the nuclear norm to formulate a novel visual recovery. Section V describes the results of our experimental validation. Section VI concludes the paper.

## II. RELATED WORKS

We first briefly introduce related works on visual completion based on the conventional nuclear norm and TV regularizer minimization.

Since an image can be seen as a continuous integration, whose main structure and texture are held by the first several largest singular values. Given a corrupted observation  $Y$ , nuclear norm minimization-based matrix completion [10] is formulated as

$$\hat{X} = \arg \min_X \|X\|_*, \quad s.t. \quad X_\Omega = Y_\Omega, \quad (1)$$

where  $X$  is the estimated image,  $\|X\|_*$  denotes the sum of all the singular values of  $X$  and  $\hat{X}$  is the optimal estimation of the corrupted observation  $Y$ . Since useless information in  $Y$  is commonly contained by singular values whose values are relatively small, minimize the above function can effectively

eliminate the noise and completely corrupted observations. Fcn. 1 is often solved by the classical soft-thresholding method.

However, shrinkage all singular values simultaneously is not perfect in any case. Thus, [26] proposed the truncated nuclear norm which maintains the first  $r$  largest singular values. The truncated nuclear norm minimization problem is defined as:

$$\hat{X} = \arg \min_X \|X\|_r, \quad s.t. \quad X_\Omega = Y_\Omega, \quad (2)$$

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

Additionally, the low-rank assumption is not sufficient when the observed image has been damaged terribly. Thus we should introduce more prior knowledge as regularizations. Another effective tool for image recovery is the TV regularizer [38], which is widely known that it is an efficient smoothness regularization which accumulates all the gradients of a given image  $X$ :

$$\|X\|_{TV} = \sum_{i,j} \sqrt{|X_{i+1,j} - X_{i,j}|^2 + |X_{i,j+1} - X_{i,j}|^2}, \quad (3)$$

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

$$\hat{X} = \arg \min_X \|X\|_{TV}, \quad s.t. \quad X_\Omega = Y_\Omega. \quad (4)$$

Since both TV regularizer and nuclear norm have their own advantages the TV regularizer provides the local smoothness and the nuclear norm can maintain the overall structure. Thus [21] proposed to simultaneously use the TV regularizer and nuclear norm for image recovery:

$$\hat{X} = \arg \min_X \|X\|_* + \lambda \|X\|_{TV}, \quad s.t. \quad X_\Omega = Y_\Omega, \quad (5)$$

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, [29] proposed an anisotropic version:

$$\|X\|_{anisoTV} = \sum_{i,j} |X_{i+1,j} - X_{i,j}| + |X_{i,j+1} - X_{i,j}|. \quad (6)$$

However, gradients of the original TV regularizer and the anisotropic TV regularizer are complex which makes the problem hard to analyze and optimize. Thus in [25], a modified linear total variation was defined as:

$$\|X\|_{LTV} = \sum_{i,j} |X_{i+1,j} - X_{i,j}|^2 + |X_{i,j+1} - X_{i,j}|^2. \quad (7)$$

It is obvious that the above modified TV regularizer is linear thus it can be easily solved by some traditional gradient based approach, which leads to a smooth low-rank matrix completion problem:

$$\hat{X} = \arg \min_X \|X\|_* + \lambda \|X\|_{LTV}, \quad s.t. \quad X_\Omega = Y_\Omega. \quad (8)$$

An ADMM-like optimization scheme [6] can be adopted to solve Fcn. 8. Actually, the introduction of the conventional

TV regularizer minimization in Fcn. 5 and Fcn. 8 can be seen as a supplement of the low-rank completion scheme, which makes the nuclear norm more comprehensive. However, all types of traditional TV regularizers are calculated by accumulating neighboring pixels, which neglect the internal relationship between any two pixels in an image. Additionally, the traditional TV regularizer only can guarantee an estimation presenting a locally smooth visualization, when in reality a natural image should be smoothed at every scale. Since any component in the frequency domain is calculated by weighted summarizing all pixels of the original image, thus a process in the frequency domain will involve all pixel values. In the next section, we propose a multi-scale DCT regularizer in the frequency domain.

### III. THE DCT REGULARIZER

Generally, most regions in a natural image are smooth and conjunct in the spatial domain. When we transfer the image into the frequency domain, there is less high-frequency information than low-frequency information in the image's frequency domain [30]. This section presents a novel Discrete Cosine Transform (DCT) norm for smoothing the objective variable, which provides advantages over the TV regularizer due to its linear and convex properties.

#### A. The DCT for a 2D image

DCT is widely used in image compression and is an approximate KL transformation [3], [40], [20], which also be employed in the field of visual description [43] and denoising [44]. For an arbitrary matrix  $X \in \mathbb{R}^{N \times M}$ , its DCT coefficient matrix  $\mathcal{C}$  is:

$$\begin{aligned} \mathcal{C}_{j_1, j_2} &= \alpha_{j_1} \alpha_{j_2} \sum_{i_1=0}^{N-1} \sum_{i_2=0}^{M-1} c(i_1, i_2, j_1, j_2) X_{i_1, i_2} \\ &= C_{j_1, j_2} * X, \end{aligned} \quad (9)$$

where  $N$  and  $M$  are height and width of  $X$ , respectively.  $*$  is the image convolution, and  $C_{j_1, j_2}(i_1, i_2) = \alpha_{j_1} \alpha_{j_2} c(i_1, i_2, j_1, j_2)$  with the same size of  $X$  is generated to calculate the  $j_1, j_2$ -th DCT coefficient of  $X$ .  $\alpha_{j_1} = \sqrt{1/N}$  if  $j_1 = 0$  and  $\alpha_{j_1} = \sqrt{2/N}$ , otherwise.  $\alpha_{j_2} = \sqrt{1/M}$  if  $j_2 = 0$  and  $\alpha_{j_2} = \sqrt{2/M}$ , otherwise.  $c(\cdot, \cdot, \cdot, \cdot)$  represents the cosine basis function:

$$\begin{aligned} c(i_1, i_2, j_1, j_2) &= \\ &= \cos\left(\frac{\pi(2i_1 + 1)j_1}{2N}\right) \cos\left(\frac{\pi(2i_2 + 1)j_2}{2M}\right). \end{aligned} \quad (10)$$

In order to have explicit description and explanation, let  $X = \text{vec}(X)$ , we can simply reformulate the DCT in matrix form as  $\mathcal{C} = \mathbf{C}X$ , where  $\mathbf{C} = [\text{vec}(C_{1,1}), \dots, \text{vec}(C_{m,n})]$  is an orthogonal matrix, *i.e.*,  $\mathbf{C}^T \mathbf{C} = \mathbf{I}$ .  $\mathcal{C}$  has the same dimensionality with  $X$ , where  $\mathcal{C}_0$  is the DC (direct current) coefficient which only consists of the overall illumination information [12], the other coefficients in  $\mathcal{C}$  are AC (alternating current) components denote the energies of every frequency level, *i.e.*, the weights of the DCT blocks as showed in Fig. 1(c). Additionally, it is instructive to note that DCT is

a linear lossless transformation and the original data can be restored by  $X = \mathbf{C}^T \mathcal{C}$ . DCT is also applied to a variety of computer vision tasks, such as image denoising [15] and image representation [41], [24].

Since Fcn. 9 involves all elements in  $X$  we can access the overall structural information of the entire matrix, *i.e.*, any operation in the frequency domain involves all pixels in the spatial domain. Moreover, the DCT and its gradient can be quickly calculated using linear transformations such that the proposed DCT regularizer-based optimization problems can be efficiently and easily solved.

