

# Image Cartoon-Texture Decomposition and Feature Selection using the Total Variation Regularized $L^1$ Functional

Wotao Yin<sup>1</sup>, Donald Goldfarb<sup>1</sup>, and Stanley Osher<sup>2</sup>

<sup>1</sup> Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, USA. {wy2002, goldfarb}@columbia.edu

<sup>2</sup> Department of Mathematics, University of California at Los Angeles, Los Angeles, CA, USA. sjo@math.ucla.edu

**Abstract.** This paper studies the model of minimizing total variation with an  $L^1$ -norm fidelity term for decomposing a real image into the sum of cartoon and texture. This model is also analyzed and shown to be able to select features of an image according to their scales.

## 1 Introduction

Let  $f$  be an observed image which contains texture and/or noise. Texture is characterized as repeated and meaningful structure of small patterns. Noise is characterized as uncorrelated random patterns. The rest of an image, which is called *cartoon*, contains object hues and sharp edges (boundaries). Thus an image  $f$  can be decomposed as  $f = u + v$ , where  $u$  represents image cartoon and  $v$  is texture and/or noise. A general way to obtain this decomposition using the variational approach is to solve the problem  $\min \{ \int |Du| \mid \|u - f\|_B \leq \sigma \}$ , where  $Du$  denotes the generalized derivative of  $u$  and  $\|\cdot\|_B$  is a norm (or semi-norm). The total variation of  $u$ , which is  $\int |Du|$ , is minimized to regularize  $u$  while keep edges like object boundaries of  $f$  in  $u$  (i.e. allow discontinuities in  $u$ ). The fidelity term  $\|t(u, f)\|_B \leq \sigma$  forces  $u$  to be close to  $f$ . Among the recent total variation-based cartoon-texture decomposition models, Meyer [13] and Haddad & Meyer [10] proposed to use the  $G$ -norm, Vese & Osher [21] approximated the  $G$ -norm by the  $\text{div}(L^p)$ -norm, Osher & Sole & Vese [18] proposed to use the  $H^{-1}$ -norm, Lieu & Vese [12] proposed to use the more general  $H^{-s}$ -norm, and Le & Vese [11] proposed to use the  $\text{div}(BMO)$ -norm. In addition, Alliney [2–4], Nikolova [14–16], and Chan & Esedoglu [8] used the  $L^1$ -norm together with total variation. In this paper, we study the TV- $L^1$  model.

The rest of the paper is organized as follows. In Section 2 we define certain fundamental function spaces and norms. In Section 3 we present and analyze the TV- $L^1$  model. In particular, we relate the level sets of the input to the solution of the TV- $L^1$  model using a geometric argument and discuss the scale-selection and morphologically invariant properties of this model. The proofs of the lemmas, theorems, and corollaries are given in the technical report [22]. In Section 4 we briefly give the second-order cone programming (SOCP) formulation of this model. Numerical results illustrating the properties of the model are given in Section 5.

## 2 Preliminaries

Let  $u \in L^1$ , and define the total variation of  $u$  as

$$\|Du\| := \sup \left\{ \int u \operatorname{div}(\mathbf{g}) \, dx : \mathbf{g} \in C_0^1(\mathbb{R}^n; \mathbb{R}^n), |\mathbf{g}(x)|_{l^2} \leq 1 \, \forall x \in \mathbb{R}^n \right\},$$

and the  $BV$ -norm of  $u$  as  $\|u\|_{BV} := \|u\|_{L^1} + \|Du\|$ , where  $C_0^1(\mathbb{R}^n; \mathbb{R}^n)$  denotes the set of continuously differentiable vector-valued functions that vanish at infinity. The Banach space of functions with bounded variation is defined as  $BV := \{u \in L^1 : \|u\|_{BV} < \infty\}$  and is equipped with the  $\|\cdot\|_{BV}$ -norm.  $\|Du\|$  is often written in a less mathematically strict form  $\int |\nabla u|$ .

