

Semi-sparsity Priors for Image Structure Analysis and Extraction

Junqing Huang, Haihui Wang, Michael Ruzhansky

Abstract—Image structure-texture decomposition is a long-standing and fundamental problem in both image processing and computer vision fields. In this paper, we propose a generalized semi-sparse regularization framework for image structural analysis and extraction, which allows us to decouple the underlying image structures from complicated textural backgrounds. Combining with different textural analysis models, such a regularization receives favorable properties differing from many traditional methods. We demonstrate that it is not only capable of preserving image structures without introducing notorious staircase artifacts in polynomial-smoothing surfaces but is also applicable for decomposing image textures with strong oscillatory patterns. Moreover, we also introduce an efficient numerical solution based on an alternating direction method of multipliers (ADMM) algorithm, which gives rise to a simple and maneuverable way for image structure-texture decomposition. The versatility of the proposed method is finally verified by a series of experimental results with the capability of producing comparable or superior image decomposition results against cutting-edge methods.

Index Terms—Semi-sparsity model, image structure extraction, structure/cartoon-texture decomposition, image texture filtering.

1 INTRODUCTION

Natural images always contain various well-organized objects together with some regular/irregular textural patterns, which convey rich information for both machine and human vision. In many psychology and perception [1], [2], it has also found that the contribution of visual cues mostly comes from geometrical structures of objects rather than local individual details. Interestingly, it is always easy for human visual system (HSV) to distinguish main objects from the complex textural backgrounds. Based on these observations, it is always desirable to decouple an image into different meaningful components, for example, a structural part and one or multiple textural parts [3], [4]. The problem is known as image structure-texture decomposition and has gained great attention in research fields because it helps to extract and analyze the preferred information of images. Such an image decomposition is also proved to be beneficial for a variety of vision-based applications such as image edge extraction [3], [5], cartoon and abstraction [5], tone mapping and enhancement [6], [7], and so on.

Manuscript received XX, XXXX, 2023; revised XX, XXXX, XX and accepted XX, XXXX, XX. Date of publication XX, XXXX, XX; date of current version XX, XXXX, XX. This work was supported in part by the Research Foundation – Flanders (FWO) Odysseus 1 under Grant G.0H94.18N; Methusalem Programme of the Ghent University Special Research Fund (BOF) under Grant 01M01021; and in part by the National Science and Technology Major Project, China, under Grant J2019-I-0001-0001 and Grant J2019-I-0019-0018. Michael Ruzhansky was also supported by Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/R003025/2. (Corresponding author: Michael Ruzhansky.)

Junqing Huang and Michael Ruzhansky are with the Department of Mathematics: Analysis, Logic and Discrete Mathematics, Ghent University, 9000 Ghent, Belgium; Michael Ruzhansky is also with the School of Mathematical Sciences, Queen Mary University of London, E1 4NS London, UK (e-mail: {Junqing.Huang, Michael.Ruzhansky}@UGent.be).

Haihui Wang is with the School of Mathematical Sciences, Beihang University (BUAA), China (e-mail: wwhmath@buaa.edu.cn).

Junqing Huang and Haihui Wang contributed equally to this work.

Digital Object Identifier no. XX.XXXX/TIP.XXXX.XXXXXXX.

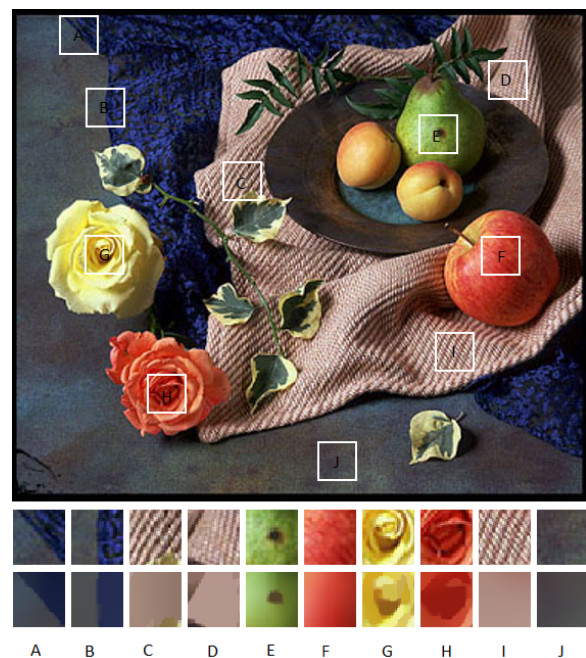


Figure 1. An illustration of image structures and textures with different characteristics in local close-ups (bottom). It is clear that image textures may have various forms with coarse-to-fine oscillating patterns arranged in spatial (non-)uniformly, (ir-)regularly, (an-)isotropically, and so on; while the underlying image structures are in general referred to as strong edges, piecewise constant and/or smoothing surfaces.

Decomposing an image into meaningful structural and textural components is generally a challenging problem due to the complex visual analysis processes. As shown in Fig. 1, visual objects could be exhibited in various and complicated forms — for example, arranged in spatial (non-)uniformly, (ir-)regularly, (an-)isotropically. The structural and textural parts, as illustrated in [8], [9], could be vague under the semantic descriptions, because they are highly dependent on the scale — that is, image structures in one scale can

be regarded as the textures in another scale. As a result, it is not easy to model them based on a simple model. In practice, it is necessary to characterize both structural and textural parts to make the image decomposition problem to be deterministic when modeling them mathematically.

In a general sense, an image is assumed to be composed of two fundamental parts: a structural part — referring to the geometrical clues, and a textural part — representing these coarse-to-fine scale details or noise [3], [4], [10], [11]. It may also suggest that the structural part contains semantic meaning elements such as homogeneous regions, object contours, and sharpening edges, while the textural part could have more abundant and diverse local features, at least the periodic and oscillating information [3], [8]. This assumption has been adopted in many existing methods such as partial differential equation (PDE) diffusion methods [4], [12], and structure-aware filtering methods [6], [13], [14], variational-based methods [8], [10], [15] and higher-order extensions [16], [17], and so on. Despite the great success, there is still considerable interest to exploit powerful and efficient methods to achieve visual-appealing decomposition results both theoretically and practically.

In this paper, we propose a novel semi-sparse model for image structure-texture decomposition, which devotes to extracting piece-wise smoothing structures from complex textural backgrounds. This new model is derived from an optimization framework consisting of L^1 -norm data fidelity and semi-sparse regularization, which takes advantage of two-fold benefits for high-quality decomposition results. On the one hand, it has shown in traditional methods [18]–[20] that L^1 -norm data fidelity helps to capture the oscillatory patterns and is empirically suitable for structure-texture decomposition. On the other hand, it also turns out in cutting-edge filtering methods [21] that semi-sparsity prior knowledge is beneficial for fitting polynomial-smoothing surfaces while preserving these sparse features such as edges and singularities. As illustrated hereafter, such a combination indeed possesses two-fold benefits of L^1 -norm data fidelity and semi-sparse regularization and is empirically suitable for image structure-texture decomposition. The main contribution of the paper is summarized as follows:

- A simple and effective semi-sparsity minimization model is designed for image structure-texture decomposition, which is an optimization-based framework by means of two-fold benefits: L^1 -norm data fidelity and semi-sparse regularization.
- An efficient numerical solution based on a multi-block alternating direction method of multipliers (ADMM) algorithm is introduced to solve the non-convex and non-smooth minimization problem, which gives rise to a powerful tool for image structure-texture decomposition.
- A number of experimental results are also presented and discussed to demonstrate its versatility and many benefits to image structure-texture decomposition, where our semi-sparsity model performs more favorable image decomposition results on natural images against many state-of-the-art methods.

We further point out that the proposed approach has a very simple form and possesses attractive properties in

decoupling main structures from complex textures. Despite the non-convex nature of the optimization problem, the proposed method can be efficiently solved using a multi-block ADMM algorithm and significantly improves the performance of image decomposition, in particular towards natural images. Notice also that the semi-sparsity regularization is also possible to be combined with other functions for a more complex image decomposition — for example, image structure, texture, and noise decomposition. More details will be discussed below.

The rest paper is organized as follows. In Sec. 2, we review the typical methods for modeling the textural and structural parts. In Sec. 3, we derive the proposed model in detail and illustrate the properties of L^1 -norm data fidelity and semi-sparsity regularization. An efficient ADMM numerical solution is then introduced in Sec. 4. The performance of the proposed method is compared in Sec. 5 with a series of experimental results. After that, we give a brief analysis of the possible extensions in Sec. 6. The conclusion and further work are drawn in Sec. 7.

2 STRUCTURE AND TEXTURE ANALYSIS

In the literature, many computational models have been developed to either extract main structures from an image or dedicate an explicit image decomposition. We review some related work regarding structural and textural analysis models for favorable image decomposition results.

2.1 Problem Formulation

We consider an additive form of the image decomposition problem — that is, given an image f , it is possible to form the problem, either explicitly or implicitly, into a general optimization-based framework:

$$\min_{u,v} \mathcal{S}(u) + \mathcal{T}(v), \quad \text{subject to } u + v = f, \quad (1)$$

where u and v are the structural and textural counterparts defined on $\Omega \in \mathbf{R}^2$, typically the same rectangle or a square as that of f . The functional $\mathcal{S}(u)$ and $\mathcal{T}(v)$ are defined in some proper spaces to characterize the properties of v and u , respectively. In general, v and u could be exhibited in varying and complicated forms in natural images. In practice, $\mathcal{S}(u)$ is advocated to capture the well-structural components such as homogeneous regions and salient edges, and $\mathcal{T}(v)$ may be expected to have the capacity of extracting the textural part such as repeating or oscillating patterns or random noise. The generalization of $\mathcal{S}(u)$ and $\mathcal{T}(v)$ contributes to various existing methods in the literature.

2.2 Modeling $\mathcal{S}(u)$ for Structural Analysis

Image “structure” is a vague semantic terminology and hard to give a mathematical definition. Intuitively, the structural part contributes to the majority of the semantic information of an image, which is always assumed to be composed of piece-wise constants or smoothing surfaces with discontinuous or sharpening edges. The structural analysis is devoted to finding suitable functions $\mathcal{S}(u)$ to represent image structures. As shown in Tab. 1, $\mathcal{S}(u)$ is typically formulated as a regularization term in the context of Eq. 1.

Table 1
Methods of modeling structural parts

Methods	Structural functional $\mathcal{S}(u)$	Description and References
Structure-aware filters	$u = Wf$ $\mathcal{S}(u) := g(u, \nabla u) \ \nabla u\ _2^2$	Local filters [13], [22], [23] Global filters [5], [6]
Regularized methods	$\mathcal{S}(u) := \alpha \ \nabla u\ _1$, $\mathcal{S}(u) := \alpha \ \nabla u\ _0$, $\mathcal{S}(u) := \alpha \ \nabla u\ _1 + \beta \ \nabla^2 u\ _1$, $\mathcal{S}(u) := \alpha \ \nabla u - v\ _1 + \beta \ \nabla v\ _1$	TV regularization [3], [10], [15], [19] L_0 gradient regularization [14], [24] TV-TV ² models [25], [26] TGV models [27], [28]
Patch-based Models	$\nabla_w u := (u(y) - u(x)) \sqrt{w(x, y)}$, $\mathcal{S}(u) := \alpha \ \mathbf{W} \circ u\ _1 + \beta \ \mathbf{L} \circ \mathbf{W} \circ u\ _1$	low-rank [29] non-local [30]

The early work for structure analysis can be traced back to nonlinear PDE methods for anisotropic diffusion [3], [4], [12]. The diffusion process helps to preserve main image structures and remove local textural details, leading to structure-preserving smoothing effects. This idea is further explored by structure-aware filters for simplification and ease of implementation. The existing filters are generally divided into local ones: bilateral filter [13], guided filter [22] and many variants, and global ones: weighted least square (WLS) filter [6] and extensions [34], [35], and so on. Many structure-aware filters such as Laplacian filter [36], and rolling guidance filter (RGF) [37] also employ multi-scale analysis techniques to characterize coarse-to-fine structures. Although these filters are not originally designed for the goal of structural analysis, they have been used as practical tools for image decomposition. However, many of them may cause blur effects around strong edges and are not adequate to decouple main structures from strong oscillating textural backgrounds. Recently, some special filters are also designed for image structure extraction such as bilateral texture filter [23], relative total variation (RTV) model [5] and variants [38], [39], and so on.

Another trend for image structural analysis is based on variational theory in functional spaces. For example, the pioneering Rudin–Osher–Fatemi (ROF) model [10] assumes that the structural part u belongs to a bounded variation (BV) space, allowing for piece-wise constant functions to represent image structures. The BV space assumption of natural images has been extensively studied as a powerful mathematical tool in many computational models since the induced total variational (TV) regularization has a famous structure-preserving property. In theory, it is possible for TV-based methods to capture oscillating signals for better decomposition results in accordance with some concisely-designed textural models — as interpreted hereafter, L^1 space [19], weaker G space [3], Hilbert space [15], and so on. As demonstrated therein, these spaces allow to decouple the complicated textures from image structures, nevertheless, the TV-based regularization for image structures has two limitations due to the properties of the BV space: (1) the TV-based regularization enforces piece-wise constant results which may cause staircase artifacts, especially in the region of polynomial-smoothing surfaces, and (2) the non-linear thresholding operator of TV-based models give rise to over-smoothing effects around sharpening edges. The defects are further illustrated in the experimental results.