#### B. The DCT regularizer for smoothing

**The global smooth DCT regularizer.** Neighboring pixels in a natural image are generally significantly correlated. On the other hand, the abnormal signal (*e.g.*, noise, missing values) 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 high-frequency information of the original image is much lower than that of the corrupted observation  $Y$ . Based on this observation, we design a DCT regularizer in the frequency domain:

$$\|X\|_{DCT}^q = \|S_q * \mathbf{C}X\|_F^2, \quad (11)$$

where  $*$  denotes the Hadamard product,  $\|\cdot\|_F$  denotes the Frobenius norm, and  $S_q$  is a selection mask:

$$S_q = \begin{cases} S_{ij} = 0, & i \leq q \ \& \ j \leq q \\ S_{ij} = 1, & \text{otherwise.} \end{cases} \quad (12)$$

where  $q < N$  is the cut-off position. As is discussed above, Fcn. 11 aims to remove the abnormal high-frequency information thus  $S_q$  can be seen as a penalty acting on high-frequency coefficients.

The smoothing-oriented visual recovery problem by exploiting the proposed DCT regularizer can thus be formulated as:

$$\hat{X} = \arg \min_X \|X\|_{DCT}^q, \quad \text{s.t. } X_\Omega = Y_\Omega. \quad (13)$$

Although we can obtain a smooth estimation by solving the above problem, the optimal solution of Fcn. 13 will have some deformations as shown in Fig. 2 and Fig. 4. This is due to some remaining frequency coefficients with positions lower than  $q$  still needing to be refined and they are difficult to estimate. Obviously, a local smoothness regularization can eliminate these deformations. Therefore, we expand the global smoothed DCT regularizer to a more comprehensive model that can also represent the local smoothness of the given image.

**The locally smooth DCT regularizer.** Inspired by non-local denoising methods [8], [23], we divide the corrupted observation into several small patches. For an  $N \times M$  image, there are  $n = (N - p + 1) \times (M - p + 1)$  patches with size  $p \times p$ . Patches extracted at every pixel are overlapping. Then, we propose to employ the DCT regularizer to every small patch to form the locally smooth DCT regularizer:

$$\begin{aligned} \|X\|_{DCT}^{p,q} &= \sum_{l=1}^n \|S_{p,q} * \mathbf{C}x_p^{(l)}\|_F^2 \\ &= \|S_{p,q} * \mathbf{C}X_p\|_F^2, \end{aligned} \quad (14)$$

where  $p$  is the scale parameter, *e.g.*, we divide the input observation into  $2 \times 2$  patches when  $p = 2$ . Thus,  $\|X\|_{DCT}^{p,q}$  will be equivalent to Fcn. 11 given  $p = N$ ,  $S_{p,q}$  denotes the mask generated according to  $p$ ,  $x_p^{(l)}$  is a  $p \times p$  matrix denoting the  $l$ -th patch extracted from  $X$ ,  $\mathbf{X}_p = [\text{vec}(x_p^{(1)}), \text{vec}(x_p^{(2)}), \dots, \text{vec}(x_p^{(l)})] \in \mathbb{R}^{p^2 \times n}$  stacks  $x_p^{(l)}$  into a matrix, and  $\mathbf{S}_{p,q} = [\text{vec}(S_{p,q}), \dots, \text{vec}(S_{p,q})] \in \mathbb{R}^{p^2 \times n}$ , which has the same dimensionality to that of  $\mathbf{X}$ .

Correspondingly, the resulting optimization function w.r.t. the local DCT regularizer is:

$$\hat{X} = \arg \min_X \|X\|_{DCT}^{p,q}, \quad s.t. X_\Omega = Y_\Omega. \quad (15)$$

Since the input image may have different degrees of damage, we can select the local DCT regularizer and the global DCT regularizer according to the real demand. Moreover, the functionality of the proposed local DCT regularizer is similar to that of the conventional TV regularizer. Both of them are designed to make local regions in the corrupted observation smooth. Actually, the TV regularizer is a particular case of the local DCT regularizer when  $p = 2$  and  $q = 1$ , while calculated in the spatial domain. We will further discuss it in Theorem 1. **The multi-scale DCT regularizer.** Both the locally smooth DCT regularizer and the global smooth DCT regularizer have pros and cons and can be further refined. Inspired by research on the local descriptor [33], [13], detecting and describing a key point based on multiple scales can capture more geometric information. Similarly, we also can make the estimated image smooth in different views by combining the proposed local DCT regularizer and global DCT regularizer. Hence, we propose integrating local and global smoothness by integrating DCT regularizers from multiple image scales:

$$\hat{X} = \arg \min_X \sum_{i=1}^s \|X\|_{DCT}^{p_i, q_i}, \quad s.t. X_\Omega = Y_\Omega. \quad (16)$$

where  $s$  indicates the number of scales. If  $s = 1$  and  $p = N$ , Fcn. 16 will focus on global smoothness. Given  $s = 1$  and  $p = 2$ , Fcn. 16 will be reduced to the local DCT regularizer minimization problem in Fcn. 14. Hence we can design a multi-scale DCT regularizer with different functionalities by giving various  $p_i$  and  $q_i$  and utilize it to form a visual recovery problem.

### C. Relationship to the TV regularizer

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

**Theorem 1.** *Given  $X \in \mathbb{R}^{N \times M}$ , the optimal solution of Fcn. 4 with the linear TV regularizer is exactly that of the DRM (DCT regularizer minimization) problem in Fcn. 15 with  $p = 2$  and  $q = 1$ .*

*Proof.* Since both the local DCT regularizer and the linear TV regularizer can be regarded as aggregations of pixel differences that are calculated in all  $2 \times 2$  patches in  $X$ . Thus we first take

an arbitrary  $2 \times 2$  patch divided from  $X$  here as an example, denote as  $x = [x_{0,0}, x_{1,0}, x_{0,1}, x_{1,1}]^T$ . The DCT coefficient matrix of  $x$  is  $\mathcal{C} = \mathcal{T}(x)$  and the corresponding local DCT regularizer with  $q = 1$  is

$$\|x\|_{DCT}^{2,1} = \mathcal{C}_{0,1}^2 + \mathcal{C}_{1,0}^2 + \mathcal{C}_{1,1}^2, \quad (17)$$

where

$$\begin{aligned} \mathcal{C}_{0,1} &= \frac{1}{2} ((x_{0,0} - x_{0,1}) + (x_{1,0} - x_{1,1})), \\ \mathcal{C}_{1,0} &= \frac{1}{2} ((x_{0,0} - x_{1,0}) + (x_{0,1} - x_{1,1})), \\ \mathcal{C}_{1,1} &= \frac{1}{2} ((x_{0,0} + x_{1,1}) - (x_{0,1} + x_{1,0})). \end{aligned} \quad (18)$$

Since the linear TV regularizer is superior to the conventional TV regularizer [25] and has a more obvious analyzability, thus the discussion here is focusing on the linear TV regularizer, who is calculated as:

$$\|x\|_{LTV} = (x_{00} - x_{01})^2 + (x_{00} - x_{10})^2 + (x_{10} - x_{11})^2 + (x_{01} - x_{11})^2. \quad (19)$$