$\|Du\|$  and  $BV(\Omega)$  limited to  $\Omega$  are defined in analogy using  $\mathbf{g} \in C_0^1(\Omega; \mathbb{R}^n)$ .

Sets in  $\mathbb{R}^n$  with finite perimeter are often referred to as  $BV$  sets. The perimeter of a set  $S$  is defined by  $\operatorname{Per}(S) := \|D\mathbf{1}_S\|$ , where  $\mathbf{1}_S$  is the indicator function of  $S$ .

Next, we define the space  $G$  [13]. Let  $G$  denote the Banach space consisting of all *generalized functions*  $v(x)$  defined on  $\mathbb{R}^n$  that can be written as

$$v = \operatorname{div}(\mathbf{g}), \quad \mathbf{g} = [g_i]_{i=1, \dots, n} \in L^\infty(\mathbb{R}^n; \mathbb{R}^n), \quad (1)$$

and equipped with the norm  $\|v\|_G$  defined as the infimum of all  $L^\infty$  norms of the functions  $|\mathbf{g}(x)|_{l^2}$  over all decompositions (1) of  $v$ . In short,  $\|v\|_G := \inf\{\|\mathbf{g}(x)\|_{L^\infty} : v = \operatorname{div}(\mathbf{g})\}$ .

$G$  is the dual of the closed subspace  $\mathcal{BV}$  of  $BV$ , where  $\mathcal{BV} := \{u \in BV : |Du| \in L^1\}$  [13]. We note that finite difference approximations to functions in  $BV$  and in  $\mathcal{BV}$  are the same. For the definition and properties of  $G(\Omega)$ , where  $\Omega \subset \mathbb{R}^n$ , see [6].

It follows from the definitions of the  $BV$  and  $G$  spaces that

$$\int u v = \int u \nabla \cdot \mathbf{g} = - \int Du \cdot \mathbf{g} \leq \|Du\| \|v\|_G, \quad (2)$$

holds for any  $u \in \mathcal{BV}$  with a compact support and  $v \in G$ . We say  $(u, v)$  is an *extremal pair* if (2) holds with equality.

## 3 The TV- $L^1$ model.

The TV- $L^1$  model is define as a variational problem

$$\min_{u \in BV} TVL1_\lambda(u) = \min_{u \in BV} \int_\Omega |\nabla u| + \lambda \int |f - u|. \quad (3)$$

Although this model appears to be simple, it is very different to the ROF model [19]: it has the important property of being able to separate out features of a certain scale in an image as we shall show in the next section.

## 4 Analysis of the TV- $L^1$ model

In this section we first relate the parameter  $\lambda$  to the  $G$ -norm of the texture output  $v$ , then we focus on the TV- $L^1$  geometry and discuss the properties of the TV- $L^1$  model for scale-based feature selection in subsection 3.1.

Meyer [13] recently showed that the  $G$  space, which is equipped with the  $G$ -norm, contains functions with high oscillations. He characterized the solution  $u$  of the ROF model using the  $G$ -norm: given any input  $f$  defined on  $\mathbb{R}^n$ ,  $u$  satisfies  $\|f - u\|_G = \frac{1}{2\lambda}$  if  $\lambda > (2\|f\|_G)^{-1}$ , and  $u$  vanishes (i.e.,  $u \equiv 0$ ) if  $0 \leq \lambda \leq (2\|f\|_G)^{-1}$ . We can interpret this result as follows. First, no matter how regular  $f$  is,  $u$  is always different to  $f$  as long as  $f \neq 0$ . This is a major limitation of the ROF model, but it can be relaxed by applying the ROF model iteratively [17] or use the inverse TV flow [7]. Second, the texture/noise output  $v$  has its  $G$ -norm given by  $\min\{\frac{1}{2\lambda}, \|f\|_G\}$ . Therefore, the oscillating signal with  $G$ -norm less than  $\frac{1}{2\lambda}$  is removed by the ROF model. A similar characterization is given below for the TV- $L^1$  model in Theorems 1 and 2.