Table 2
Methods of modeling textural parts

Models	Textural functional $\mathcal{T}(v)$	Description and Reference
$\int u + n = \int$	$\mathcal{T}(v) := \lambda \ v\ _p^p$, $\ v\ _p^p := \begin{cases} \ u - f\ _2^2, & \text{ROF model [10]} \\ \ u - f\ _1, & L^1 \text{ space [19], [31]} \end{cases}$ $\mathcal{T}(v) := \lambda \ v\ _G$, $\ v\ _G := \begin{cases} \inf_g \{\ g\ _{L^\infty} v \in G\}, & G \text{ space [3]} \\ G = \{v v = \text{div } g\}, \end{cases}$ $\mathcal{T}(v) := \lambda \ v\ _{H^{-1}}$, $\ v\ _{H^{-1}} := \ \nabla (\Delta^{-1}) v\ _2^2$, Hilbert space [15]	
$\int u + c + n = \int$	$\lambda \ f - u - v\ _2^2 + \gamma \ v\ _{G_p}$, $\ v\ _{G_p} := \begin{cases} \inf \{\ g\ _{L^p} v \in G_p\}, & G_p \text{ space [3]} \\ G_p = \{v v = \text{div } g\}, \end{cases}$ $\lambda \ f - K(u + v)\ _2^2 + \gamma \ v\ _*$, $\ v\ _* := \begin{cases} v \in W^{s,p}, & \text{Sobolev space [32]} \\ v \in H^{-s}, & \text{Hilbert space [33]} \end{cases}$	

Regarding the staircase artifacts in TV-based models, an effective remedy is based on higher-order methods for image structural analysis [25], [32], [40], [41]. For example, it turns out in the total generalized variation (TGV) [42] that a higher-order gradient regularization is beneficial for signal/image denoising, as it helps reduce staircase effects substantially when preserving the jump and discontinuities (edges). Based on the property of TGV regularization, the TGV- L^1 [27], TGV-OSV [32], TGV-Gabor [17] methods have also been proposed to replace the TV-based ones for image structure analysis. The TV-TV² model is another higher-order modification based on [25], which advocates a similar property as the TGV-based models with resemble effects. Despite the improvements of these higher-order methods in suppressing the notorious staircase artifacts, they still suffer from over-smoothing problems around the sharpening discontinuous edges [28], [40]. More recently, a semi-sparsity model [21] is also proposed based on a higher-order regularization model, which is proven to have powerful simultaneous-fitting abilities in both sharpening edges and polynomial-smoothing surfaces. However, it is not suitable for image structure extraction because of the limitation in large oscillatory textural components.

The efforts of avoiding blurry artifacts around strong edges have also received considerable attention in image structural analysis. The sparsity-inducing L_0 regularization has been demonstrated to be effective in preserving strong edges in structure-aware smoothing methods [14], [24]. The advantage is further demonstrated in a higher-order extension for signal/image filtering [21]. However, they are not suitable for image structure-texture decomposition directly due to the limitation of removing high-contrast oscillating details. Recently, a new L_0 minimization based on the RTV metric [5] has been proposed for image structure retrieval, which turns out to be, in particular, applicable for image structure extraction [38]. Despite the advance of L_0 minimization in capturing sharpening edges, it is worth noting that such a minimization problem is a non-convex and non-smooth optimization problem and the numerical solution is computationally expensive in the case of large-scale scenarios, despite the recent endeavor in developing acceleration techniques. Inspired by the properties of L_0 minimization [5] and higher-order extension [21], we will further discuss the properties of L_0 norm regularization in higher-order gradient domains and demonstrate its advantages for image structural analysis.

2.3 Modeling $\mathcal{T}(v)$ for Textures or Noises

The textural part v , as aforementioned, can be exhibited in diverse and complicated forms in natural images. In general, image texture is recognized as repeated patterns consisting of oscillating details, random noise, and so on. In many existing methods, the textural part v is modeled as the residual of an image f subtracting the structural part u — that is, $v = f - u$. This simplified assumption has been adopted by many existing methods since it provides an easy but acceptable approximation for modeling image textures mathematically. The typical functional $\mathcal{T}(v)$ for textural analysis in the literature are presented in Tab. 2. Accordingly, $\mathcal{T}(v)$ is always treated as a data-fidelity term in the context of Eq. 1. We briefly discuss their advantages and limitations as follows.

In the well-known ROF model [10], v is simply assumed to be in L^2 space with image structures u in BV space. The ROF model takes advantage of BV space to penalize oscillating signals when representing the structural part u with piece-wise smoothing functions. However, it is not suitable for large oscillatory textural modes despite using multi-scale analysis [43]. Along with the ROF model, many image decomposition methods were subsequently proposed for better textural modeling results. A simple but attractive remedy is to model the texture part v in the L^1 space instead of the L_2 , which leads to the well-known TV- L^1 model for image textural analysis [31], [44]. The benefits underpin the appealing properties of L^1 space that help to decouple large-scale oscillating details, outliers and impulse noise, while preserving geometrical features without eroding main structures [18], [19]. Notice that the TV- L^1 model needs to solve a type of non-smooth minimization problem and the numerical solution is time-consuming even with a concisely-designed algorithm [19], [45].

It has also witnessed many efforts to find more suitable functions to model the textural part v in the literature. The TV- G model [3], for example, takes into account a weaker space endowed G -norm to replace $L^p(p \geq 1)$ space for oscillating patterns. The theoretical justification shows that the G space identifies the Banach space containing signals with large oscillations, in particular image textures and noise. However, a direct solution to the TV- G model is not possible due to the unavailability of the corresponding Euler-Lagrange equation [8], [46]. An alternative way is using $G_p(1 \leq p \leq \infty)$ to approximate the problem [8]. The experimental results also verify its better property for capturing the oscillating patterns [47]. Based on the G -norm space assumption, the model has been further explored and implemented along with different structural analysis models, for example, higher-order TGV model [32], edge-driven high-order BV model [48] and adaptive directional variant [49]. Despite the advantages, they may partially decouple image structures into textures and cause blurry effects around strong or sharpening edges.

Similar to the G -norm assumption, the oscillating patterns in the Osher-Sole-Vese (OSV) model are measured in H^{-1} space given by the second-order derivative of functions in a homogeneous Sobolev space [15]. The numerical solution of the OSV model turns out to be easy to implement based on the corresponding Euler-Lagrange

equation. In a case of more generalized theoretical analysis [33], it is shown that the component v can be modeled in the Hilbert-Sobolev space H^{-s} . Along with the strategy of finding admissible textural spaces, there also exist many other models for texture analysis — for example, the TV-div(BMO) [50] and TV-Besov [51], which are also combined with different regularization terms for better results. In the practical implementations, the use of different spaces for textural analysis has shown limited improvements despite the different properties of these spaces. The phenomenon may be largely due to the complex textural patterns of images, in particular for natural images.

2.4 Extensive Analysis

Although we discuss a range of models for structure and texture analysis, there exist many other methods that can be also understood under the interpretation of Eq. 1 explicitly or implicitly. The patch-based models [29], [30], [52], for example, characterize a low-rank nature of image structures or textures over non-local image patches. It is possible for them to produce better decomposition results but the performance is at the expense of searching local similar patches. Besides, the frequency filters, transformed domain analysis [53], and sparse representation techniques [54] are also proposed for structural and textural analysis, while the performance is generally limited in some typical cases and not universally applicable in many practical applications. In addition, it has also been witnessed that some existing methods [55] strive to decompose an image into more than two basic components — for example, structures, textures, and random noise. Despite their different forms, they share a similar assumption of image structures and textures as listed in Tab. 1 and 2. More detailed discussions of them are out of the scope here. The reader is also referred to the related research points for more theoretical analysis and their internal connections. It is worth noting that it is still a fertile research point to explore more advanced models for image structure and texture analysis.

3 PROPOSED METHOD

In this section, we further analyze the characteristics of image structures and textures and propose a generalized framework for image structure-texture decomposition. The model is based on a semi-sparsity-inducing regularization for structural analysis. We interpret that such a model is preferable to produce better image decomposition results in accordance with different textural analysis models.

3.1 Observations and Motivations

We have discussed several established image structure and texture analysis methods and highlighted their strengths and limitations for image decomposition applications. To further elucidate their properties, we first take into account an example in 2, where a natural image is given with the structural component formed by singularities, sharpening edges, polynomial-smoothed surfaces, and the textural component composed of various characteristics spanning from textural patterns, scales, and directions to amplitudes.

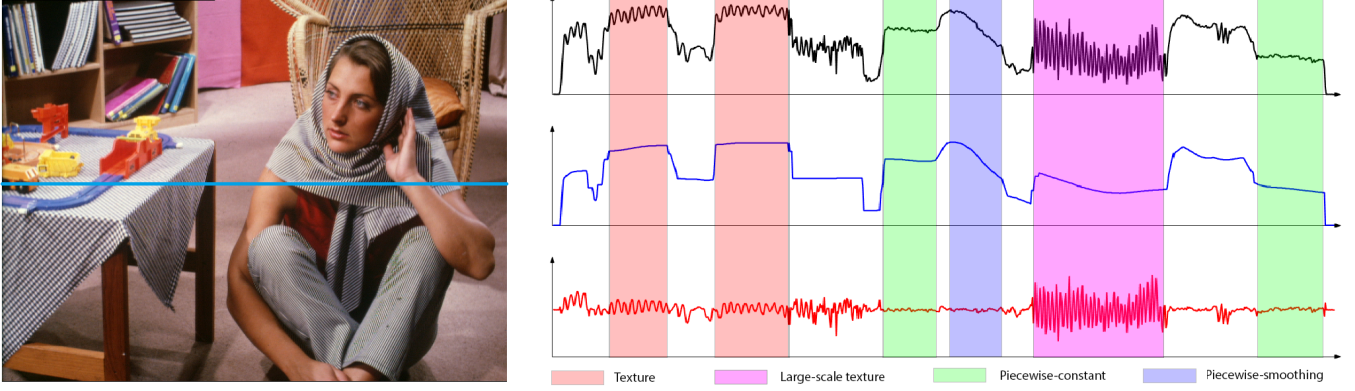


Figure 2. An illustration of the characteristics of natural image structures and textures. The input image signal (Black) along a row is plotted with the decomposed structures (Blue) and textures (Red) based on the proposed method. It is clear that the structural part consists of some strong edges, piece-wise constant and polynomial smoothing surfaces, while the textural part contains different levels of oscillation or repeated details.

It is always expected for an image decomposition to ideally distinguish the structural and textural parts. But, it is a challenging problem due to the facts: (1) the sharpening edges (singularities) and oscillation patterns are highly intertwined for their high-frequency nature, but the former always belongs to image structures, and the latter is mostly attributed to image textures, (2) the piece-wise constant and polynomial smoothing surfaces may coexist in image structures, but they have very different properties and it is difficult for a simple model to capture their characteristics simultaneously, and (3) the textural part may be exhibited in various forms which are not adequate for a simple model to identify them mathematically. In summary, a plausible image decomposition method should be able to distinguish the three-fold characteristics of signals:

- **Piece-wise constant or smoothing surfaces:** Many natural or synthesized images have visual objects with homogeneous regions consisting of piece-wise constant and smoothing surfaces simultaneously. It is necessary for image decomposition methods to have the capacity of representing and extracting them without introducing artifacts such as staircase results in the polynomial-smoothing surfaces.
- **Sharpening edges/singularities:** The discontinuous boundaries between homogeneous objects give rise to sharpening edges or singularities of an image, which is important for visual understanding and analysis. An effective image decomposition model should be able to precisely preserve these features without introducing blur artifacts, and distinguish these sharpening edges and singularities from large oscillating textures¹.
- **Coarse-to-fine oscillatory details:** The characteristics of oscillatory details can be complicated because of the diverse patterns, multiple directions, varying amplitudes, and coarse-to-fine oscillation scales. It is crucial for an image decomposition method to capture these various characteristics and distinguish them from sharpening edges and singularities.

1. The assignment of singularities may depend on the appropriate semantic scale of the interested visual information.

As verified in Fig. 2, the key point of image structure-texture decomposition is to distinguish the three types of signals while maintaining a fine balance. The dilemma also exists in many existing methods in Tab. 1 and Tab. 2. Structure-aware filters, for example, are capable of keeping main structures but they do not well distinguish sharpening edges and large oscillating textures. The TV-based regularization is proposed to identify the structural part with piece-wise constant and sharpening edges, while it may result in stair-case artifacts in polynomial-smoothing surfaces [3], [8], [10]. Higher-order TV extensions [25], [32] alleviate the stair-case artifacts, but they cannot get rid of blurry edges. The textural analysis models perform better textural results with appropriate structural models but they may still have problems in decoupling the coarse-to-fine textures. The L^1 [31], [44], G -norm [3], and H^{-1} [15] spaces help to characterize large-scale oscillating textures, while the performance is somehow still limited in complicated textural patterns and they may still cause side-effects such as over-smoothing results around strong edges.

3.2 (Semi)-Sparsity Inducing Regularization

“Sparsity” prior knowledge has been extensively studied in many signal and image processing fields, for example, sparse image recovering [58] sparse representation [59], compressed sensing [60], and so on. It has shown that L_0 norm regularization can be used as a favorable and powerful mathematical tool to identify sparsity-induced priors because of the capacity in capturing the minority of key features (archetypes) of signals. In many linear system backgrounds, “sparsity” priors can be posed into a L_0 quasi-norm regularized optimization model with the form,

$$\min_u \Phi(u) + \alpha \|Hu\|_0, \quad (2)$$

where $\Phi(u)$ is defined in a proper function space modeling the forward process of a physical system. The second term provides a “sparsity” constraint under a linear operator H with a positive weight α . The notation $\|\cdot\|_0$ is the so-called L_0 quasi-norm denoting the number of non-zero entries of a vector, which provides a simple and easily-grasped measurement of sparsity.