Let  $g(x) = \|x\|_{LTV}$ ,  $d_1 = [1, 0, -1, 0]^T$ ,  $d_2 = [1, -1, 0, 0]^T$ ,  $d_3 = [0, 1, 0, -1]^T$ , and  $d_4 = [0, 0, 1, -1]^T$ . We have

$$g(x) = \sum_{i=1}^4 x^T d_i d_i^T x. \quad (20)$$

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

$$\begin{aligned} f(x) &= (e_1^T x)^2 + (e_2^T x)^2 + (e_3^T x)^2 \\ &= x^T e_1 e_1^T x + x^T e_2 e_2^T x + x^T e_3 e_3^T x \\ &= \sum_{i=1}^3 x^T e_i e_i^T x. \end{aligned} \quad (21)$$

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

$$\begin{aligned} \nabla_x g(x) &= 2 \sum_{i=1}^4 d_i d_i^T x = Gx = 0, \\ \nabla_x f(x) &= 2 \sum_{i=1}^3 e_i e_i^T x = Fx = 0, \end{aligned} \quad (22)$$

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

Furthermore, for the  $N \times M$  image  $X$  we can also construct its  $\hat{G}$  and  $\hat{F}$  as:

$$\begin{aligned} \hat{G} &= 2 \sum_{i=1}^n \sum_{m=1}^4 \hat{d}_{i,m} \hat{d}_{i,m}^T, \\ \hat{F} &= 2 \sum_{i=1}^n \sum_{m=1}^3 \hat{e}_{i,m} \hat{e}_{i,m}^T, \end{aligned} \quad (23)$$

where  $n = (N-1) \times (M-1)$  is the number of patches divided by  $\|x\|_{DCT}^{2,1}$ ,  $\hat{d} = [\hat{d}_{1,1}, \dots, \hat{d}_{1,4}, \dots, \hat{d}_{n,1}, \dots, \hat{d}_{n,4}] \in \mathbb{R}^{NM \times 4n}$  and  $\hat{e} = [\hat{e}_{1,1}, \dots, \hat{e}_{1,3}, \dots, \hat{e}_{n,1}, \dots, \hat{e}_{n,3}] \in \mathbb{R}^{NM \times 3n}$ .  $\hat{d}_{n,m} = \text{vec}(D_{n,m})$  with

$$P_{\Omega_n}(D_{n,m}) = \text{vec}(d_k), \quad (24)$$

and  $\hat{e}_{n,m} = \text{vec}(E_{n,m})$  with

$$P_{\Omega_n}(E_{n,m}) = \text{vec}(e_k), \quad (25)$$

where  $D_{n,m} \in \mathbb{R}^{N \times M}$  and  $E_{n,m} \in \mathbb{R}^{N \times M}$ .  $P_{\Omega_n}(X)$  denote the components of  $X$  in  $\Omega_n$  which is the area of the  $n$ -th patch in  $X$ . It is obvious that the null spaces of  $\hat{G}$  and  $\hat{F}$  are identical according to Fcn. 22. Hence the optimal solution of DRM is equal to that of the TV regularizer minimization for any image.  $\square$

According to Theorem 1, we also can estimate images with local smoothness by replacing the traditional TV regularizer with the proposed DCT regularizer.

#### IV. DCT REGULARIZER FOR VISUAL RECOVERY

This section discusses the visual recovery theory based on the proposed multi-scale DCT regularizer. Image completion aims to restore a clean image from its corrupted observation  $Y$  with an observed region  $\Omega$ . Recovering missing values in a matrix with limited observed information has recently attracted considerable interest [32], [11], [26].

##### A. Model

This problem is commonly addressed with inpainting [26] or denoising [23] methods, especially when the missing ratio is not too high ( $\leq 80\%$ ). The non-local means [8] and its variations [23], [17] 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 without a reconstruction by interpolation.

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

$$\begin{aligned} \hat{X} = \arg \min_X \|X\|_* + \sum_{i=1}^s \lambda_i \|X\|_{DCT}^{p_i, q_i} \\ + \frac{\gamma}{2} \|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(Y)\|_F^2, \end{aligned} \quad (26)$$

where  $\lambda_i$  denotes the weighting parameter for the DCT regularizer  $\|X\|_{DCT}^{p_i, q_i}$  of  $X$  in the  $i$ -th scale.  $\gamma > 0$  is a relaxation factor that converts the original problem into an unconstrained minimization [26].

##### B. Optimization.

The relaxed problem can be solved using the accelerated proximal gradient line method (APGL) [28]. We first reformulate Fcn. 26 as

$$\hat{X} = \arg \min_X F(X) = \|X\|_* + f(X), \quad (27)$$

where

$$f(X) = \sum_i \lambda_i \|X\|_{DCT}^{p_i, q_i} + \frac{\gamma}{2} \|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(Y)\|_F^2, \quad (28)$$

---

#### Algorithm 1 DRM for visual recovery.

---

**Input:** Observation  $Y$ ,  $\Omega$ ,  $\lambda$ ,  $\gamma$ ;

1:  $t_1 = 1$ ,  $E_1 = X_1 = Y$ ;

2: **repeat**

3:   **for** each scale  $p_i$  **do**

4:     Divide  $X^k$  into  $p_i \times p_i$  patches to form  $\mathbf{X}_{p_i}$ ;

5:   **end for**

6:   Calculate the gradient  $\nabla f(X)$  according to Fcn. 29;

7:    $(U, \Sigma, V) \leftarrow \text{svd}(E_k - t_k \nabla f(X_k))$ ;

8:    $X_{k+1} \leftarrow U \mathcal{T}_{t_k}(\Sigma) V^T$ ;

9:    $t_{k+1} \leftarrow \frac{1 + \sqrt{4t_k^2 + 1}}{2}$ ;

10:    $E_{k+1} \leftarrow X_{k+1} + \frac{t_k - 1}{t_{k+1}}(X_{k+1} - X_k)$ ;

11:    $k \leftarrow k + 1$ ;

12: **until**  $\|E_{k+1} - X_k\|_F \leq \epsilon$

**Output:** The estimated image  $\hat{X} = E_{k+1}$ ;

---

and the gradient of  $f(X)$  is

$$\begin{aligned} \nabla f(X) = \gamma (\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(Y)) + \\ \sum_i \lambda_i \mathcal{I}(\mathbf{C}^T [\mathbf{S}_{p_i, q_i} * (\mathbf{C} \mathbf{X}_{p_i})]), \end{aligned} \quad (29)$$

where  $\mathcal{I}(\cdot)$  is an operation that recovers the stacking matrix  $\mathbf{X}_{p_i}$  into the original image and  $\mathcal{I}(\mathbf{X}_{p_i}) = X$ .

For any  $t > 0$  and a given point  $E$  of  $F(X)$ , APGL constructs an approximation

$$\begin{aligned} Q(X, Y) = \|X\|_* + \langle X - E, \nabla f(E) \rangle + \\ \frac{1}{2t} \|X - E\|_F^2 + f(E), \end{aligned} \quad (30)$$

and for the  $k$ -th iteration, we can update  $X_{k+1}$  by:

$$\begin{aligned} X_{j+1} = \arg \min_X Q(X, E_k) \\ = \arg \min_X \|X\|_* + \frac{1}{2t_k} \|X - (E_k - t_k \nabla f(X_k))\|_F^2. \end{aligned} \quad (31)$$

Fcn. 31 can be easily solved by utilizing the soft-thresholding technique, *i.e.*,

$$X_j = U \mathcal{T}_{t_k}(\Sigma) V^T, \quad (32)$$

where  $U^T \Sigma V = E_k - t_k \nabla f(X_k)$  is the singular value decomposition (SVD) and

$$\mathcal{T}_{t_k}(\Sigma_{ii}) = \max(\Sigma_{ii} - t_k, 0), \quad (33)$$

where  $\Sigma_{ii}$  is the  $i$ -th largest singular value and  $t_{k+1}$  and  $E_{k+1}$  are updated as

$$\begin{aligned} t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ E_{k+1} = X_{k+1} + \frac{t_k - 1}{t_{k+1}}(X_{k+1} - X_k). \end{aligned} \quad (34)$$

We describe the image completion method by exploiting the proposed multi-scale DCT regularizer as showed in Algorithm 1.