In order to use the  $G$ -norm, we first consider the approximate TV- $L^1$  model in which a perturbation  $\epsilon$  has been added to the fidelity term  $\|f - u\|_{L^1}$  to make it differentiable:

$$\min_{u \in BV(\Omega)} \int_{\Omega} |\nabla u| + \lambda \int_{\Omega} \sqrt{(f - u)^2 + \epsilon}, \quad (4)$$

where the image support  $\Omega$  is assumed to be compact. Since  $TVL1_{\lambda, \epsilon}(u)$  is strictly convex, problem (4) has a unique solution  $u_{\lambda, \epsilon}$ .

**Theorem 1.** *The solution  $u_{\lambda, \epsilon} (= f - v_{\lambda, \epsilon}) \in BV(\Omega)$  of the approximate TV- $L^1$  model satisfies*

$$\|\text{sign}_{\epsilon}(v_{\lambda, \epsilon})\|_G \leq 1/\lambda,$$

where  $\text{sign}_{\epsilon}(\cdot)$  is defined point-wise by  $\text{sign}_{\epsilon}(g)(x) := g(x)/\sqrt{|g(x)|^2 + \epsilon}$  for any function  $g$ .

Moreover, if  $\|\text{sign}_{\epsilon}(f)\|_G \leq 1/\lambda$ ,  $u_{\lambda, \epsilon} \equiv 0$  is the solution of the approximate TV- $L^1$  model.

If  $\|\text{sign}_{\epsilon}(f)\|_G > 1/\lambda$ , then there exists an optimal solution  $u_{\lambda, \epsilon}$  satisfying

- $\|\text{sign}_{\epsilon}(v_{\lambda, \epsilon})\|_G = 1/\lambda$ ;
- $\int u_{\lambda, \epsilon} \text{sign}_{\epsilon}(v_{\lambda, \epsilon}) = \|Du_{\lambda, \epsilon}\|/\lambda$ , i.e.,  $u_{\lambda, \epsilon}$  and  $\text{sign}_{\epsilon}(v_{\lambda, \epsilon})$  form an extremal pair.

Next, we relate the solution of the perturbed TV- $L^1$  model to the solution of the (unperturbed) TV- $L^1$  model.

**Theorem 2.** *Assuming the TV- $L^1$  model (3) using parameter  $\lambda$  has a unique solution  $u_{\lambda}$ , then the solution of approximate TV- $L^1$  model (4) using the same parameter  $\lambda$  satisfies*

$$\lim_{\epsilon \downarrow 0^+} \|u_{\lambda, \epsilon} - u_{\lambda}\|_{L^1} = 0, \quad \lim_{\epsilon \downarrow 0^+} \|v_{\lambda, \epsilon} - v_{\lambda}\|_{L^1} = 0.$$

We note that Chan and Esedoglu [8] proved that (4) has a unique solution for almost all  $\lambda$ 's with respect to the Lebesgue measure.

In the above two theorems, for  $\epsilon$  small enough, the value of  $\text{sign}_{\epsilon}(v)(x)$  can be close to  $\text{sign}(v)(x)$  even for small  $v(x)$ . In contrast to  $\|v\|_G = \min\{\frac{1}{2\lambda}, \|f\|_G\}$  for the

solution  $v$  of the ROF model, Theorems 1 and 2 suggest that the solution  $v$  of the TV- $L^1$  model can be much smaller. In other words, the TV- $L^1$  may not always remove some oscillating signal from  $f$  and erode the structure. This is supported by the following analytic example from [8]: if  $f$  equal to the disk signal  $B_r$ , which has radius  $r$  and unit height, then the solution  $u_\lambda$  of the TV- $L^1$  model is 0 if  $0 < \lambda < 2/r$ ,  $f$  if  $\lambda > 2/r$ , and  $cf$  for any  $c \in [0, 1]$  if  $\lambda = 2/r$ . Clearly, depending on  $\lambda$ , either 0 or the input  $f$  minimizes the TV- $L^1$  functional. This example also demonstrates the ability of the model to select the disk feature by its “scale”  $r/2$ . The next subsection focuses on this scale-based selection.