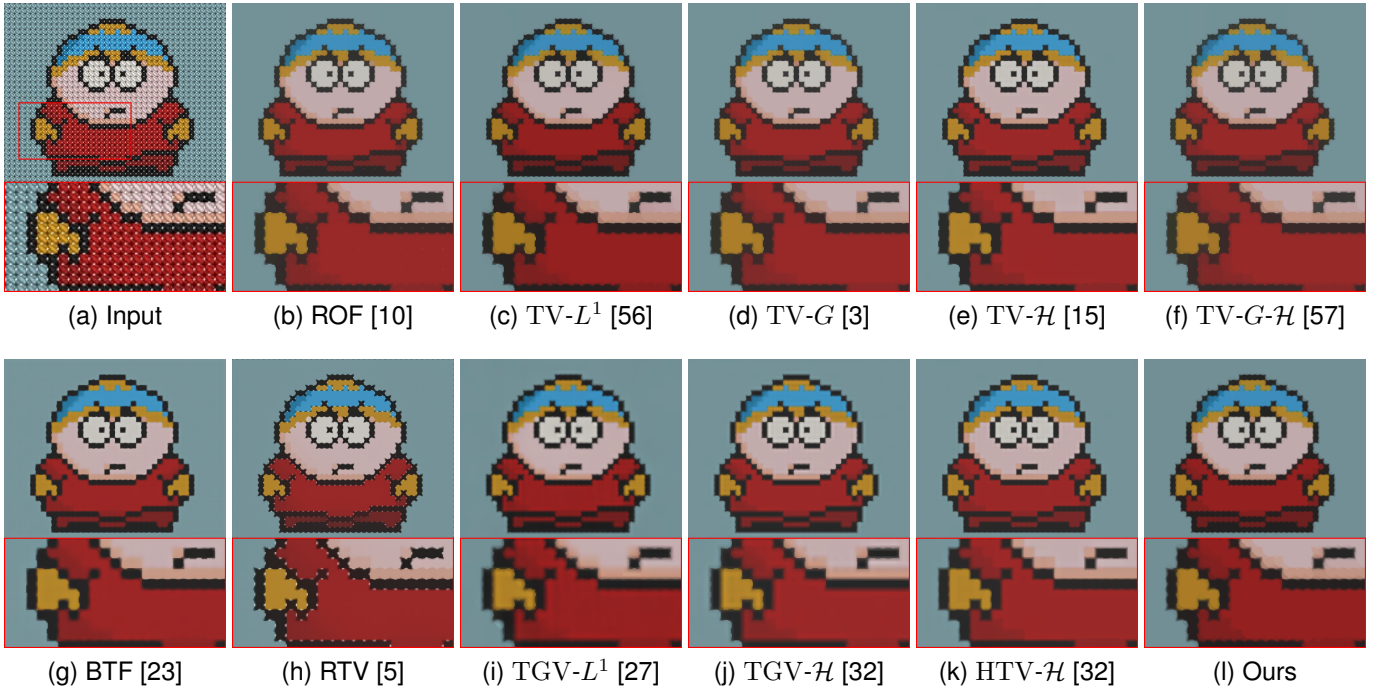


Figure 3. Visual results of image structures in case of **piece-wise constant structures** and **large oscillating textures**. (a) Input, (b) ROF ($\lambda = 0.005, \alpha = 0.001$) [10], (c) TV- L^1 ($\lambda = 0.003, \alpha = 0.0045$) [56], (d) TV- G ($\lambda = 0.003, \alpha = 0.0005, \gamma = 0.0002$) [3], (e) TV- H ($\lambda = 0.002, \alpha = 0.005$) [15], (f) TV- $G-H$ ($\lambda = 0.004, \alpha = 0.001, \gamma = 0.002$) [57], (g) BTF ($\sigma = 5.0, \text{iter} = 4$) [23], (h) RTV ($\lambda = 0.01, \sigma = 3.0$) [5], (i) TGV- L^1 ($\lambda = 0.0007, \alpha = 0.0008, \beta = 0.0008$) [27], (j) TGV- H ($\lambda = 0.004, \alpha = 0.01, \beta = 0.03$) [32], (k) HTV- H ($\lambda = 0.003, \alpha = 0.006, \beta = 0.0015$) [32], and (l) Ours ($\lambda = 0.005, \alpha = 0.006, \beta = 0.001$). For fairness, all methods are fine-tuned for a similar level of smoothness. $SSR(C_0/C_1)$, (b)~(j): 17.24 (0.1138/0.0194), 17.16 (0.0301/0.0340), 17.15 (0.1035/0.0075), 17.18 (0.0364/0.0316), 17.11 (0.1298/0.0059), 17.05 (0.0369/0.0727), 17.17 (0.0427/0.1089), 17.04 (0.0645/0.0609), 17.06 (0.0717/0.0627), 17.07 (0.0560/0.0829), 17.12 (0.0203/0.0022). (Zoom in for better view.)

Many sparsity-inducing models can be understood in the context of Eq. 2. The inverse imaging reconstruction, for example, takes into account $\Phi(u) = \|Au - f\|_2^2$ where A is observation matrix and $H = \nabla$ is the gradient operator because the L_0 norm gradient regularization help to restore piece-wise constant image surfaces. This idea has also drawn great attention in some filtering techniques [14], [24] because it helps to preserve the minority of singularities and discontinuous sharpening features of signals. Recently, such an idea is also found in image structure restoration [38], where a sparsity-induced regularization is introduced for image structure analysis.

Despite the progress in preserving discontinuous and sharp edges, the occurrence of notorious stair-case artifacts is not negligible in existing sparsity-inducing models. To address this contradiction, there has been considerable attention towards using regularization techniques in higher-order gradient domains, in particular focusing on TGV-based methods [32], [42] and TV-TV²-based regularization [25], [32]. More recently, a so-called semi-sparsity model has been also explored for smoothing filters [21], [42], in which the established L_0 regularization in higher-order gradient domains allows simultaneously fitting the regions coexisting sharp edges and polynomial-smoothing surfaces. Motivated by the edge-preserving property of L_0 regularization [14], [24] and higher-order methods [21], [42] in avoiding stair-case artifacts, we combine them in a unified framework and illustrate that they are also universally applicable for image structural and textural analysis.

Formally, we recall the semi-sparse notations and show how to identify the semi-sparse property in the higher-order gradient domains effectively. Without loss of the generality [21], we say u is a semi-sparse signal if its higher-order gradients satisfy the following relationship,

$$\begin{cases} \|\nabla^{n-1}u\|_0 > N, \\ \|\nabla^n u\|_0 < M, \end{cases} \quad (3)$$

where ∇^n is the n -th (partial) differential operator, and M, N are some appropriate natural number satisfying $N \gg M$. The merit of Eq. 3 is easy to understand, that is, the non-zeros entries of $\nabla^n u$, measured by L_0 quasi-norm, is much smaller than that of $\nabla^{n-1}u$, which occurs if and only if $\nabla^n u$ is much more sparse than $\nabla^{n-1}u$. Taking into account a n degree piece-wise polynomial function for example, it is easy to verify that the n -th order gradient $\nabla^n u$ is sparse, while it not holds for the k -th ($k < n$) order gradient $\nabla^k u$. It is also clear that a polynomial-smoothing signal with degree $d < n$ satisfies Eq. 3 essentially.

3.3 Generalized Semi-Sparse Decomposition

It has shown in the work [21] that natural images always exhibit notable semi-sparsity properties when attributing the characteristics of higher-order gradient distributions. Notice that image structures could characterize more obvious semi-sparsity properties, as they can be effectively represented as a combination of a series of piece-wise constant regions, polynomial-smoothing surfaces, and sharpening discontinuous edges. This observation motivates us

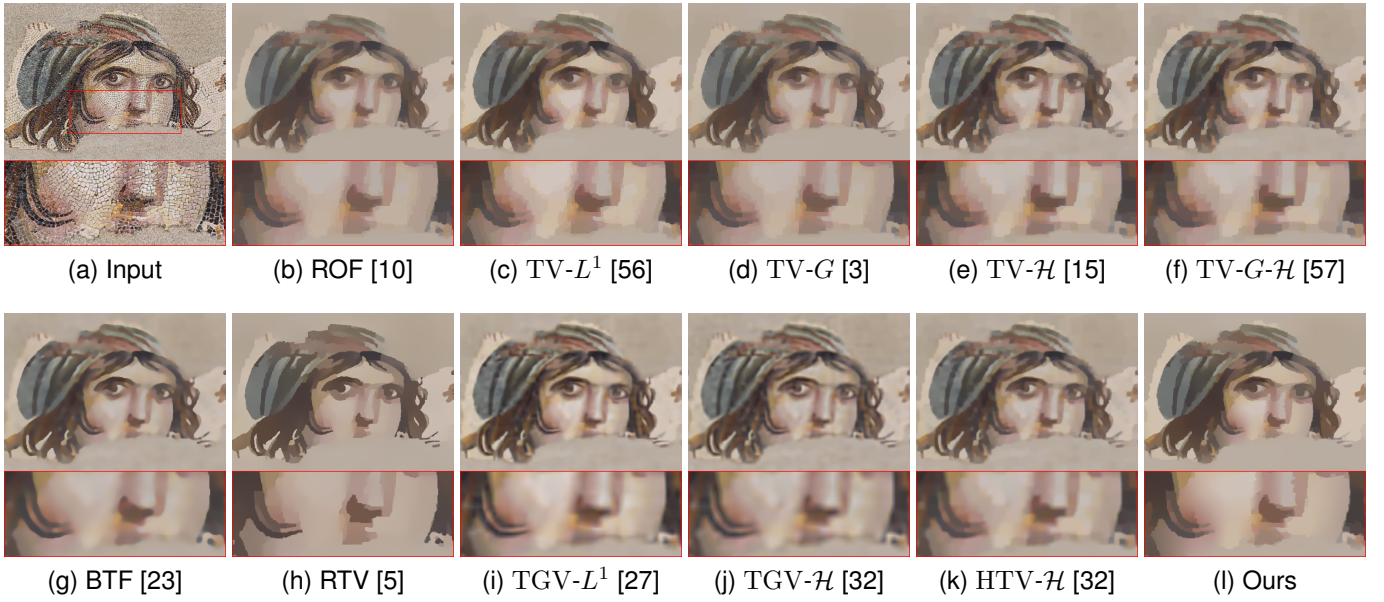


Figure 4. Visual results of image structures in case of **piece-wise constant/smoothing structures and large oscillating textures**. (a) Input, (b) ROF ($\lambda = 0.0035$, $\alpha = 0.001$) [10], (c) $TV-L^1$ ($\lambda = 0.001$, $\alpha = 0.002$) [56], (d) $TV-G$ ($\lambda = 0.003$, $\alpha = 0.0007$, $\gamma = 0.0002$) [3], (e) $TV-H$ ($\lambda = 0.002$, $\alpha = 0.003$) [15], (f) $TV-G-H$ ($\lambda = 0.004$, $\alpha = 0.008$, $\gamma = 0.008$) [57], (g) BTF ($\sigma = 5.0$, $iters = 7$) [23], (h) RTV ($\lambda = 0.015$, $\sigma = 4.0$) [5], (i) $TGV-L^1$ ($\lambda = 0.001$, $\alpha = 0.002$, $\beta = 0.002$) [27], (j) $TGV-H$ ($\lambda = 0.002$, $\alpha = 0.05$, $\beta = 0.04$) [32], (k) $HTV-H$ ($\lambda = 0.002$, $\alpha = 0.0045$, $\beta = 0.004$) [32], and (l) Ours ($\lambda = 0.004$, $\alpha = 0.007$, $\beta = 0.002$). Quantitative results with $STR(C_0/C_1)$ metrics, (b)~(j): 18.88(0.1220/0.0158), 18.89(0.0365/0.0204), 18.90(0.1101/0.0259), 18.89(0.0486/0.0429), 18.84(0.0444/0.0500), 19.01(0.0373/0.0834), 18.97(0.0539/0.0195), 18.67(0.1003/0.1742), 18.86(0.0320/0.2002), 18.87(0.0431/0.1421), 18.95(0.0292/0.0188). (Zoom in for better view.)

to use semi-sparsity priors to capture structural features and avoid blurry and stair-case artifacts. Specifically, we propose a general semi-sparsity framework based on the L_0 regularization in higher-order gradient domains,

$$\min_{u,v} \mathcal{T}(v) + \alpha_k \sum_{k=1}^{n-1} \|\nabla^k u\|_1 + \alpha_n \|\nabla^n u\|_0, \quad s.t. \quad u + v = f, \quad (4)$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ are the positive weights. The semi-sparse regularization part advocates that the k -th ($k < n$) order gradients $\nabla^k u$ have a small measure in L^1 space, and the highest-order gradient $\nabla^n u$ tends to be fully sparse in views of the polynomial-smoothing property of image structures. Differing from the semi-sparsity smoothing [21], a new semi-sparse regularization is introduced here². As we illustrate hereafter, such a semi-sparse regularization combined with an appropriate structural analysis model $\mathcal{T}(v)$ enjoys better fitting performance in preserving polynomial-smoothing surfaces and sharp edges.

In principle, $\mathcal{T}(v)$ in Eq. 4 can be arbitrarily chosen from Tab. 2. However, the textural part v , as illustrated in Fig. 2, varies from image to image with diverse and multiple forms spanning from textural patterns, scales, directions to amplitudes, and so on. It is usually difficult to discriminate and describe them independently in a simple model. In this paper, we suggest using the L^1 -norm to model image textures due to its simplicity and effectiveness. Firstly, it has been demonstrated in several $TV-L^1$ models [19], [31] that L^1 space allows representing image textures with a variety of patterns, including coarse-to-fine scales, anisotropic

directions, varying oscillating amplitudes frequencies, and so on. Secondly, the contrast of image structures is also important information that needs to be preserved during the decomposition process, however, it is not adequate for many decomposition methods such as the L_2 -norm model. Instead, $\mathcal{T}(v)$ in L^1 space has been proven to have the capacity of preserving the contrast information even in large oscillating textural cases. Moreover, the L^1 data fidelity provides a fine balance between effectiveness and efficiency due to the simple form and numerical implementation, and affordable computational cost. As demonstrated in experimental results, such a semi-sparse regularization combined with L^1 data-fidelity enjoys better fitting performance in polynomial-smoothing surfaces and sharpening edges.