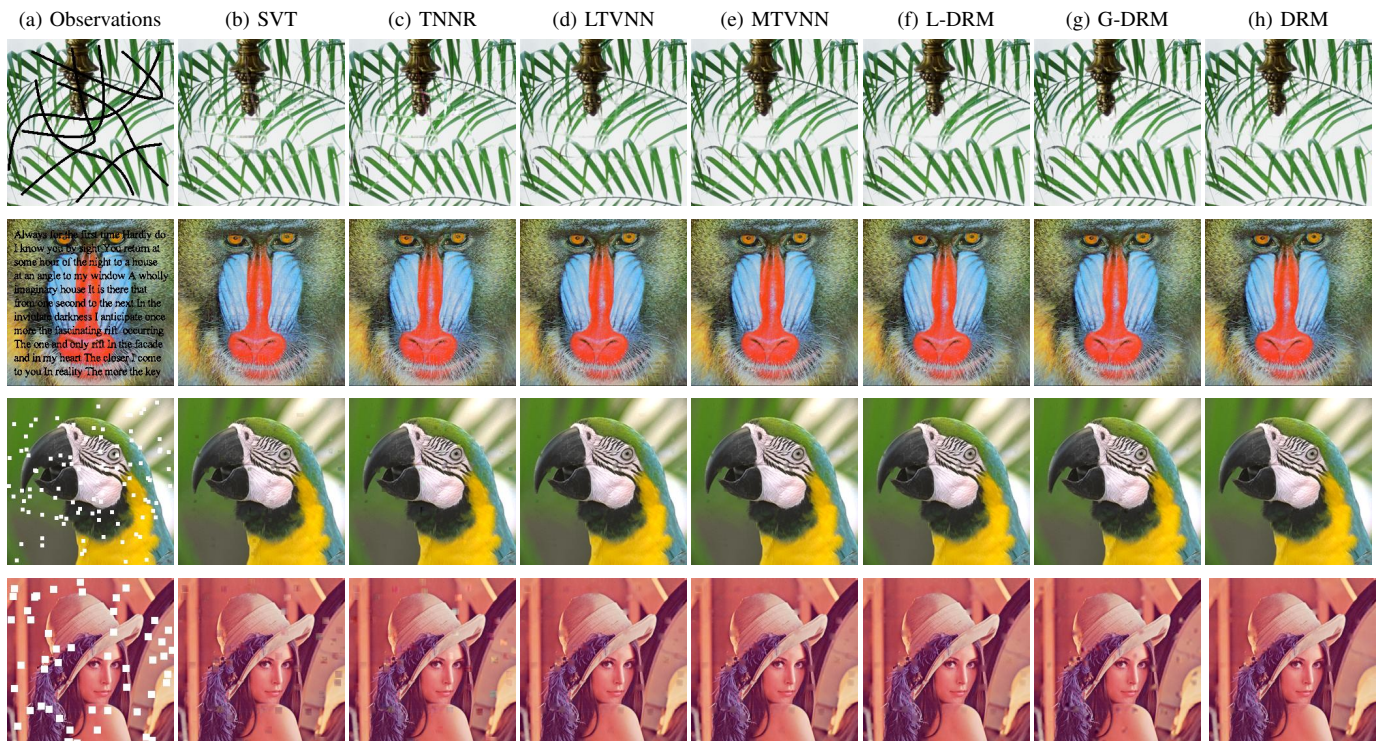


Fig. 2: A comparison of the DCT regularizer and other schemes in removal problems. From left to right, the figures show corrupted observations and reconstructed images using the state-of-the-art methods.

## V. EXPERIMENTS

### A. Experimental setup

**The test dataset.** Experiments were carried on 14 images [18], [23] 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, *e.g.*, human faces, buildings, plants, animals. In addition, they are RGB color images with size of  $256 \times 256$ .



Fig. 3: The test images.

**Compared schemes.** We conducted experiments using different image completion algorithms to compare the performance of the proposed DCT regularizer with current methods. Therefore, completion experiments were carried using the proposed DRM and state-of-the-art comparison algorithms: SVT [9], TNNR [26], and LTVNN [25]. Most of the comparative methods were conducted using source codes 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 regularizer into the low rank minimization. In order to have an explicit demonstration of the superiority of the proposed multi-scale DCT regularizer, we also report the results exploiting the global and local DCT regularizers, marked as G-DRM and L-DRM.

**Parameter settings.** The proposed completion algorithm has several important parameters:  $p_i$ ,  $q_i$  and  $\lambda_i$  with  $i = 1, \dots, s$ , it is obvious that a DCT regularizer with a larger  $s$  has a big computational complexity, but has smoothness in more views. In the experiments,  $p_i$  was set as 2 and 256 (the size of each image) to obtain a better smoothness estimation and  $q_1 = 1, q_2 = 128$ , which denote the locally and globally smooth regularizations, respectively. Accordingly, in order to have a fair comparison,  $p$  and  $q$  in the L-DRM and G-DRM are set to be  $p = 2, q = 1$ , and  $p = 256, q = 128$ , respectively. The weight parameter  $\gamma$  was set to 0.06,  $\lambda_1 = 0.25$ , and  $\lambda_2 = 0.5$ . For the algorithm stop conditions, we set the tolerance  $\epsilon$  in Alg. 1 to  $10^{-4}$  and the maximum iteration number  $K = 1000$ . In fact, 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 [26]. Since TNNR will be degenerated into SVT while the parameter  $r$  in  $\|X\|_r$  is set to 0, we set  $r = 5$  in TNNR to compare the results of TNNR and SVT. Note that the proposed DCT regularizer adopts the original nuclear norm since when the corruption has been terribly damaged the first several largest singular values are also corrupted significantly, we will further illustrate this phenomenon in Fig. 4.

TABLE I: An overall comparison on the performance of the removal experiment.