#### 4.1 TV- $L^1$ Geometry

To use the TV- $L^1$  model to separate large-scale and small-scale features, we are often interested in an appropriate  $\lambda$  that will allow us to extract geometric features of a given scale. For general input, the TV- $L^1$  model, which has only one scalar parameter  $\lambda$ , returns images combining many features. Therefore, we are interested in determining a  $\lambda$  that gives the whole targeted features with the least unwanted features in the output.

For simplicity, we assume  $\Omega = \mathbb{R}^2$  in this section. Our analysis starts with the decomposition of  $f$  using level sets and relies on the co-area formula (5) [?] and “layer cake” formula (6) [8], below. Then, we derive a TV- $L^1$  solution formula (9), in which  $u^*$  is built slice by slice. Each slice is then characterized by feature scales using the  $G$ -value, which extends the  $G$ -norm, and the *slopes* in Theorem 3, below. Last, we relate the developed properties to real-world applications. In the following we let  $U(g, \mu) := \{x \in \text{Dom}(g) : g(x) > \mu\}$  denote the (upper) level set of a function  $g$  at level  $\mu$ .

The co-area formula [?] for functions of bounded variation is

$$\int |Du| = \int_{-\infty}^{\infty} \text{Per}(U(u, \mu)) d\mu. \quad (5)$$

Using (5), Chan and Esedoglu [8] showed that the  $TVL1_\lambda$  functional can be represented as an integral over the perimeter and weighted areas of certain level sets by the following “layer cake” formula:

$$TVL1_\lambda(u) = \int_{-\infty}^{\infty} (\text{Per}(U(u, \mu)) + \lambda |U(u, \mu) \setminus U(f, \mu)| + \lambda |U(f, \mu) \setminus U(u, \mu)|) d\mu, \quad (6)$$

where  $|S|$  for a set  $S$  returns the area of  $S$ . Therefore, an optimal solution  $u_\lambda$  to the TV- $L^1$  model can be obtained by minimizing the right-hand side of (6). We are interested in finding a  $u^*$  such that  $U(u^*, \mu)$  minimizes the integrand for almost all  $\mu$ .

Let us fix  $\lambda$  and focus on the integrand of the above functional and introduce the following notation:

$$C(\Gamma, \Sigma) := \text{Per}(\Sigma) + \lambda |\Sigma \setminus \Gamma| + \lambda |\Gamma \setminus \Sigma| \quad (7)$$

$$\min_{\Sigma} C(\Gamma, \Sigma) \quad (8)$$

where  $\Gamma$  and  $\Sigma$  are sets with bounded perimeters in  $\mathbb{R}^2$ . Let  $\Sigma_{f, \mu}$  denote a solution of (8) for  $\Gamma = U(f, \mu)$ . From the definition of the upper level set, for the existence of a

$u$  satisfying  $U(u, \mu) = \Sigma_{f, \mu}$  for all  $\mu$ , we need  $\Sigma_{f, \mu_1} \supseteq \Sigma_{f, \mu_2}$  for any  $\mu_1 < \mu_2$ . This result is given in the following lemma:

**Lemma 1.** *Let the sets  $\Sigma_1$  and  $\Sigma_2$  be the solutions of (6) for  $\Gamma = \Gamma_1$  and  $\Gamma = \Gamma_2$ , respectively, where  $\Gamma_1$  and  $\Gamma_2$  are two sets satisfying  $\Gamma_1 \supset \Gamma_2$ .*