Furthermore, the choice of n is also an important factor for the image composition model of Eq. 4, which is in nature determined by the property of image structures. For natural images, it has been shown in [21] that $n = 2$ is enough to give plausible filtering results in practice. Recalling the similar properties of image structures and the observations in Sec. 3.1, we here choose $n = 2$ as the highest order for semi-sparsity regularization, reducing Eq. 4 into the form,

$$\min_u \lambda \|u - f\|_1 + \alpha \|\nabla u\|_1 + \beta \|\nabla^2 u\|_0, \quad (5)$$

where, λ, α, β are positive parameters. Notice that λ is not necessarily varying in theory, for example, fixing $\lambda = 1$, and it is introduced here to rescale Eq. 5 into an appropriate scale for a more accurate numerical solution. Despite the simple form, it enjoys some startling properties for image structure-texture decomposition. We additionally discuss the advantages and benefits of Eq. 5 compared with many existing methods.

2. In the semi-sparsity smoothing model [21] the k -th ($k < n$) order gradients $\nabla^k u$ is assumed to be close to that of image f in L^2 space.

On the one hand, it has been shown that high-quality image decomposition results are not easy to guarantee in many existing methods, including structure-aware filters [13], [14], TV-based regularization [10] and higher-order extensions [27], [32] since they either give over-smoothing results in spikes and sharpening edges — taking total variation (TV) method [10] for example, and/or introduce stair-case artifacts in polynomial smoothing surfaces such as the famous L_0 -norm gradient regularization [14]. On the other hand, several textural analysis models [3], [15], [31] faithfully decouple the large oscillating patterns, while they still suffer from over-smoothing problems, especially around sharpening and discontinuous edges. In a nutshell, many existing methods have limitations in recovering image structures that coexist with sparse features and polynomial-smoothing surfaces, especially in large oscillating textural situations. It is actually difficult for them to weigh a balance to avoid two aspects of deficiencies simultaneously. In contrast, our semi-sparsity model as demonstrated in the below experiments retains edge-preserving properties in strong edges and spikes, and also produces more appropriate results in polynomial-smoothing surfaces, giving a so-called simultaneous-fitting ability in both sharpening and smoothing regions. Moreover, the textural analysis in L^1 space also enables us to process image decomposition in large oscillating textures. We conclude that it provides a way to deal with both cases in high-level fidelity, which makes it essentially different from the traditional filtering methods.

3.4 Differing from Existing Higher-order Methods

There are two existing works highly related to the proposed semi-sparsity regularization model, in particular, the total generalization variation (TGV) [42] and high-order total variations [25], [26], as well as many of their extensions [27], [32], [40]. It is necessary to discuss them and show their differences both theoretically and experimentally.

Mathematically, the TGV-based model [42] suggested the following regularization,

$$\text{TGV}(u) = \min_u \alpha_1 \|\nabla u - w\|_1 + \alpha_2 \|\xi(w)\|_1,$$

where $\alpha_1, \alpha_2 \in \mathbb{R}^+$ are weights, and $\xi(w) = \frac{1}{2} (\nabla w + \nabla w^T)$ denotes the distributional symmetrized derivative. Such a regularization is always combined with different textural analysis functions [27], [32] to reduce the staircase artifacts in many total variational (TV) methods.

The TGV-based method is essentially a second-order regularization model by introducing an auxiliary variable w . Notice that $\nabla u \approx w$ when $\alpha_1 \rightarrow \infty$, while the exact equality is not attainable in practice, thus there exists a discrepancy between ∇u and w , in particular around sharpening edges. In this case, w can be viewed as a smoothing approximation of ∇u . Notice also that the second term $\alpha_2 \|\xi(w)\|_1$ also has a smoothing role when imposing a regularization for w , which may lead to over-smoothing results in sharpening edges of u . The defect may be alleviated by increasing α_1 and reducing α_2 , but it is not easy to find a balance in case of varying scales of oscillating textures. The intuitive conclusion is also observed in our visual comparisons. As illustrated therein, it is usually difficult to find a group of α and β to preserve the strong edges precisely.

The TV-TV² regularization is another higher-order method that has drawn attention in image restoration [25], [26], which is also proposed as a higher-order extension of the TV regularizer to overcome the staircase artifacts. By definition, the TV-TV² regularization has the form,

$$\text{TV-TV}^2(u) = \min_u \alpha_1 \|\nabla u\|_1 + \alpha_2 \|\nabla^2 u\|_1,$$

The first term $\alpha_1 \|\nabla u\|_1$ is a TV-based regularization that encourages piece-wise constant u , and the second term $\alpha_2 \|\nabla^2 u\|_1$ plays a similar role but acts on the second-order gradient ∇u . The TV-TV² regularization helps suppress staircase artifacts since ∇u is imposed to be piece-wise constant. However, it is also known that the TV regularization would cause blur effects around sharpening edges, and the second-order regularization would further strengthen the smoothing effect. In summary, the TV-TV² regularization indeed helps reduce the stair-case artifacts but the blur results in strong edges are also not avoidable, especially in cases of large oscillating textures or strong noise. Essentially, it has a similar property as the TGV regularization.

The semi-sparsity regularization inherits the advantages of higher-order methods in terms of mitigating stair-case artifacts. In parallel, the inclusion of an L_0 -norm constraint in higher-order gradients also enables the restoration of sharpening edges without compromising the fitting ability of higher-order models in polynomial smoothing surfaces. The analysis of image textures in L^1 space further enhances the capacity to decompose large oscillating patterns. In summary, the proposed method combines the advantages of both TGV and TV-TV², and manages to overcome their weakness, allowing image decomposition for a broad spectrum of natural images.

4 EFFICIENT MULTI-BLOCK ADMM SOLVER

The semi-sparsity image decomposition model in Eq. 5 poses challenges in direct solution due to the non-smooth L^1 -norm and non-convex nature of L_0 -norm regularization. It has proved that such a minimization problem can be approximately solved based on several efficient numerical algorithms such as iterative thresholding algorithm [61], half-quadratic (HQ) splitting technique [62], alternating direction method of multipliers (ADMM) [24], [63], and so on.

We here employ a multi-block ADMM algorithm to solve the proposed model because it is applicable for a type of non-convex minimization problem [63], [64]. Moreover, it is easy to implement and has a low computational cost even for large-scale problems. For completeness, we briefly introduce the multi-block ADMM algorithm with the M separable objective functions and N constraints,

$$\begin{aligned} & \min_{\{x_i\}} f_1(x_1) + f_2(x_2) + \cdots + f_M(x_M) \\ \text{s.t. } & A_{1,1}x_1 + A_{1,2}x_{1,2} + \cdots + A_{1,M}x_M = b_1 \\ & \vdots \\ & A_{N,1}x_1 + A_{N,2}x_{N,2} + \cdots + A_{N,M}x_M = b_N, \end{aligned} \quad (6)$$

where $x_i \in \mathcal{X}_i \subset \mathbb{R}^{n_i}$ are closed convex sets, the coefficient matrix $A_{n,i} \in \mathbb{R}^{p \times n_i}$, $b_k \in \mathbb{R}^p$ and $f_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^p$ are some proper functions.

The multi-block ADMM algorithm is a general case of the ADMM algorithm. Accordingly, we rewrite the Eq. 6 in its augmented Lagrangian form,

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}) := \sum_{i=1}^M f_i(x_i) - \sum_{n=1}^N \left\langle \mathbf{z}_n, \sum_{i=1}^M A_{n,i} x_i - \mathbf{b}_j \right\rangle + \sum_{n=1}^N \left(\frac{\rho_n}{2} \left\| \sum_{i=1}^M A_{n,i} x_i - \mathbf{b}_j \right\|^2 \right), \quad (7)$$

where $\mathbf{z} = \{z_1, \dots, z_N\}$ is the Lagrange multiplier and $\rho = \{\rho_1, \rho_2, \dots, \rho_N; \rho_i > 0\}$ is a penalty parameter. The augmented Lagrangian problem can be solved by iteratively updating the following sub-problems:

$$\begin{cases} x_1^{k+1} := \operatorname{argmin}_{x_1} \mathcal{L}_\rho(x_1, x_2^k, \dots, x_M^k; z^k) \\ \vdots \\ x_M^{k+1} := \operatorname{argmin}_{x_M} \mathcal{L}_\rho(x_1^{k+1}, x_2^{k+1}, \dots, x_{M-1}^{k+1}, x_M; z^k) \\ z_n^{k+1} := z_n^k - \rho_n \left(\sum_{i=1}^N A_{n,i} x_i^{k+1} - \mathbf{b}_j \right) \end{cases} \quad (8)$$

Now, it is easy to reformulate the proposed semi-sparsity image decomposition model of Eq. 5 into a three-block ADMM model. Specifically, we consider the discrete case of Eq. 5 and formulate it as a constrained optimization problem by introducing the variables $\mathbf{h}$, $\mathbf{g}$ and $\mathbf{w}$,

$$\begin{aligned} \min_{\{\mathbf{u}, \mathbf{h}, \mathbf{g}, \mathbf{w}\}} \quad & \lambda \|\mathbf{u} - \mathbf{f}\|_1 + \alpha \|\mathbf{D}\mathbf{u}\|_1 + \beta \|\mathbf{L}\mathbf{u}\|_0, \\ \text{s.t.} \quad & \mathbf{u} - \mathbf{f} = \mathbf{h}, \quad \mathbf{D}\mathbf{u} = \mathbf{g}, \quad \mathbf{L}\mathbf{u} = \mathbf{w}, \end{aligned} \quad (9)$$

where $\mathbf{D} = \{\mathbf{D}_x; \mathbf{D}_y\}$ and $\mathbf{L} = \{\mathbf{L}_{xx}; \mathbf{L}_{xy}; \mathbf{L}_{yx}; \mathbf{L}_{yy}\}$ are the discrete differential operators in the first-order and second-order cases along the x - and y - directions, respectively. The augmented Lagrangian form of the three-block ADMM model of Eq. 9 has the form,

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{u}, \mathbf{h}, \mathbf{g}, \mathbf{w}; \mathbf{z}) := & \lambda \|\mathbf{h}\|_1 + \alpha \|\mathbf{g}\|_1 + \beta \|\mathbf{w}\|_0 \\ & - \langle \mathbf{z}_1, \mathbf{u} - \mathbf{f} - \mathbf{h} \rangle + \frac{\rho_1}{2} \|\mathbf{u} - \mathbf{f} - \mathbf{h}\|_2^2 \\ & - \langle \mathbf{z}_2, \mathbf{D}\mathbf{u} - \mathbf{g} \rangle + \frac{\rho_2}{2} \|\mathbf{D}\mathbf{u} - \mathbf{g}\|_2^2 \\ & - \langle \mathbf{z}_3, \mathbf{L}\mathbf{u} - \mathbf{w} \rangle + \frac{\rho_3}{2} \|\mathbf{L}\mathbf{u} - \mathbf{w}\|_2^2, \end{aligned} \quad (10)$$

where $\mathbf{z} = \{z_1, z_2, z_3\}$ and $\rho = \{\rho_1, \rho_2, \rho_3\}$ are the dual variables and penalty parameters adapted from Eq. 7. Clearly, the problem Eq. 10 can be solved based on the iterative procedure presented in Eq. 8. The benefit of Eq. 9 is that each sub-problem can be efficiently solved even for a large-scale problem. We briefly explain the sub-problems and show their property for efficient solutions.

Sub-problem $\mathbf{u}$: The augmented Lagrangian objective function Eq. 10 reduces to a quadratic function with respect to $\mathbf{u}$ when fixing the other variables, which gives rise to an equivalent minimization problem,

$$\begin{aligned} \min_{\mathbf{u}} \quad & \rho_1 \left\| \mathbf{u} - \mathbf{f} - \mathbf{h} - \frac{\mathbf{z}_1}{\rho_1} \right\|_2^2 + \rho_2 \left\| \mathbf{D}\mathbf{u} - \mathbf{g} - \frac{\mathbf{z}_2}{\rho_2} \right\|_2^2 \\ & + \rho_3 \left\| \mathbf{L}\mathbf{u} - \mathbf{w} - \frac{\mathbf{z}_3}{\rho_3} \right\|_2^2 + c, \end{aligned} \quad (11)$$

where c is a constant not related to $\mathbf{u}$. The optimal solution of 10 is then attained by solving the following linear system:

$$\begin{aligned} (\rho_1 \mathbf{I} + \rho_2 \mathbf{D}^T \mathbf{D} + \rho_3 \mathbf{L}^T \mathbf{L}) \mathbf{u} = & \rho_1 (\mathbf{f} + \mathbf{s}) + \rho_2 \mathbf{D}^T \mathbf{g} + \rho_3 \mathbf{L}^T \mathbf{w} \\ & + \mathbf{z}_1 + \mathbf{D}^T \mathbf{z}_2 + \mathbf{L}^T \mathbf{z}_3, \end{aligned} \quad (12)$$

where $\mathbf{I}$, $\mathbf{D}^T$ and $\mathbf{L}^T$ are identity matrix and the transpose of $\mathbf{D}$ and $\mathbf{L}$. Due to the symmetric and positive coefficient matrix, Eq. 12 can be directly solved using several linear solvers, for example, the Gauss-Seidel method and preconditioned conjugate gradients (PCG) method. However, these methods may be still time-consuming when the problem has a large number of variables in our cases. Notice that the differential operators $\mathbf{D}$ and $\mathbf{L}$ are linear invariant operators, it is possible to use fast Fourier transforms (FFTs) to solve Eq. 12 for further acceleration. It is possible to reduce the computational cost significantly, in particular encountering a large-scale problem such as millions of variables.