Corruption type		SVT	TNNR	LTVNN	MTVNN	L-DRM	G-DRM	DRM
Text	PSNR( <i>dB</i> )	28.64	28.63	31.60	31.75	32.01	30.91	<b>32.59</b>
	SSIM	0.9612	0.9617	0.9791	0.9799	0.9815	0.9748	<b>0.9832</b>
	Time(s)	<b>8.31</b>	26.00	17.62	23.83	18.86	11.53	17.40
Scratch	PSNR( <i>dB</i> )	29.01	29.20	33.05	33.21	33.37	31.71	<b>34.27</b>
	SSIM	0.9539	0.9579	0.9826	0.9834	0.9844	0.9741	<b>0.9858</b>
	Time(s)	8.92	29.82	9.31	14.49	8.70	<b>4.26</b>	7.32
block 5	PSNR( <i>dB</i> )	32.64	32.90	35.15	35.29	35.55	33.83	<b>36.03</b>
	SSIM	0.9828	0.9836	0.9907	0.9910	0.9918	0.9869	<b>0.9921</b>
	Time(s)	<b>7.56</b>	14.51	14.70	20.85	16.59	10.94	14.92
block 10	PSNR( <i>dB</i> )	28.82	28.22	31.03	31.19	31.58	29.24	<b>32.23</b>
	SSIM	0.9658	0.9648	0.9776	0.9783	0.9801	0.9703	<b>0.9822</b>
	Time(s)	<b>9.19</b>	45.98	23.61	36.30	32.86	49.48	38.36

### B. Compared with the TV regularizer

As we have discussed above, the low-rank assumption is not sufficient when the observed image has been corrupted seriously. Thus a powerful regularization should be introduced to make the estimated image smooth. Since the traditional TV regularizer regularization can also be utilized for estimating a clean and smooth image we first compare the proposed DCT regularizer with the TV regularizer on several benchmark experiments for evaluating the smoothnesses of their desired images.

According to Theorem 1, we conclude that the optimization problem formulated by  $\|x\|_{DCT}^{2,1}$  is exactly that of the TV regularizer. Moreover, it is obvious that the proposed multi-scale DCT regularizer has more functionalities than the conventional TV regularizer. Thus we suggest that the proposed approach in Fcn. 16 has benefits over that of the TV regularizer since there is flexible to control over the degree of smoothness by operating at multiple scales.

The TV regularizer is noted for its strong capacity for image inpainting, *e.g.*, text removal, scratch removal [26]. 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. Additionally, these removal tasks can also be performed by the conventional nuclear norm algorithms [26]. Thus we first compare the proposed DCT regularizer with others by conducting these removal experiments. Generally, there are three types of masks for testing the performance of removal algorithms.

**Text & scratch masks.** Text & scratch removal is a classical challenge for visual recovery since pixels covered by text or scratch are distributed unevenly, and some important structure and texture of the original image may be damaged.

**Block mask.** Block missing is another common situation, due to images and videos are usually compressed after dividing into several patches (*e.g.*, JPEG [42] & JPEG2000 [14]). Although an image can be seen as a matrix with continuous pixels, there is still some subtle independent texture in any block area. Thus, it is hard to estimate the desired image when there are many missing blocks on the corrupted observation.

We generated corrupted observations by exploiting three

masks as showed in Fig. 2. Then, we conduct the removal experiment using the proposed DCT regularizer and the state-of-the-art methods. We used two standard criteria to evaluate the recovery performance: PSNR and structural similarity (SSIM) [45]. Results are showed in Tab. I, these values are averages of the 14 test images, where block 5 denotes there are 100 missing blocks on the corrupted observation and their size is  $5 \times 5$ , and block 10 denotes there are 50 missing blocks on the corrupted observation and their size is  $10 \times 10$ , respectively (see 3rd and 4th lines in Fig. 2). The proposed DRM for visual recovery is superior to the others. Obviously, peak signal-to-noise ratio (PSNR) values of the proposed DCT regularizer are about  $1dB$  higher than those of the TV regularizer and SSIM values of the proposed DCT regularizer are about 0.01 higher than those of others. Wherein, estimated results of L-DRM are similar to those of LTVNN and superior to those of SVT, TNNR, and G-DRM, which means that the local smoothness is more important than the global smoothness for images. We also reported the MATLAB running times of various methods. Obviously, running times of SVT are smaller than those of others. In addition, it is worth mentioned that the more difficult the task is, the longer running time becomes, *e.g.*, block 10 > block 5 for every method. Here we also test the visual recovery performance of the multi-scale TV regularizer [19], [7], [4], which is formulated by ranging the stride of pixels and can produce more smooth estimations, namely MTVNN. As shown in Tab. I, results of MTVNN are slightly higher than those of LTVNN but require more running time.

Besides the quantitative results, we also conduct some qualitative comparisons between different recovery schemes as showed in Fig. 2. Visual recovery results obtained with the proposed multi-scale DCT regularizer are superior to those of the traditional TV regularizer and nuclear norm. Especially, results by exploiting the proposed DCT regularizer are more clear and smooth than those of using the TV regularizer because it takes all pixels into consideration.

### C. Visual Recovery

Sec. V-B shows that the proposed DCT regularizer is superior to the traditional TV regularizer and nuclear norm based algorithms for mask removing. Here we employ the