*If either one or both of  $\Sigma_1$  and  $\Sigma_2$  are unique minimizers, then  $\Sigma_1 \supseteq \Sigma_2$ ; otherwise, i.e., both are not unique minimizers,  $\Sigma_1 \supseteq \Sigma_2$  may not hold, but in this case,  $\Sigma_1 \cup \Sigma_2$  is a minimizer of (8) for  $\Gamma = \Gamma_1$ .*

*Therefore, there always exists a solution of (8) for  $\Gamma = \Gamma_1$  that is a superset of any minimizer of (8) for  $\Gamma = \Gamma_2$ .*

Using the above lemma, we get the following geometric solution characterization for the TV- $L^1$  model:

**Theorem 3.** *Suppose that  $f \in BV$  has essential infimum  $\mu_0$ . Let function  $u^*$  be defined point-wise by*

$$u^*(x) := \mu_0 + \int_{\mu_0}^{\infty} \mathbf{1}_{\Sigma_{f, \mu}}(x) d\mu, \quad (9)$$

where  $\Sigma_{f, \mu}$  is the solution of (8) for  $\Gamma = U(f, \mu)$  that satisfies  $\Sigma_{f, \mu_1} \supseteq \Sigma_{f, \mu_2}$  for any  $\mu_1 < \mu_2$ , i.e.,  $\Sigma_{f, \mu}$  is monotonically decreasing with respect to  $\mu$ . Then  $u^*$  is an optimal solution of the TV- $L^1$  model (3).

Next, we illustrate the implications of the above theorem by applying the results in [20] to (8). In [20], the authors introduced the  $G$ -value, which is an extension of Meyer's  $G$ -norm, and obtained a characterization to the solution of the TV- $L^1$  model based on the  $G$ -value and the Slope [5]. These results are presented in the definition and the theorem below.

**Definition 1.** *Let  $\Psi : \mathbb{R}^2 \rightarrow 2^{\mathbb{R}}$  be a set-valued function that is measurable in the sense that  $\Psi^{-1}(S)$  is Lebesgue measurable for every open set  $S \subset \mathbb{R}$ . We do not distinguish  $\Psi$  between a set-valued function and a set of measurable (single-valued) functions, and let*

$$\Psi := \{\text{measurable function } \psi \text{ satisfying } \psi(x) \in \Psi(x), \forall x\}.$$

The  $G$ -value of  $\Psi$  is defined as follows:

$$G(\Psi) := \sup_{h \in C_0^\infty : \int |\nabla h| = 1} - \sup_{\psi \in \Psi} \int \psi(x) h(x) dx. \quad (10)$$

**Theorem 4.** *Let  $\partial|f|$  denote the set-valued sub-derivative of  $|f|$ , i.e.,  $\partial|f|(x)$  equals  $\text{sign}(f(x))$  if  $f(x) \neq 0$  and equals the interval  $[-1, 1]$  if  $f(x) = 0$ . Then, for the TV- $L^1$  model (3),*

1.  $u_\lambda = 0$  is an optimal solution if and only if  $\lambda \leq \frac{1}{G(\partial|f|)}$ ;
2.  $u_\lambda = f$  is an optimal solution if and only if  $\lambda \geq \sup_{h \in BV} \frac{\|Df\| - \|Dh\|}{\int |f-h|}$ ,

where  $\frac{1}{G(\partial|f|)} \leq \sup_{h \in BV} \frac{\|Df\| - \|Dh\|}{\int |f-h|}$ ,  $\forall f \in BV$ .

It follows from the ‘‘layer cake’’ formula (6) that solving the geometric problem (8) is equivalent to solving the TV- $L^1$  model with input  $f = \mathbf{1}_\Gamma$ . Therefore, by applying Theorem 4 to  $f = \mathbf{1}_\Gamma$ , we can characterize the solution of (6) as follows:

**Corollary 1.** *For the geometric problem (8) with a given  $\lambda$ ,*

1.  $\Sigma_\lambda = \emptyset$  is an optimal solution if and only if  $\lambda \leq \frac{1}{G(\partial|\mathbf{1}_\Gamma|)}$ ;
2.  $\Sigma_\lambda = \Gamma$  is an optimal solution if and only if  $\lambda \geq \sup_{h \in BV} \frac{\|D\mathbf{1}_\Gamma\| - \|Dh\|}{\int |\mathbf{1}_\Gamma - h|}$ .