Sub-problems $\mathbf{h}$ and $\mathbf{g}$: By analogy, the sub-problems with respect to $\mathbf{h}$ and $\mathbf{g}$ have the forms

$$\min_{\mathbf{h}} \lambda \|\mathbf{h}\|_1 - \langle \mathbf{z}_1, \mathbf{u} - \mathbf{f} - \mathbf{h} \rangle + \frac{\rho_1}{2} \|\mathbf{u} - \mathbf{f} - \mathbf{h}\|_2^2 \quad (13)$$

$$\min_{\mathbf{g}} \alpha \|\mathbf{g}\|_1 - \langle \mathbf{z}_2, \mathbf{D}\mathbf{u} - \mathbf{g} \rangle + \frac{\rho_2}{2} \|\mathbf{D}\mathbf{u} - \mathbf{g}\|_2^2 \quad (14)$$

It is known that the optimization problem of Eq. 13 and 14 is a typical L^1 -norm regularization minimization that can be efficiently solved for the separable property. Specifically, we reduce them into a one-dimensional minimization problem and estimate each variable $\mathbf{h}_i$ and $\mathbf{g}_i$ individually, giving the closed-form solutions,

$$\begin{aligned} \mathbf{h} &= \mathcal{S}(\mathbf{u} - \mathbf{f} + \frac{\mathbf{z}_1}{\rho_1}, \frac{\lambda}{\rho_1}), \\ \mathbf{g} &= \mathcal{S}(\mathbf{D}\mathbf{u} + \frac{\mathbf{z}_2}{\rho_2}, \frac{\alpha}{\rho_2}), \end{aligned} \quad (15)$$

with the soft shrinkage operator defined as:

$$\mathcal{S}(x_i, \tau) = \operatorname{sgn}(x_i) \cdot (|x_i| - \tau)_+$$

where sgn is a sign function, and $(a)_+$ is defined as 0 if $a < 0$ and a otherwise.

Sub-problems $\mathbf{w}$: Similarity, the sub-problem with respect to $\mathbf{w}$ has the form,

$$\min_{\mathbf{w}} \beta \|\mathbf{w}\|_0 - \langle \mathbf{z}_3, \mathbf{L}\mathbf{u} - \mathbf{w} \rangle + \frac{\rho_3}{2} \|\mathbf{L}\mathbf{u} - \mathbf{w}\|_2^2 \quad (16)$$

which is a L_0 -norm minimization problem and has the same separable property as L^1 -norm minimization with each individual variable given by the closed-form solution

$$\mathbf{w} = \mathcal{H}(\mathbf{L}\mathbf{u} + \frac{\mathbf{z}_3}{\rho_3}, \frac{\beta}{\rho_3}) \quad (17)$$

where $\mathcal{H}(x, \tau)$ is the hard-shrinkage operator defined as:

$$\mathcal{H}(x_i, \tau) = \begin{cases} 0, & |x_i| < \tau, \\ x_i, & \text{otherwise.} \end{cases}$$

The dual variables $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$ are updated as follows:

$$\begin{aligned} \mathbf{z}_1 &= \mathbf{z}_1 + \rho_1 (\mathbf{u} - \mathbf{f} - \mathbf{h}), \\ \mathbf{z}_2 &= \mathbf{z}_2 + \rho_2 (\mathbf{D}\mathbf{u} - \mathbf{g}), \\ \mathbf{z}_3 &= \mathbf{z}_3 + \rho_3 (\mathbf{L}\mathbf{u} - \mathbf{w}). \end{aligned} \quad (18)$$

Algorithm 1 ADMM Solver for Semi-sparse Model

- 1: **Input:** signal/image $\mathbf{f}$, parameters λ, α, β , and weights ρ_1, ρ_2, ρ_3 ;
- 2: **Initialization:** $\mathbf{u}^0 \leftarrow \mathbf{0}$; $\mathbf{h}^0, \mathbf{g}^0, \mathbf{w}^0 \leftarrow \mathbf{0}$; $\mathbf{z}_1^0, \mathbf{z}_2^0, \mathbf{z}_3^0 \leftarrow \mathbf{0}$, and $k \leftarrow 0$;
- 3: **while** not $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2^2 / \|\mathbf{u}^k\|_2^2 \leq \varepsilon$ **do**
- 4: Solve Eq. 12 for $\mathbf{u}^{k+1}$ with the FFT acceleration;
- 5: Solve Eq. 13 for $\mathbf{h}^{k+1}$ with soft-shrinkage operator;
- 6: Solve Eq. 14 for $\mathbf{g}^{k+1}$ with soft-shrinkage operator;
- 7: Solve Eq. 16 for $\mathbf{w}^{k+1}$ with hard-shrinkage operator;
- 8: Update Eq. 18 for Lagrange multipliers $\mathbf{z}_i^{k+1}$:

$$\mathbf{z}_1^{k+1} = \mathbf{z}_1^k + \rho_1(\mathbf{u}^k - \mathbf{f} - \mathbf{h}^k),$$

$$\mathbf{z}_2^{k+1} = \mathbf{z}_2^k + \rho_2(\mathbf{D}\mathbf{u}^k - \mathbf{g}^k),$$

$$\mathbf{z}_3^{k+1} = \mathbf{z}_3^k + \rho_3(\mathbf{L}\mathbf{u}^k - \mathbf{w}^k);$$
- 9: Increment $k : k = k + 1$,
- 10: **end while**
- 11: **Output:** the structural part $\mathbf{u}$.

The three-block ADMM algorithm alternatively solves the sub-problems and updates the Lagrange multipliers until the given stop criteria are met, which leads to an iterative procedure for the proposed image decomposition model. Since all sub-problems have closed-form solutions in low computational complexity, the challenging non-convex L_0 -norm regularized problem is empirically solvable even under a large-scale of variables. The numerical results will further demonstrate the effectiveness and efficiency of the multi-block ADMM algorithm.

The convergent analysis of multi-block ADMM solver can be directly built from the general ADMM case [64]. Notice that Eq. 4 and Eq. 5 have the non-convex L_0 -norm regularization, it is not easy to construct a convergence analysis. Recently, it has been shown in the study [65] that the ADMM algorithm is also applicable for a wide range of non-smooth and non-convex problems. For example, the convergent analysis can be built when one or more of f_i in Eq. 6 is lower semi-continuous with the remainders are the Lipschitz differentiable functions, which is our case and claims the convergence of **Algorithm 1**, as the first and second terms in Eq. 9 are convex L^1 -norm with Lipschitz continuous property, and the non-convex L_0 norm with has lower semi-continuous property. Actually, it is also easy to see that each sub-problem convergent to its criteria point, which makes it possible to find a Cauchy sequence of the solution $\{\mathbf{u}^k\}_{k=0}^\infty$. Here, the sub-problem of Eq. 16 is L_0 -regularization and the iterative hard-thresholding operator is convergence under the Restricted Isometry Property (RIP) condition [61]. The reader may refer to [63], [65], [66] for more convergence details.

5 EXPERIMENTAL RESULTS

In this section, we further demonstrate our semi-sparse model and its high-quality decomposition performance in scenarios of images composed of different structures and textures. We start the discussion with the interpretations of a new dataset, parameter configurations, compared methods, and so on. A series of numerical results are also presented and compared with the state-of-the-art methods both qualitatively and quantitatively.

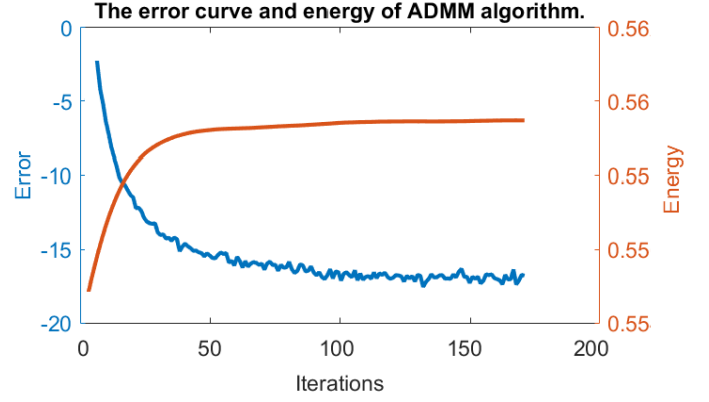


Figure 5. The (log) **error curve** and **energy** of $\mathbf{u}^k$ during the ADMM iterative procedure.

5.1 Dataset and Compared Methods

For the sake of comprehensive analysis, we first collect a new dataset for image structure and texture decomposition. This dataset has 32 images chosen from the community of image decomposition [5], [23], [67]. As shown in the later experiments, the dataset consists of a set of images that contain various structural patterns such as sharpening edges, piece-wise constant and smoothing surfaces, as well as diverse textural patterns with large oscillating values, multiple scales and directions, and so on. The inclusion of such images allows us to thoroughly assess the performance of our method across different types of visual content.

To ensure a systematic comparison, we elaborate on our semi-sparsity model with its superior performance against a series of image decomposition methods. Specifically, we compare the model with the well-known ROF model [10], TV- L^1 [56], TV- G [3], TV- $\mathcal{H}$ [15], TV- $G\text{-}\mathcal{H}$ [57], BTF [38], RTV [5], TGV- L^1 [27], TGV- $\mathcal{H}$ [32], HTV- $\mathcal{H}$ [32], which cover most of the structural and textual models listed in Tab. 1 and Tab. 2. As illustrated hereafter, we will consider three typical scenarios consisting of different image structures and textures: 1) piece-wise constant structures with uniform but large oscillating textures, 2) piece-wise smoothing structures with multi-scale oscillating textures, and 3) a mixed case with complex structural and textural patterns.

5.2 Parameters and Configurations

Recalling the model in Eq. 4 and 5, λ , α , and β control the amount of smoothness penalty applied to output image structures. As aforementioned, λ is introduced to rescale Eq. 5 into an appropriate scale, depending on the precision of discrete images and the threshold operators in Eq. 15 and 17. For 8-bit color images³, we empirically found that $\lambda \in [0.0001, 1.0]$, $\alpha \in [0.0001, 0.1]$ and $\beta \in [0.0001, 0.1]$ are always suitable for a high-accuracy numerical solution when taking into account weights $\rho_i = 1 (i = 1, 2, 3)$ without any specification. The parameters may vary from image to image depending on the density and oscillating level of image textures. The above configurations provide a good trade-off between accuracy and computational efficiency.

3. All images are normalized into the range $[0, 1]$ in our experiments.

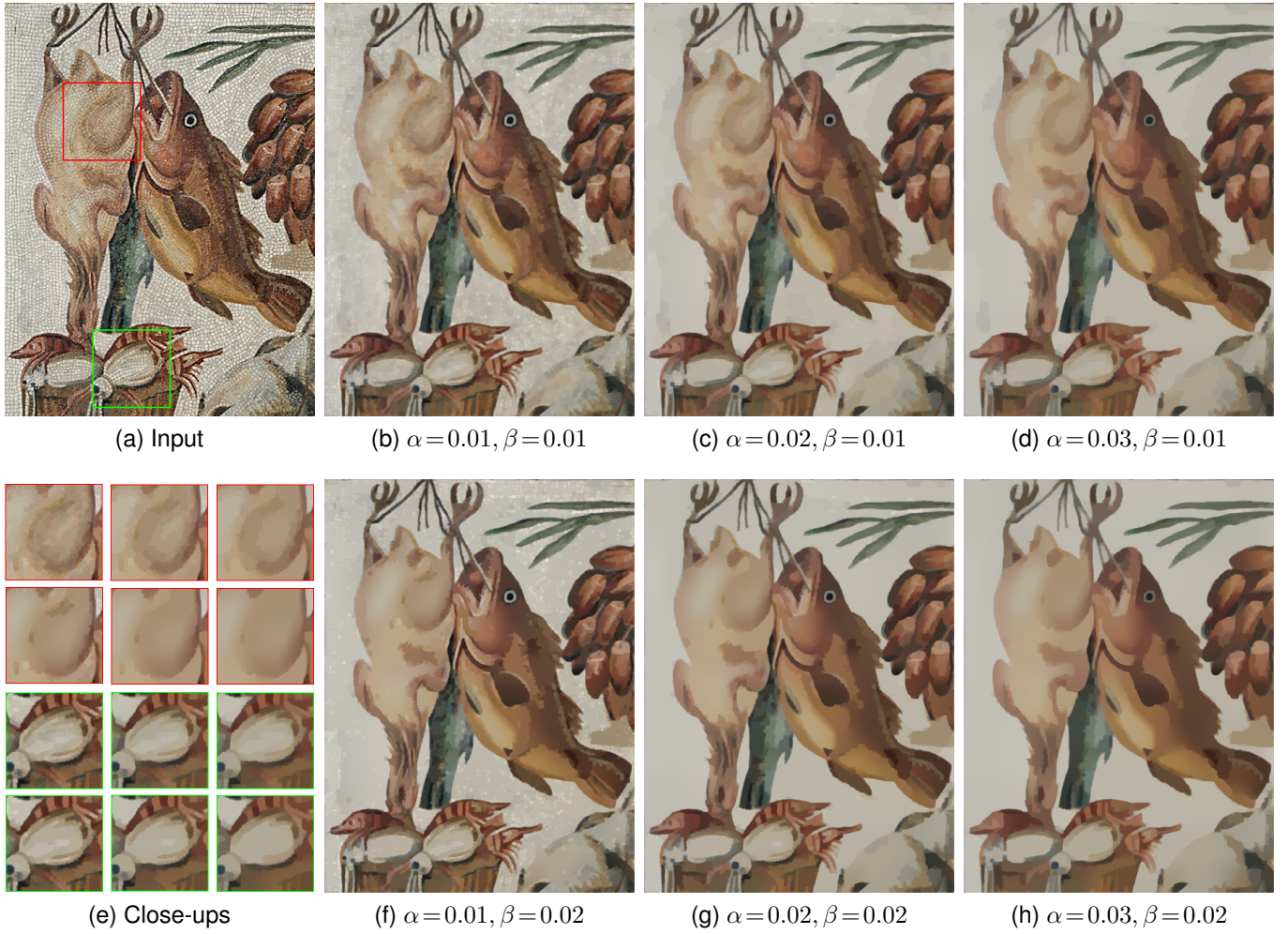


Figure 6. Visual comparison of image structures with different parameters. $\lambda = 0.01$ is fixed here and α is gradually increased in each row to give more and more smoothing results, while a similar smoothing trend is also observed with increasing β in each column. Notice that α mainly controls the global smoothness of output structures, and β tends to smooth the local features such as staircase results in polynomial smoothing surfaces.