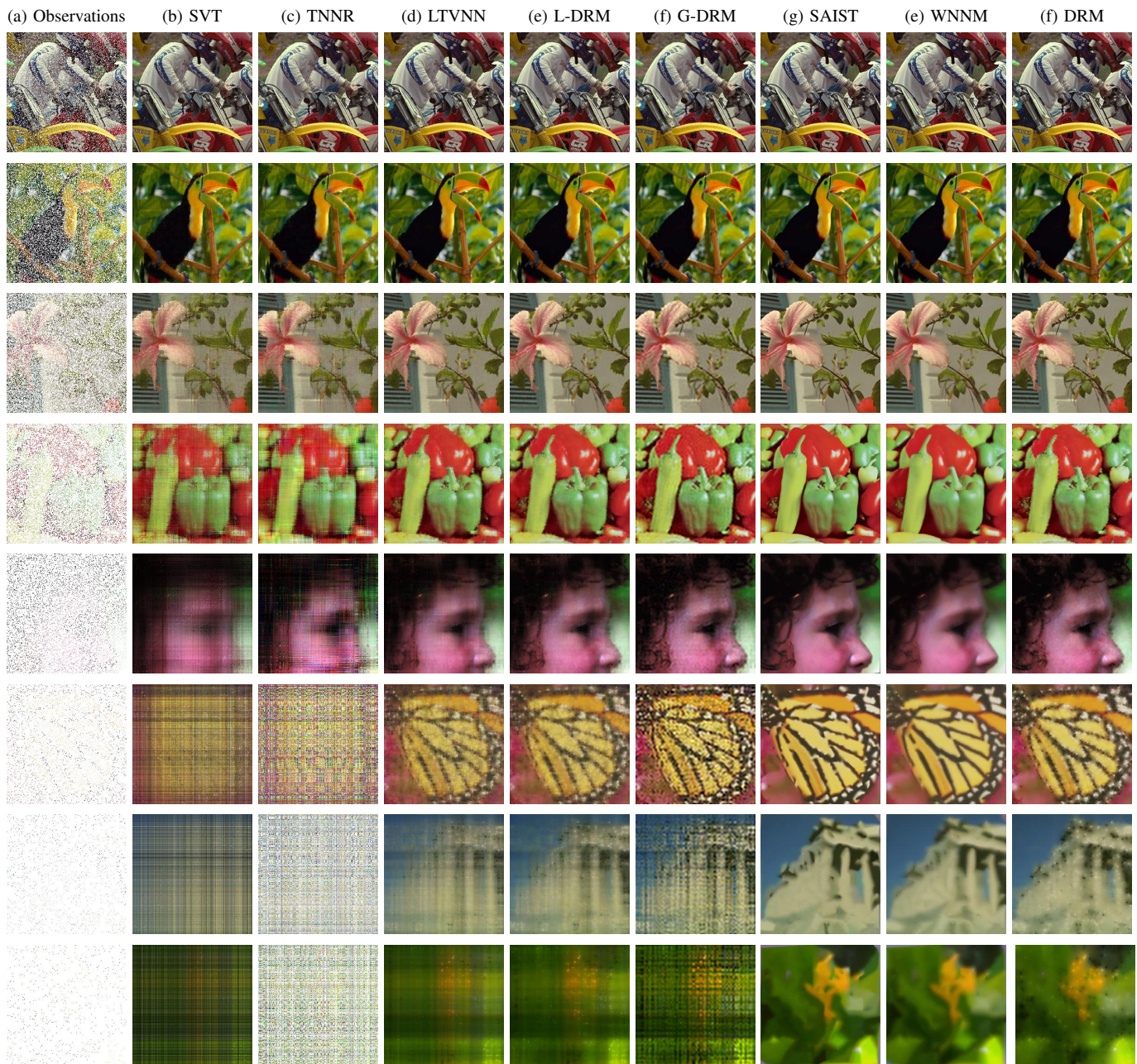


Fig. 4: The completion results for various corrupted images with different missing rate (20% ~ 99%). Corrupted observations are shown in the first column and the reconstructed images using SVT, TNNR, LTVNN, L-DRM, G-DRM, SAIST, WNNM, and the proposed DRM are shown from left to right. DRM results are more visually pleasing than those of previous state-of-the-art methods.

proposed DCT regularizer to the extreme visual recovery task. As have been discussed above, some important information and structure of an image will be covered by text or scratch, which makes the recovery task difficult. While in realistic, if the transmission or photographing conditions deteriorate further, missing pixels will constantly increase such as terribly-damaged images due to unstable online transmission, cameras covered by noises, and outdoor pictures taken behind a screen window. More essential structure and texture may be destroyed in these situations. Fig. 4 shows some synthetic images with pixel missing. Observations were generated by randomly sam-

pling a small proportion of pixels from the images subject to a Gaussian distribution. From top to bottom and their missing pixel rates: Bike (20%), Bird (40%), Flower (60%), Peppers (80%), Girl (90%), Butterfly (95%), Parthenon (98%), and Plants (99%).

As can be seen in Fig. 4, it is obvious that all methods are effective for estimating desire images when the missing rate is not too high (*e.g.*,  $\leq 80\%$ ). However, outputs of these algorithms are various when the missing rate is larger than 80%. Although the conventional rank minimization scheme can be utilized to estimate a complete observation, it is obvious

TABLE II: A comparison on the visual recovery performance of DRM and other state-of-the-art methods.

Missing Rate $\phi$		SVT	TNNR	LTVNN	L-DRM	W-DRM	SAIST	WNNM	DRM
20%	PSNR( <i>dB</i> )	31.17	31.45	35.26	35.03	32.24	<b>39.14</b>	38.93	35.54
	SSIM	0.9663	0.9692	0.9884	0.9884	0.9754	<b>0.9911</b>	0.9910	0.9879
	Time(s)	3.27	3.44	5.28	3.22	<b>2.00</b>	447.90	511.62	2.03
40%	PSNR( <i>dB</i> )	26.73	27.07	30.99	30.96	28.95	<b>34.77</b>	33.47	31.83
	SSIM	0.9161	0.9219	0.9698	0.9707	0.9484	<b>0.9787</b>	0.9731	0.9731
	Time(s)	5.22	6.07	7.68	4.94	<b>3.40</b>	445.47	1137.43	3.64
60%	PSNR( <i>dB</i> )	23.14	23.61	27.68	27.84	26.80	<b>31.10</b>	29.01	28.88
	SSIM	0.8377	0.8462	0.9381	0.9416	0.9173	<b>0.9567</b>	0.9359	0.9510
	Time(s)	8.11	13.08	11.16	7.76	6.74	447.94	951.01	<b>6.38</b>
80%	PSNR( <i>dB</i> )	19.31	19.75	24.21	24.61	24.63	<b>26.94</b>	25.16	25.66
	SSIM	0.7084	0.7044	0.8743	0.8844	0.8729	<b>0.9063</b>	0.8682	<b>0.9063</b>
	Time(s)	<b>11.20</b>	33.29	20.52	15.93	23.44	447.88	1021.28	15.09
90%	PSNR( <i>dB</i> )	16.73	14.16	21.58	22.19	22.40	<b>23.83</b>	22.83	23.28
	SSIM	0.5964	0.4435	0.8010	0.8200	0.8161	0.8419	0.8053	<b>0.8529</b>
	Time(s)	<b>10.40</b>	60.49	28.01	27.24	53.20	444.06	1088.97	27.81
95%	PSNR( <i>dB</i> )	15.15	9.17	19.40	20.17	19.93	<b>21.67</b>	21.19	21.44
	SSIM	0.5061	0.1677	0.7273	0.7558	0.7362	0.7767	0.7513	<b>0.8001</b>
	Time(s)	<b>11.85</b>	67.40	35.76	46.68	104.69	440.55	1211.31	57.17
98%	PSNR( <i>dB</i> )	14.00	6.13	16.92	17.71	16.86	19.43	19.48	<b>19.52</b>
	SSIM	0.3946	0.0636	0.6280	0.6692	0.6079	0.7003	0.6925	<b>0.7357</b>
	Time(s)	<b>15.43</b>	29.50	53.52	92.79	174.89	434.99	1242.26	118.66
99%	PSNR( <i>dB</i> )	13.42	5.52	15.60	15.93	15.15	17.99	18.21	<b>18.33</b>
	SSIM	0.3237	0.0495	0.5635	0.6059	0.5236	0.6499	0.6526	<b>0.6941</b>
	Time(s)	<b>17.41</b>	20.05	64.94	146.54	173.92	433.03	1237.84	174.00

that they will not be effective since there are too less useful pixels in the observations. Detailed comparisons are shown in Tab. II, where the highest evaluation result in each case is highlighted in bold. The proposed visual recovery algorithm based on the DCT regularizer clearly outperforms other low-rank matrix completion methods.

It is interesting to note that TNNR's estimation is better than that of SVT 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 depicts the major structure of the input image, but most likely reflects the structure of the corrupted image with a considerable missing rate; hence, TNNR is inefficient. Since the proposed DCT regularizer can conductively recover image structure, we set a larger threshold  $r = 128$  in DRM to avoid undesired lines caused by excessive minimization of the nuclear norm.

Moreover, G-DRM produces some uncoordinated grids due excessive reduction of necessary high-frequency information, and the results of L-DRM are similar to that of the TV regularizer. Since images produced by DRM are smooth and natural, they have a better perceptual quality. It is obvious that the result of G-DRM has a clearer overall structure, while its local regions are not smooth enough. The result of L-DRM presents an opposite phenomenon. The multi-scale DRM 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 are better than those of the previous two algorithms, but the estimated images by utilizing the LTVNN are not particularly clear, there are also some undesired patterns. According to Theorem 1, the proposed local DCT regularizer has the same functionality of the linear TV regularizer. Thus, the results of the LTVNN are similar to those of the L-DRM. It is obvious that the estimated images of the proposed algorithm are clear, sharp, and visually pleasing because the multi-scale DCT regularizer 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 regularizer and it can also obtain an estimation that produces a globally smooth output by exploiting the globally smooth DCT regularizer.

#### D. Comparison with Nonlocal Based Approaches

In the above section, we have shown that the proposed DRM approach has strong ability for extreme visual recovery tasks. Apparently, recovery tasks with less missing rates are directly related to the traditional image inpainting theory. Thus, it is necessary to compare the proposed DRM with the non-local based image restoration schemes which hold the highest performance in both image denoising and image inpainting [23], [22], [18], [17]. To this end, in Fig. 4 and Tab. II, we also report results of SAIST [17] and WNNM [23], which utilize low-rank assumptions and hold the highest performance.

It is worth mentioning that, these two methods need reconstruct the corrupted observations by interpolation first [17]



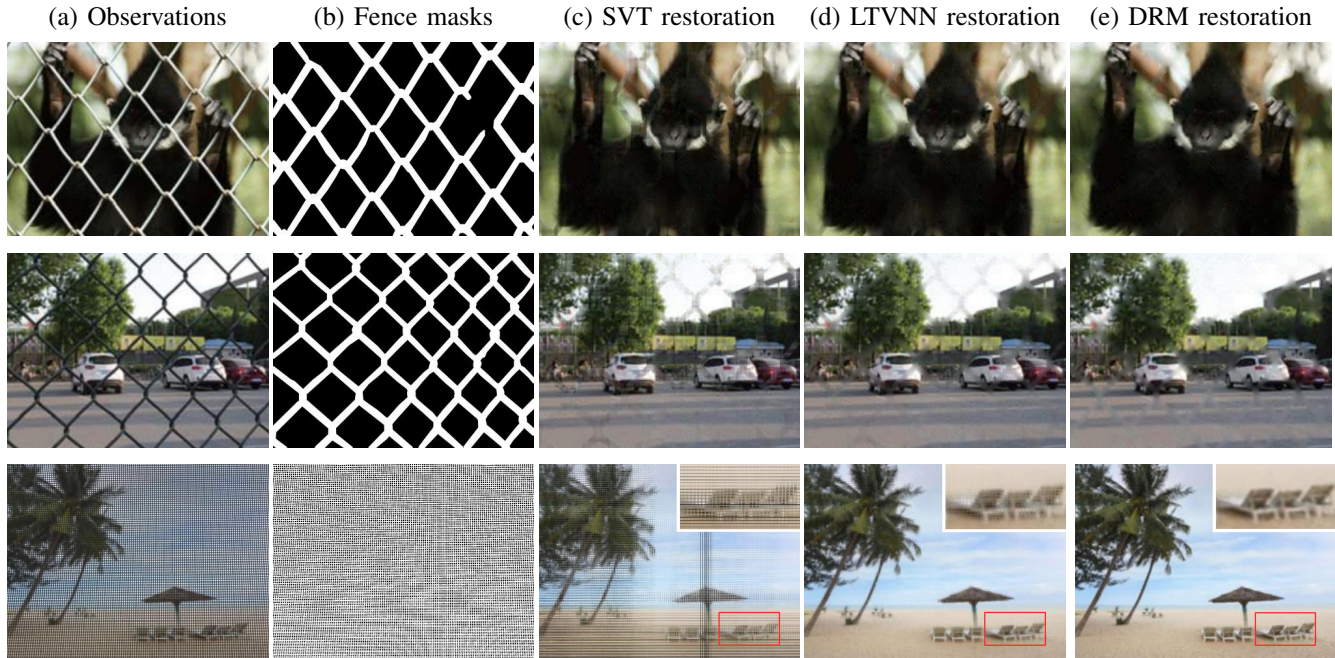


Fig. 5: A comparison of the performance of various algorithms on real world images restoration, observations are taken behind fences. (b) are masks for removing fence occlusions. (c)-(e) show the estimated images by exploiting different method.

since the “self-similarity” of original images has been damaged terribly. In addition, we used  $100 \times (1 - \phi)$  as the noise level for conducting WNNM. As shown in Fig. 4, both SAIST and WNNM can estimate clean and smooth images. However, estimated images through SAIST and WNNM have some grids when  $\phi > 90$  since blank areas are filled with surrounding pixels. Thus, their PSNR and SSIM values are inferior to those of DRM. While when  $\phi < 90$ , *i.e.*, the non-extreme case, PSNR and SSIM values of SAIST are higher than those of others benefiting from the strong restoration ability of the self-similarity priori.

Furthermore, we also reported their MATLAB running times in Tab. II. Since each patch in corrupted observations need to be processed with several similar patches simultaneously, non-local based image restoration methods have larger complexities and longer running times.

### E. Real-word Image Restoration

The above experiments are conducted on a set of benchmark test images and all the corrupted observations are generated synthetically. Actually, corrupted images are exactly the combinations of two independent images, *i.e.*, the original image and the occlusion mask. However, the requirement may be more complex in realistic. We do not only need to recovery some noise pixels of the observed image, but also need to remove some unwanted objects, *e.g.*, to remove some annoying texts and advertisements in the top of a magazine, recovery a photo taken behind a screen window or fence [46], [35] (see Fig. 5). Since the desired image gets tangled in other objects, *e.g.*, a gibbon is grasping the fence, the corrupted areas and clean areas are usually continuous, thus to implement this type

of visual recovery is much more difficult than handling the synthetic missing problem.

We take several real images taken behind fences as examples. In order to estimate the original scenes, we first mask all pixels of fences manually as shown in Fig. 5b, which is similar to the removal experiment in Fig. 2 but has more complex masks. Then, we recover these missing pixels by using the proposed DRM and the state-of-the-art methods, as illustrated in Fig. 5(c)-Fig. 5(d). Obviously, the estimated images by utilizing the proposed DRM are more clear and natural than those of the other methods. Especially, when fence occlusion is extremely dense, the traditional low-rank minimization scheme will be not applicable while the proposed DCT regularizer can also handle the problem perfectly.

### F. Discussion on Weight Matrix

The global DCT regularizer is defined by utilizing a binary mask  $S_q$  in Fcn. 11, which gives various high-frequency coefficients the same weight. While the higher frequency coefficients are generally much larger than lower frequency coefficients which should be given larger weights. Thus, we also test a weighted global DCT regularizer with the mask by exploiting the following function which gives weights of frequency coefficients from 0 to 1. For the given matrix  $X \in \mathbb{R}^{N \times M}$ , its mask is defined as

$$S_{ij} = \frac{ij}{NM}, \quad (35)$$

namely W-DRM.

Fig. 6 shows a comparison between the proposed G-DRM and W-DRM. The performance of W-DRM is slightly higher than that of G-DRM. Since coefficients of natural images are

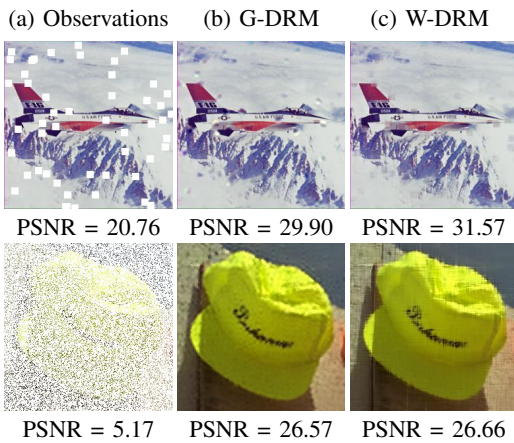


Fig. 6: A comparison between G-DRM and W-DRM.

increasing from high-frequency to low-frequency, Fcn. 35 is more suitable for handling some image restoration tasks. We will further investigate the weight matrix for constructing the global DCT regularizer in future works.