For clarity, we show the role of λ , α and β in determining the characteristics of output image structures. As illustrated in Fig. 6, we fix $\lambda = 0.01$ and compare the output structures with varying α and β . The output image structures in each row tend to be more and more smoothing when increasing α , but the staircase artifacts can not be avoided effectively in some poly-nominal smoothing regions, as α mainly controls the first-order gradient to be piece-wise constant. In contrast, the staircase artifacts can be alleviated when increasing β in each column. This conclusion also provides the guidance for choosing λ , α , and β in practice — that is, α is first adjusted to produce acceptable decomposition results and then fine-tuning β to reduce the remaining staircase artifacts to reach better local results.

To demonstrate the efficiency of ADMM iterative procedure, we show its convergence based on the relative error $Q_r^k = \|u^{k+1} - u^k\|_2^2 / \|u^*\|_2^2$. Notice that an “exact” solution u^* is not available here, we instead take an extremely tight stop tolerance, for example, $\varepsilon = 10^{-16}$, and treat the output as u^* for evaluation. Afterward, the algorithm is rerun under the same configuration with a smaller stop criterion $\varepsilon = 1.0 \times 10^{-12}$ in **Algorithm 1**. We also take into

account the energy $E_{u^k} = \|u^k\|_2^2$ to show the decomposition performance. The error curve Q_r and energy E_u are plotted in Fig. 5. As we can see, both of them converge quickly to the stable solution and it roughly needs 20 ~ 50 iterations to produce high-quality image decomposition results with appropriate parameters.

Benefiting from the multi-block ADMM procedure, the proposed model can be solved efficiently. As shown in **Algorithm 1**, the process has three main phases: linear system solver, soft/hard-threshold shrinking operators, and the update of Lagrangian multiplier in each iteration. The computation is dominated by the linear system solver, while it can be also efficiently computed with the FFT acceleration because both the FFT and its inverse transform have the computational cost of $O(N \log(N))$, where N is the number of pixels of an image. The threshold shrinking operator can be computed independently for their separable property, and the Lagrange multiplier can be also updated in-place efficiently. As a result, the number of iterations controls the total time of the proposed method. The quantitative analysis of computational cost is also compared with the other methods in the next section.

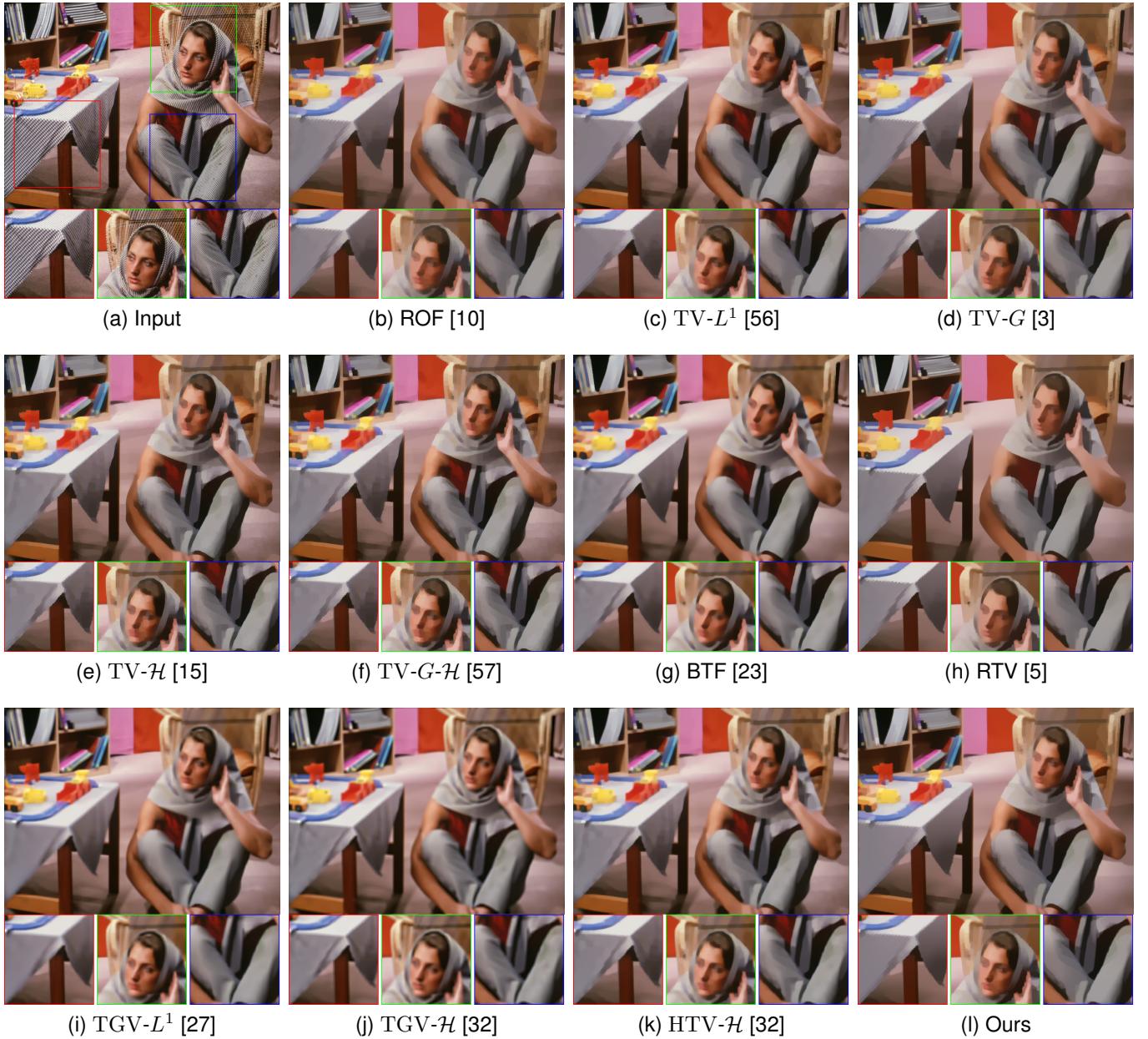


Figure 7. Visual results of image structures in case of **hybrid and complicated textural and structural components**. (a) Input, (b) ROF ($\lambda = 0.0045, \alpha = 0.001$) [10], (c) $TV-L^1$ ($\lambda = 0.001, \alpha = 0.002$) [56], (d) $TV-G$ ($\lambda = 0.003, \alpha = 0.0006, \gamma = 0.0004$) [3], (e) $TV-H$ ($\lambda = 0.001, \alpha = 0.01$) [15], (f) $TV-G-H$ ($\lambda = 0.004, \alpha = 0.008, \gamma = 0.005$) [57], (g) BTF ($\sigma = 7.0, iter = 3$) [23], (h) RTV ($\lambda = 0.012, \sigma = 3.0$) [5], (i) $TGV-L^1$ ($\lambda = 0.001, \alpha = 0.002, \beta = 0.002$) [27], (j) $TGV-H$ ($\lambda = 0.002, \alpha = 0.045, \beta = 0.025$) [32], (k) $HTV-H$ ($\lambda = 0.004, \alpha = 0.006, \beta = 0.0015$) [32], and (l) Ours ($\lambda = 0.002, \alpha = 0.005, \beta = 0.001$). Quantitative results with the $STR(C_0/C_1)$ metrics, (b)~(j): 22.84 (0.1908/0.0918), 22.83 (0.0550/0.1161), 22.89 (0.1771/0.0954), 22.78 (0.0392/0.0424), 22.81 (0.0416/0.0326), 22.88 (0.0271/0.0526), 22.85 (0.1371/0.0447), 22.84 (0.0077/0.1087), 22.78 (0.0370/0.1994), 22.85 (0.0507/0.0869), 23.01 (0.0774/0.0644). (Zoom in for better view.)

5.3 Structure Extraction Performance

To demonstrate the benefits of the proposed method, we conduct a comparative analysis of the decomposed results. Firstly, we introduce image structure-to-texture ratio (STR) as an objective index to evaluate the smoothness of output structures, which is defined in decibels (dB) as:

$$STR = 10 \log_{10} \frac{\|u\|_2^2}{\|v\|_2^2},$$

where u and v represent the decomposed image structures and textures, respectively. The STR index has the same

definition as signal-to-noise (SNR) in our case when treating $v = f - u$ as noise. For fairness, all compared methods are either configured with a greedy strategy to produce visual-friendly results or fine-tuned by hand to reach a similar smoothing level of image structures evaluated by the STR index. In what follows, we show the structural decomposition results and compare them in different scenarios of image structures and textures.

In Fig. 3, we show the decomposition structures for the first type of image. In this situation, it is usually difficult for traditional structure-aware filtering methods [13], [14] to

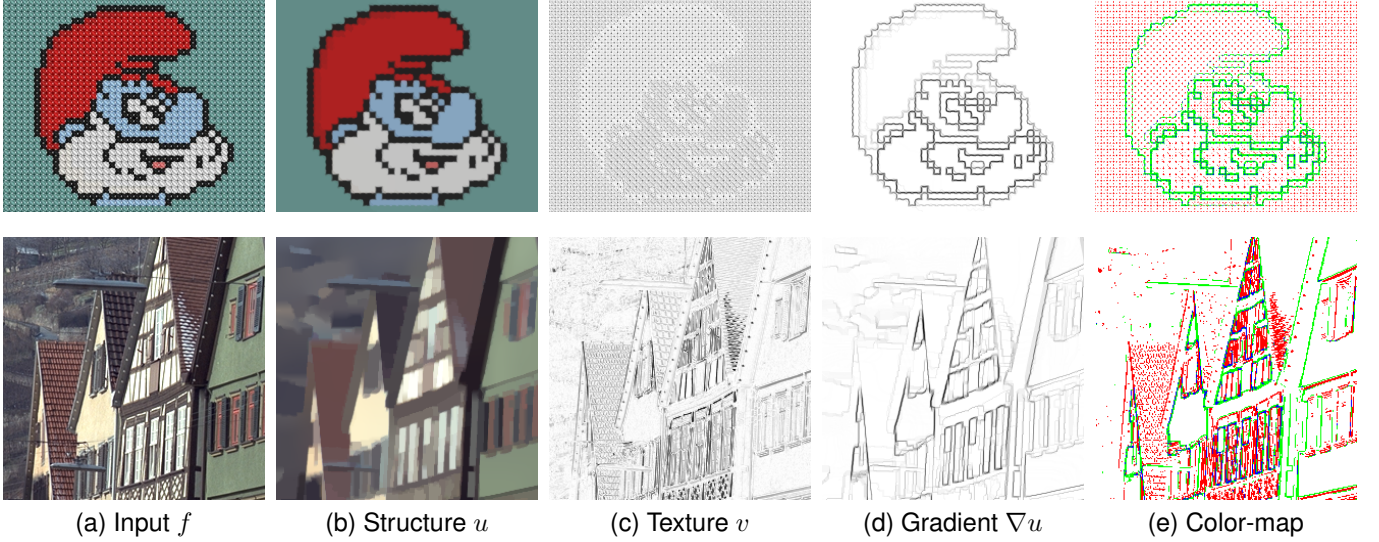


Figure 8. Visual illustrations of the correlation between image structures u (gradient ∇u) and textures v . Given (a) input images f , our method produce (b) structures u and (c) textures $v = f - u$ and (d) gradients ∇u , which are also plotted in (e) in **Red** and **Green** to indicate their tiny overlap regions in **Blue**. The correlation coefficients $C_0 = 0.001$ and $C_1 = 0.001$. (c)~(e) are scaled for better view.

give high-quality results, because of the presence of large oscillating textures. The TV-based models such as TV- G [3], TV-L1 [45] and TV- $\mathcal{H}$ [15], in contrast, demonstrate their reasonable performance in achieving fine-balanced decomposition results. The benefits mainly arise from the capacity of TV space in capturing piece-wise constant structures in spite of large oscillating textures. The special-designed BTF method [23] and RTV model [5] are known for their ability to persevere strong edges when removing large oscillating textures. Despite much better performance, they may introduce slightly blurry effects around strong edges, in particular, the BTF method for the nature of using local spatial weights to suppress periodic textures. Notice also in Fig. 3 that the RTV algorithm may encounter difficulties when removing (white) spikes next to strong edges. The higher-order TGV- $\mathcal{H}$ and HTV- $\mathcal{H}$ models [32] exhibit effective removal of large oscillating textures but suffer from noticeable degradation in sharpening edges. This partly arises from the fact that the TGV regularization tends to penalize the second-order gradients more than the first-order case, leading to a loss of edge fidelity. In contrast, our method receives comparable results in removing large-scale oscillating textures with the best performance in views of edge fidelity.

We continuously analyze the second type of images by considering piece-wise constant and smoothing surfaces in structures, and multi-scale oscillating patterns in textures. As shown in Fig. 4, the TV-based ROF model [10], TV- G [3] and TV-L1 [45] show the ability to decouple the large oscillating textures in piece-wise constant backgrounds but they suffer from different levels staircase effects in the polynomial-smoothing (face) surfaces. In this scenario, the BTF method gives a similar result as the TV-based methods with weak stair-case artifacts, but the over-blurring effect still exists in sharpening edges. The RTV model [5] performs much better results than other existing methods, achieving minimal staircase effects in the face regions while preserving strong edges effectively. The higher-order TGV- $\mathcal{H}$ [32] and HTV- $\mathcal{H}$ [32] partially suppress the staircase effects but give

rise to serious blur degradation around strong edges, which is similar to the case in Fig. 3. In contrast, our semi-sparsity method demonstrates significantly improved performance as the RTV model, receiving the desired smoothing results in the face regions and comparable sharpening edges.

The advantage of our semi-sparsity method is further demonstrated with a more challenging case in Fig. 7. The test image here exhibits highly complex structural and textural patterns, in particular, the oscillating textures are anisotropic in directions, and have coarse-to-fine scales, and varying amplifications, which poses a significant challenge for many image decomposition models. In this case, existing methods either reveal over-smoothing results in textureless regions or can not remove large oscillating textures, especially in view of the sharpness of discontinuous boundaries among the piece-wise constant/smoothing surfaces. Instead, the proposed semi-sparsity decomposition method gives more preferable results in this complex situation with more clean and smooth structural backgrounds and precise sharpening edges. In summary, the proposed method is applicable to a wide range of natural images and can successfully preserve sharp edges while effectively eliminating the undesired oscillating textures. The advance is mainly beneficial from the leveraging of higher-order L_0 regularization for image structures and the measurement of image textures in L^1 space.

5.4 Quantitative Evaluation

We further take a quantitative evaluation to compare the proposed semi-sparsity model with the aforementioned traditional methods. Notice that an objective evaluation is always not difficult for image decomposition methods because the ground truth of each component (structure or textures) is not available in practice.

As suggested in the work [32], [47], the structural and textural components show very little correlation for many texture-rich images. Based on this idea, the correlation coefficient of image structures and textures is then employed

Table 3

The correlation coefficient of STR metrics C_0/C_1 between image structure $u/\nabla u$ (gradient) and image texture $v = f - u$ on the new dataset, and the computation time (s) compared with the state-of-the-art methods.

Metrics STR(19.23)	ROF [10]	TV- L^1 [56]	TV- G [3]	TV- H^{-1} [15]	TV- $G-H$ [57]	BTF [38]	RTV [5]	TGV- L^1 [27]	TGV- H [32]	HTV- H [32]	Ours
$C_0(\downarrow)$	0.1250	0.0464	0.0849	0.0462	0.0774	0.1438	0.0381	0.0850	0.0666	0.0742	0.0361
$C_1(\downarrow)$	0.0528	<u>0.0393</u>	0.0400	0.0467	0.0779	0.1018	<u>0.0442</u>	0.0987	0.0740	0.1041	0.0239
Time (s)	5.91	6.31	11.46	10.82	16.25	<u>3.61</u>	2.05	18.20	22.61	15.52	12.94

to evaluate the quality of decomposed results. We here verify their relationship in Fig. 8 (a)~(c). As we can see, the preferred image structures and textures indeed reveal very different characteristics in local regions. The relationship is more clear when comparing the textures v with the gradient map of structures ∇u as indicated in Fig. 8 (c)~(d). Specifically, we show their overlap regions in Fig. 8 (e) in **Blue**, where the former textures v with the gradient map of structures ∇u are shown in **Red** and **Green**. The conclusion can be also understood from the characteristics of image structures and textures discussed in Sec. 3.1. Accordingly, the strong edges and oscillating patterns are attributed to different parts — the former belongs to image structures and the latter occurs frequently in image textures. As a result, image structures or the corresponding gradient — referred as a representation of strong edges, has little correlation with image textures, since they have tiny overlap in spatial.

The observation also motivates us to use the correlation coefficient for quantitative evaluation. We here define the correlation coefficient $C(x, y) = \frac{cov(x, y)}{\sqrt{var(x)var(y)}}$ for variables x and y , where $cov(\cdot)$ and $var(\cdot)$ refer to the covariance and variance of the counterparts, respectively. Specifically, we denote $C_0(u, v)$ as the correlation coefficient of image structures u and textures v , and $C_1(\nabla u, v)$ as that of the gradient of image structures ∇u and textures v . Both of them are adopted to indicate the correlation between image structures and textures. For color images, each metric is computed and averaged by channel. In general, C_1 is smaller than C_0 due to the less correlation of the gradient ∇u and structure v . We here evaluate our semi-sparsity model against the compared methods on the new dataset. For fairness, all parameters in each method are fine-tuned by hand to give similar smoothing structures, quantified by the STR index. The correlation coefficients C_0 and C_1 are listed in Tab. 3, which are evaluated for each method under the average structure-to-texture ratio $STR = 19.23$ for 32 images. As we can see, the proposed semi-sparse method achieves the best results in both C_0 and C_1 cases.

In addition, we also compare the running time in Tab. 3 to show the efficiency of each method. To evaluate the performance of each method thoroughly, the BTF [38] and RTV [5] methods are adopted from the official implementations, and the TV- L^1 [56] and TGV- L^1 [27] algorithms are carried out based on the primal-dual algorithm [16], [45]. The numerical solutions for other methods are based on a similar ADMM procedure in **Algorithm 1** with 100 iterations. Notice that the BTF [38] has the computational cost in several seconds depending on the filter kernel size. The RTV [5] is much faster for small-size images, while the



Figure 9. Visual comparison of image structural results with L_0 -norm regularization on the (a) second-order (STR = 18.94), and third-order (STR = 18.82) gradient domains, respectively.

time rises greatly due to the high computational cost of linear solver in large-scale cases. The primal-dual algorithm and ADMM procedure have a similar level of computation in each iteration since they are all determined by the FFT implementation in linear system solvers. The statistical results Tab. 3 are estimated by processing a 512×512 resolution color image. All the methods are implemented in Matlab 2015b without any optimization on a desktop PC with Intel Core i7-9800X CPU 3.80GHz and 64G RAM.

6 EXTENSIONS AND ANALYSIS

There are some obvious extensions for our semi-sparsity image decomposition model. Here, we briefly discuss two potential directions: applying the higher-order (order $n \geq 3$) L_0 regularization for semi-sparsity priors, and replacing the textural analysis model in other spaces.

6.1 Higher-order Regularization

In the previous, we have discussed the semi-sparsity model with L_0 -norm regularization in the second-order ($n = 2$) gradient-domain in Eq. 4 and 5, while it is straightforward to apply it in the higher-order ($n \geq 3$) cases. In general, the choice n is determined by the characteristics of image signals. As illustrated in semi-sparsity smoothing filter [21], it is usually enough to produce acceptable results by setting $n = 2$ for natural images, although a better result could be attained when the regularization is imposed on a higher-order gradient domain. The conclusion is derived from the statistic distribution of higher-order gradients of natural images. In the context of image structure and texture

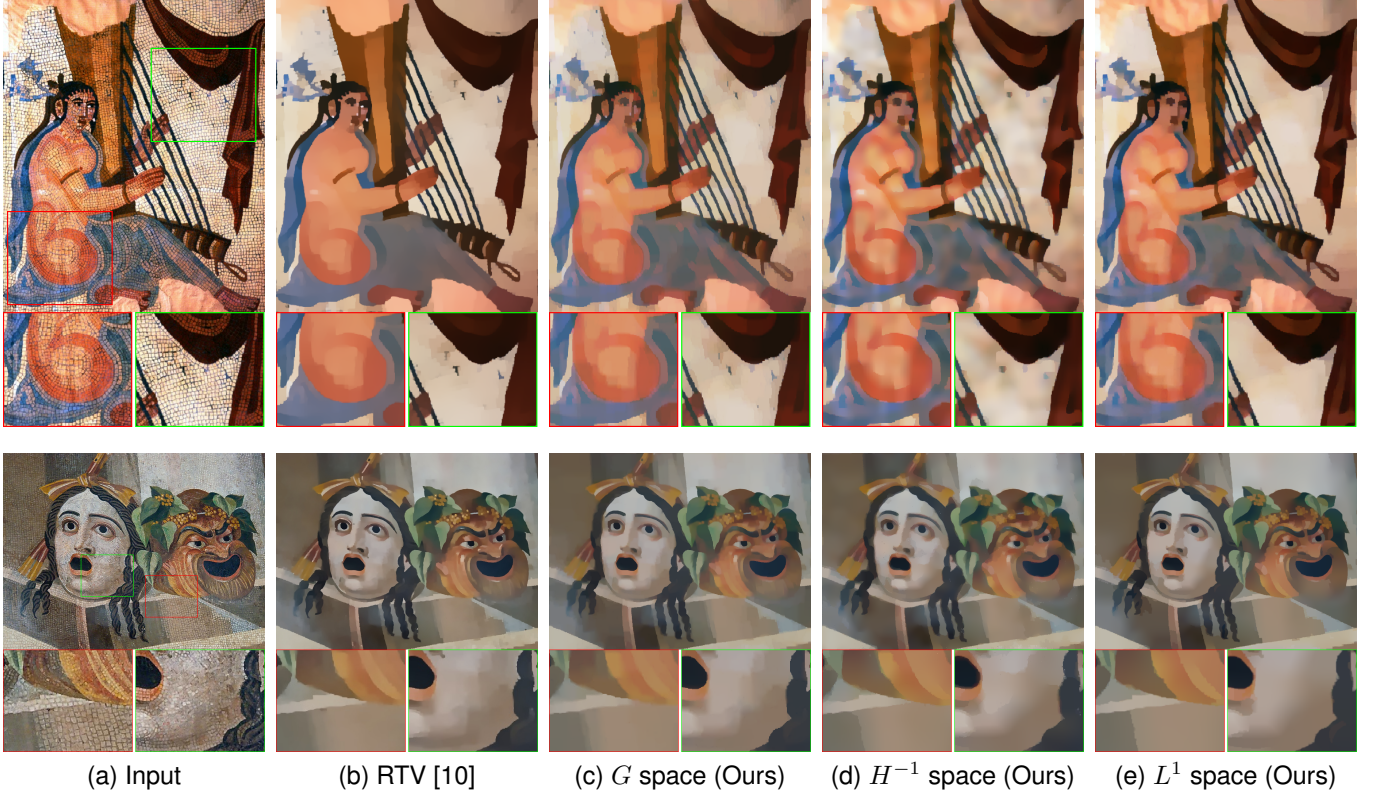


Figure 10. Visual comparison of structural decomposition results based on our semi-sparse model with the textural models in different spaces. (a) Input, (b) RTV [5], (c) G space, (d) H^{-1} space, and (e) L^1 space. The structures are compared under STR metrics: 19.18 (top) and 22.10 (bottom), respectively.

decomposition, we claim that this conclusion is still valid for the similar propriety of image structures. Actually, the structural component of an image tends to more sparse in comparison of images when the textural part is decoupled or removed from images.

We have shown the results in different scenarios of image structures and textures by using second-order L_0 -norm regularization. For complementary, we compare the results by imposing the L_0 constraint on the third-order gradient domain. As shown in Fig. 9, the given image tends to have piece-wise constant and smoothing structures. In both cases, image textures are obviously removed despite that the tiny difference in the piece-wise smoothing surfaces. This is also indicated that it is usually precise enough to use second-order L_0 -norm regularization. This conclusion is also demonstrated in other higher-order methods such as the well-known TGV-based methods [42] and TV-TV² regularized models [25], [26]. Another consideration of choosing the 2nd order gradient for regularization is to reduce the computation cost in many practical applications.

6.2 Textural Analysis in Different Spaces

In this paper, we have demonstrated the effectiveness of the higher-order L_0 -norm regularization with the textural analysis in L^1 space. Apparently, it is possible to combine such a regularization with different textural models.

As shown in Tab. 2, we have discussed several textural models that can be integrated with our structural analysis. The weak space G [3], for instance, allows capturing the

oscillatory components such as regular textures and random noise. The original space G is equipped with the norm:

$$\|v\|_G = \inf_{g=(g_1, g_2)} \left\{ \left\| \sqrt{g_1^2 + g_2^2} \right\|_{L^\infty} \mid v = \text{div } g \right\},$$

where div is the divergence operator. However, the direct numerical solution is unavailable for the involved L^∞ space. The problem is then relaxed, for example, by replacing G space with $G_p = W^{-1,p}$ ($1 \leq p < \infty$). By analogy, we combine the semi-sparse regularization with the G_p textural analysis, giving a new semi-sparsity decomposition model,

$$\min_{u,g} \lambda \|u + \text{div } g - f\|_2^2 + \gamma \|g\|_p^p + \alpha \|\nabla u\|_1 + \beta \|\nabla^2 u\|_0 \quad (19)$$

where $u = \text{div } g$ with g in L^p space, and λ, γ are positive weights. The difference of Eq. 19 from the original OSV model [8] is the utilization of semi-sparsity regularization for structural analysis.

Another well-known structural analysis space is based on the different specializations of Hilbert space $\mathcal{H}$. As stated in [47], there are several models can be deduced from $\mathcal{H}$ space, including ROF model ($\mathcal{H} = L^2$) [10], TV- H^{-1} ($\mathcal{H} = H^{-1}$) [15] and TV-Gabor method [17]. By analogy, it is straightforward to combine our semi-sparsity structural regularization with them for preferable results. We introduce a simple modification of the TV- H^{-1} model [15],

$$\min_{u,g} \lambda \|f - u\|_{H^{-1}}^2 + \alpha \|\nabla u\|_1 + \beta \|\nabla^2 u\|_0 \quad (20)$$

where $\|f - u\|_{H^{-1}}^2 = \int |\nabla (\Delta^{-1})(f - u)|^2 dx dy$ for 2D cases. In the TV- H^{-1} model, the oscillatory texture is modeled

Table 4

The quantitative comparison of the STR metrics C_0/C_1 and running time for the textural models in different spaces.