## VI. 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 regularizer. Compared to the traditional TV regularizer, the proposed scheme involves all the pixel values and can guarantee estimation smoothness at different scales. Moreover, we demonstrate that the TV regularizer can be regarded as a special case of the DCT regularizer. 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 regularizer 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] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *Image Processing, IEEE Transactions on*, 19(9):2345–2356, 2010.
- [2] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *Image Processing, IEEE Transactions on*, 20(3):681–695, 2011.
- [3] N. Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transform. *Computers, IEEE Transactions on*, 100(1):90–93, 1974.
- [4] A. Almansa, C. Ballester, V. Caselles, and G. Haro. A tv based restoration model with local constraints. *Journal of Scientific Computing*, 34(3):209–236, 2008.
- [5] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *Image Processing, IEEE Transactions on*, 18(11):2419–2434, 2009.
- [6] 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.
- [7] X. Bresson and T. F. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse problems and imaging*, 2(4):455–484, 2008.
- [8] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *CVPR*, 2005.
- [9] 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.
- [10] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.
- [11] 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.
- [12] 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.
- [13] H. Cheng, Z. Liu, N. Zheng, and J. Yang. A deformable local image descriptor. In *CVPR*, pages 1–8, 2008.
- [14] C. Christopoulos, A. Skodras, and T. Ebrahimi. The jpeg2000 still image coding system: an overview. *Consumer Electronics, IEEE Transactions on*, 46(4):1103–1127, 2000.
- [15] 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.
- [16] A. Danielyan, V. Katkovnik, and K. Egiazarian. Bm3d frames and variational image deblurring. *Image Processing, IEEE Transactions on*, 21(4):1715–1728, 2012.
- [17] 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.
- [18] W. Dong, L. Zhang, and G. Shi. Centralized sparse representation for image restoration. In *ICCV*, pages 1259–1266, 2011.
- [19] Y. Dong, M. Hintermüller, and M. M. Rincon-Camacho. Automated regularization parameter selection in multi-scale total variation models for image restoration. *Journal of Mathematical Imaging and Vision*, 40(1):82–104, 2011.
- [20] E. Feig and S. Winograd. Fast algorithms for the discrete cosine transform. *Signal Processing, IEEE Transactions on*, 40(9):2174–2193, 1992.
- [21] M. Golbabaee and P. Vanderghyest. Joint trace/tv norm minimization: A new efficient approach for spectral compressive imaging. In *ICIP*, pages 933–936, 2012.
- [22] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang. Weighted nuclear norm minimization and its applications to low level vision. *IJCV*, pages 1–26, 2016.
- [23] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, 2014.
- [24] Z. M. Hafed and M. D. Levine. Face recognition using the discrete cosine transform. *International Journal of Computer Vision*, 43(3):167–188, 2001.
- [25] 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.
- [26] 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.
- [27] P. Jain, R. Meka, and I. S. Dhillon. Guaranteed rank minimization via singular value projection. In *NIPS*, 2010.
- [28] S. Ji and J. Ye. An accelerated gradient method for trace norm minimization. In *ICML*, pages 457–464, 2009.
- [29] 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.
- [30] E. Y. Lam and J. W. Goodman. A mathematical analysis of the dct coefficient distributions for images. *TIP*, 9(10):1661–1666, 2000.
- [31] S. Li and H. Qi. A douglas-rachford splitting approach to compressed sensing image recovery using low-rank regularization. *Image Processing, IEEE Transactions on*, 24(11):4240–4249, 2015.
- [32] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *TPAMI*, 35(1):208–220, 2013.
- [33] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [34] C. Lu, Z. Lin, and S. Yan. Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization. *Image Processing, IEEE Transactions on*, 24(2):646–654, 2015.
- [35] S. McCloskey. Masking light fields to remove partial occlusion. In *ICPR*, 2014.
- [36] 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.
- [37] I. Ram, M. Elad, and I. Cohen. Image processing using smooth ordering of its patches. *Image Processing, IEEE Transactions on*, 22(7):2764–2774, 2013.



- [38] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
- [39] Z. Shen, H. Qian, T. Zhou, and S. Wang. Simple atom selection strategy for greedy matrix completion. In *IJCAI*, pages 1799–1805, 2015.
- [40] G. Strang. The discrete cosine transform. *SIAM review*, 41(1):135–147, 1999.
- [41] R. Tjahjyadi, W. Liu, S. An, and S. Venkatesh. Face recognition via the overlapping energy histogram. In *IJCAI*, pages 2891–2896, 2007.
- [42] G. K. Wallace. The jpeg still picture compression standard. *Consumer Electronics, IEEE Transactions on*, 38(1):xviii–xxxiv, 1992.
- [43] Y. Wang, M. Shi, S. You, and C. Xu. Dct inspired feature transform for image retrieval and reconstruction. *IEEE TIP*, 25(9):4406–4420, 2016.
- [44] Y. Wang, C. Xu, C. Xu, and D. Tao. Beyond rpca: Flattening complex noise in the frequency domain. In *AAAI*, 2017.
- [45] 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.
- [46] J. Yang, J. Wang, L. Liu, and C. Hou. Rifo: Restoring images with fence occlusions. In *Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on*, pages 1–6, 2015.



co-authored more than 80 publications and 5 patents in these fields.

**Chao Xu** received the B.E. degree from Tsinghua University in 1988, the M.S. degree from University of Science and Technology of China in 1991 and the Ph.D degree from Institute of Electronics, Chinese Academy of Sciences in 1997. Between 1991 and 1994 he was employed as an assistant professor by University of Science and Technology of China. Since 1997 Dr. Xu has been with School of EECS at Peking University where he is currently a Professor. His research interests are in image and video coding, processing and understanding. He has authored or



**Yunhe Wang** received the B.E degree from Xidian University in 2013. Currently, he is a Ph.D. candidate with the Key Laboratory of Machine Perception (Ministry of Education) in the Peking University. His research interests lie primarily in machine learning and computer vision.



**Chang Xu** received the B.E. degree from Tianjin University, China, and the Ph.D. degree from Peking University, China. He is currently a Lecturer with the School of Information Technologies and the Faculty of Engineering and Information Technologies in the University of Sydney. His research interests lie primarily in machine learning, multimedia search and computer vision.



**Dacheng Tao** (F'15) is Professor of Computer Science and ARC Future Fellow in the School of Information Technologies and the Faculty of Engineering and Information Technologies, and the Founding Director of the UBTech Sydney Artificial Intelligence Institute at the University of Sydney. He was Professor of Computer Science and Director of Centre for Artificial Intelligence in the University of Technology Sydney. He mainly applies statistics and mathematics to Artificial Intelligence and Data Science. His research interests spread across computer vision, data science, image processing, machine learning, and video surveillance. His research results have expounded in one monograph and 500+ publications at prestigious journals and prominent conferences, such as IEEE T-PAMI, T-NNLS, T-IP, JMLR, IJCV, NIPS, CIKM, ICML, CVPR, ICCV, ECCV, AISTATS, ICDM; and ACM SIGKDD, with several best paper awards, such as the best theory/algorithm paper runner up award in IEEE ICDM'07, the best student paper award in IEEE ICDM'13, and the 2014 ICDM 10-year highest-impact paper award. He received the 2015 Australian Scopus-Eureka Prize, the 2015 ACS Gold Disruptor Award and the 2015 UTS Vice-Chancellor's Medal for Exceptional Research. He is a Fellow of the IEEE, OSA, IAPR and SPIE.



**Shan You** received the B.E degree from Xi'an Jiaotong University in 2014. Currently, he is a Ph.D. candidate with the Key Laboratory of Machine Perception (Ministry of Education) in the Peking University. His research interests lie primarily in machine learning and computer vision.