Metrics STR(19.23)	RTV [5]	G space	H^{-1} space	L^1 space
$C_0(\downarrow)$	0.1444	0.1287	0.2145	0.0874
$C_1(\downarrow)$	0.1174	0.0966	0.1383	0.0826
Time (s)	2.11	29.52	26.83	25.92

as the second derivative of a function in a homogeneous Sobolev space, but the underlying smooth structure may be not handled well in view of the property of TV regularization, while our semi-sparsity extension in higher-order gradient domains partially avoids this problem.

For a supplement, we compare the modified models of Eq. 19 and 20 with the original one in L^1 space. The numerical solutions are also based on the similar multi-block ADMM procedure in **Algorithm 1**. As shown in Fig. 10, we take two typical images into account and both of them have large oscillating textures intertwined with piece-wise constant and smoothing structures. Clearly, the textural models in weaker space G and H^{-1} tend to give similar results as the one in L^1 space when combined with the semi-sparsity higher-order L_0 regularization, while the one in H^{-1} space reveals slight degradation in local structural regions. An advantage of these modified models is that they tend to preserve more local structural information when compared with the cutting-edge RTV method in a similar level of smoothness. Notice also that the textural model in L^1 space generally produces better visual results with more sharpening edges among the piece-wise smoothing surfaces. This may be due to the fact that the weaker space G and H^{-1} tend to capture the periodical oscillating details, while the underlying image structures and textures could be in very complicated forms. As a result, they may produce unexpected results around the irregular textural parts, for example, causing blur edges nearby the non-periodical textures. In contrast, the model in L^1 space is more robust to these complex situations and this is also one of the reasons that we model the textural part in L^1 space. The conclusion is also verified in Tab. 4, where the textural model in L^1 space has slightly better performance in both quality and efficiency. Finally, we would like to mention that the semi-sparsity higher-order regularization can be also utilized for more complex image decomposition problems, for example, the situation of structure, texture, and noise. More exploration and discussion are out of scope here. The reader is also referred to the work [3], [32], [33], [47], [49] for more related details.

7 CONCLUSION

In this work, we have proposed a simple but effective semi-sparsity model for image structure and texture decomposition. We demonstrate its advantages based on a fast ADMM solver. Experimental results also show that such a semi-sparse minimization has the capacity of preserving sharp edges without introducing the notorious staircase

artifacts in piece-wise smoothing regions and is also applicable for decomposing image textures with strong oscillatory patterns when applied to natural images. Some avenues of research for more complex decomposition models are also possible based on the semi-sparsity priors, which are leaving for further work.

REFERENCES

- [1] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Josa*, vol. 61, no. 1, pp. 1–11, 1971.
- [2] A. Gilchrist, *Seeing black and white*. Oxford University Press, 2006.
- [3] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*. American Mathematical Soc., 2001, vol. 22.
- [4] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [5] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM transactions on graphics (TOG)*, vol. 31, no. 6, pp. 1–10, 2012.
- [6] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3. ACM, 2008, p. 67.
- [7] S. Ghosh, R. G. Gavaskar, D. Panda, and K. N. Chaudhury, "Fast scale-adaptive bilateral texture smoothing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2015–2026, 2019.
- [8] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of scientific computing*, vol. 19, no. 1, pp. 553–572, 2003.
- [9] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 3606–3613.
- [10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [11] W. Yin, D. Goldfarb, and S. Osher, "The total variation regularized l1 model for multiscale decomposition," *Multiscale Modeling & Simulation*, vol. 6, no. 1, pp. 190–211, 2007.
- [12] J. Weickert et al., *Anisotropic diffusion in image processing*. Teubner Stuttgart, 1998, vol. 1.
- [13] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Iccv*, vol. 98, no. 1, 1998, p. 2.
- [14] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l0 gradient minimization," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 6. ACM, 2011, p. 174.
- [15] S. Osher, A. Solé, and L. Vese, "Image decomposition and restoration using total variation minimization and the h," *Multiscale Modeling & Simulation*, vol. 1, no. 3, pp. 349–370, 2003.
- [16] K. Bredies and M. Holler, "Higher-order total variation approaches and generalisations," *Inverse Problems*, vol. 36, no. 12, p. 123001, 2020.
- [17] X. Liu, "A new tgv-gabor model for cartoon-texture image decomposition," *IEEE Signal Processing Letters*, vol. 25, no. 8, pp. 1221–1225, 2018.
- [18] V. Duval, J.-F. Aujol, and Y. Gousseau, "The tvl1 model: a geometric point of view," *Multiscale Modeling & Simulation*, vol. 8, no. 1, pp. 154–189, 2009.
- [19] W. Yin, D. Goldfarb, and S. Osher, "Image cartoon-texture decomposition and feature selection using the total variation regularized l1 functional," in *International Workshop on Variational, Geometric, and Level Set Methods in Computer Vision*. Springer, 2005, pp. 73–84.
- [20] T. F. Chan and S. Esedoglu, "Aspects of total variation regularized l1 function approximation," *SIAM Journal on Applied Mathematics*, vol. 65, no. 5, pp. 1817–1837, 2005.
- [21] J. Huang, H. Wang, X. Wang, and M. Ruzhansky, "Semi-sparsity for smoothing filters," *IEEE Transactions on Image Processing*, vol. 32, pp. 1627–1639, 2023.
- [22] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2012.
- [23] H. Cho, H. Lee, H. Kang, and S. Lee, "Bilateral texture filtering," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, pp. 1–8, 2014.

- [24] S. Ono, " $L_{\{0\}}$ gradient projection," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1554–1564, 2017.
- [25] M. Bergounioux and L. Piffet, "A second-order model for image denoising," *Set-Valued and Variational Analysis*, vol. 18, pp. 277–306, 2010.
- [26] K. Papafitsoros and C.-B. Schönlieb, "A combined first and second order variational approach for image reconstruction," *Journal of mathematical imaging and vision*, vol. 48, pp. 308–338, 2014.
- [27] K. Bredies, K. Kunisch, and T. Valkonen, "Properties of l_1 - l_2 : The one-dimensional case," *Journal of Mathematical Analysis and Applications*, vol. 398, no. 1, pp. 438–454, 2013.
- [28] K. Bredies and M. Holler, "A l_2 -based framework for variational image decompression, zooming, and reconstruction. part i: Analytics," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2814–2850, 2015.
- [29] H. Schaeffer and S. Osher, "A low patch-rank interpretation of texture," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 226–262, 2013.
- [30] R. Xu, Y. Xu, Y. Quan, and H. Ji, "Cartoon-texture image decomposition using orientation characteristics in patch recurrence," *SIAM Journal on Imaging Sciences*, vol. 13, no. 3, pp. 1179–1210, 2020.
- [31] S. Alliney, "A property of the minimum vectors of a regularizing functional defined by means of the absolute norm," *IEEE transactions on signal processing*, vol. 45, no. 4, pp. 913–917, 1997.
- [32] M. Jung and M. Kang, "Simultaneous cartoon and texture image restoration with higher-order regularization," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 721–756, 2015.
- [33] L. H. Lieu and L. A. Vese, "Image restoration and decomposition via bounded total variation and negative hilbert-sobolev spaces," *Applied Mathematics and Optimization*, vol. 58, pp. 167–193, 2008.
- [34] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5638–5653, 2014.
- [35] Y. Kim, D. Min, B. Ham, and K. Sohn, "Fast domain decomposition for global image smoothing," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 4079–4091, 2017.
- [36] S. Paris, S. W. Hasinoff, and J. Kautz, "Local laplacian filters: Edge-aware image processing with a laplacian pyramid," *ACM Trans. Graph.*, vol. 30, no. 4, p. 68, 2011.
- [37] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part III* 13. Springer, 2014, pp. 815–830.
- [38] Y. Sun, S. Schaefer, and W. Wang, "Image structure retrieval via l_0 minimization," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 7, pp. 2129–2139, 2017.
- [39] P. Xu and W. Wang, "Structure-aware window optimization for texture filtering," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4354–4363, 2019.
- [40] K. Papafitsoros, "Novel higher order regularisation methods for image reconstruction," Ph.D. dissertation, University of Cambridge, 2015.
- [41] Y. Gao and K. Bredies, "Infimal convolution of oscillation total generalized variation for the recovery of images with structured texture," *SIAM Journal on Imaging Sciences*, vol. 11, no. 3, pp. 2021–2063, 2018.
- [42] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.
- [43] E. Tadmor, S. Nezzar, and L. Vese, "A multiscale image representation using hierarchical (bv, l_2) decompositions," *Multiscale Modeling & Simulation*, vol. 2, no. 4, pp. 554–579, 2004.
- [44] M. Nikolova, "A variational approach to remove outliers and impulse noise," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 99–120, 2004.
- [45] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, pp. 120–145, 2011.
- [46] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "Image decomposition into a bounded variation component and an oscillating component," *Journal of Mathematical Imaging and Vision*, vol. 22, no. 1, pp. 71–88, 2005.
- [47] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition—modeling, algorithms, and parameter selection," *International journal of computer vision*, vol. 67, no. 1, pp. 111–136, 2006.
- [48] J. Duan, Z. Qiu, W. Lu, G. Wang, Z. Pan, and L. Bai, "An edge-weighted second order variational model for image decomposition," *Digital Signal Processing*, vol. 49, pp. 162–181, 2016.
- [49] B. Shi, G. Meng, Z. Zhao, and Z.-F. Pang, "Image decomposition based on the adaptive direction total variation and g -norm regularization," *Signal, Image and Video Processing*, vol. 15, no. 1, pp. 155–163, 2021.
- [50] T. M. Le and L. A. Vese, "Image decomposition using total variation and div (bmo)," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 390–423, 2005.
- [51] J. B. Garnett, T. M. Le, Y. Meyer, and L. A. Vese, "Image decompositions using bounded variation and generalized homogeneous besov spaces," *Applied and Computational Harmonic Analysis*, vol. 23, no. 1, pp. 25–56, 2007.
- [52] S. Ono, T. Miyata, and I. Yamada, "Cartoon-texture image decomposition using blockwise low-rank texture characterization," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1128–1142, 2014.
- [53] H. Chen, Z. Xu, Q. Feng, Y. Fan, and Z. Li, "An l_0 regularized cartoon-texture decomposition model for restoring images corrupted by blur and impulse noise," *Signal Processing: Image Communication*, vol. 82, p. 115762, 2020.
- [54] C. Wang, L. Xu, and L. Liu, "Structure-texture image decomposition via non-convex total generalized variation and convolutional sparse coding," *The Visual Computer*, pp. 1–16, 2022.
- [55] J. C. Moreno, V. Prasath, D. Vorotnikov, H. Proença, and K. Palaniappan, "Adaptive diffusion constrained total variation scheme with application to cartoon+ texture+ edge image decomposition," *arXiv preprint arXiv:1505.00866*, 2015.
- [56] V. Le Guen, "Cartoon+ texture image decomposition by the tv- l_1 model," *Image Processing On Line*, vol. 4, pp. 204–219, 2014.
- [57] J. Xu, W. Shang, and Y. Hao, "A new cartoon+ texture image decomposition model based on the sobolev space," *Signal, Image and Video Processing*, vol. 16, no. 6, pp. 1569–1576, 2022.
- [58] J. Mairal, F. Bach, J. Ponce *et al.*, "Sparse modeling for image and vision processing," *Foundations and Trends® in Computer Graphics and Vision*, vol. 8, no. 2-3, pp. 85–283, 2014.
- [59] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, 2010, vol. 2, no. 1.
- [60] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [61] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [62] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.
- [63] D. Davis and W. Yin, "A three-operator splitting scheme and its optimization applications," *Set-valued and variational analysis*, vol. 25, pp. 829–858, 2017.
- [64] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [65] Y. Wang, W. Yin, and J. Zeng, "Global convergence of admm in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, pp. 29–63, 2019.
- [66] D.-R. Han, "A survey on some recent developments of alternating direction method of multipliers," *Journal of the Operations Research Society of China*, pp. 1–52, 2022.
- [67] R. Franzen, "Kodak lossless true color image suite," <http://r0k.us/graphics/kodak>, accessed: 2023-03-30.



Junqing Huang received the BS degree in Automation from the School of Electrical Engineering, Zhengzhou University, Zhengzhou, China, in 2011, and the MS degree in Mathematics from the School of Mathematical Sciences, Beihang University (BUAA), Beijing, China, in 2015. He is currently a Ph.D. candidate of Department of Mathematics: Analysis, Logic and Discrete Mathematics, Ghent University, Belgium. His research interests include deep learning, image processing, optimal transport and optimization.



Haihui Wang received the PhD degree in mathematics from Beijing University, Beijing, China, in 2003. She is currently a full professor with the school of Mathematics and Sciences, Beihang University (BUAA), Beijing, China. Her research interests include artificial intelligence, machine learning, signal and image processing, wavelet analysis and applications.



Michael Ruzhansky is currently a senior full professor in the Department of Mathematics and a Professorship in Special Research Fund (BOF) at Ghent University of Belgium, a Professorship in School of Mathematical Sciences at Queen Mary University of London, UK, and Honorary Professorship in Department of Mathematics at Imperial College London, UK. He was awarded by FWO (Belgium) the prestigious Odysseus 1 Project in 2018, he was recipient of several Prizes and Awards: ISAAC Award in 2007, Daiwa Adrian Prize in 2010 and Ferran Sunyer I Balaguer Prizes in 2014 and 2018. His research interests include different areas of analysis, in particular, theory of PDEs, microlocal analysis, and harmonic analysis.