Corollary 1, together with Theorem 3, implies the followings. Suppose that the mask set  $S$  of a geometric feature  $F$  coincides with  $U(f, \mu)$  for  $\mu \in [\mu_0, \mu_1]$ . Then, for any  $\lambda < 1/G(\partial|\mathbf{1}_S|)$ ,  $\Sigma_{f, \mu} = \emptyset$  for  $\mu \in [\mu_0, \mu_1]$ ; hence, the geometric feature  $F$  is not observable in  $u_\lambda$ . In the example where  $F = f = cB_r$  (recall that  $B_r$  is the disk function with radius  $r$  and unit height),  $S$  and  $U(f, \mu)$  are the circle  $\bar{B}_r$  with radius  $r$  for  $\mu \in [0, c)$ , and  $G(\partial|\mathbf{1}_S|) = G(\partial|B_r|) = r/2$ . Therefore, if  $\lambda < 1/G(\partial|\mathbf{1}_S|) = 2/r$ ,  $\Sigma_{f, \mu} = \emptyset$  for  $\mu \in [0, c)$ . Also because  $\mu_0 = 0$  and  $\Sigma_{f, \mu} = \emptyset$  for  $\mu \geq c$  in (9),  $u_\lambda \equiv 0$ , which means the feature  $F = cB_r$  is not included in  $u_\lambda$ .

If  $\lambda > 1/G(\partial|\mathbf{1}_S|)$ ,  $\Sigma_{f, \mu} \neq \emptyset$  for  $\mu \in [\mu_0, \mu_1]$ , which implies at least some part of the feature  $F$  can be observed in  $u_\lambda$ . Furthermore, if  $\lambda \geq \sup_{h \in BV} (\|D\mathbf{1}_\Gamma\| - \|Dh\|) / \int |\mathbf{1}_\Gamma - h|$ , we get  $\Sigma_{f, \mu} = U(f, \mu) = S$  for  $\mu \in [\mu_0, \mu_1]$  and therefore, the feature  $F$  is fully contained in  $u_\lambda$ . In the above example where  $F = f = cB_r$  and  $S = \bar{B}_r$ , it turns out  $2/r = 1/G(\partial|\mathbf{1}_S|) = \sup_{h \in BV} (\|D\mathbf{1}_\Gamma\| - \|Dh\|) / \int |\mathbf{1}_\Gamma - h|$ . Therefore, if  $\lambda > 2/r$ ,  $\Sigma_{f, \mu} = S$  for  $\mu \in [0, c)$ , and  $u_\lambda = cB_r = f$ .

In general, although a feature is often different from its vicinity in intensity, it cannot monopolize a level set of the input  $f$ , i.e., it is represented by an isolated sets in  $U(f, \mu)$ , for some  $\mu$ , which also contains isolated sets representing other features. Consequently,  $u_\lambda$  that contains a targeted feature may also contain many other features. However, from Theorem 3 and Corollary 1, we can easily see that the arguments for the case  $S = U(f, \mu)$  still hold for the case  $S \subset U(f, \mu)$ .

**Proposition 1.** *Suppose there are a sequences of features in  $f$  that are represented by sets  $S_1, S_2, \dots, S_l$  and have distinct intensity values. Let*

$$\lambda_i^{\min} := \frac{1}{G(\partial|\mathbf{1}_{S_i}|)}, \quad \lambda_i^{\max} := \sup_{h \in BV} \frac{\|D\mathbf{1}_{S_i}\| - \|Dh\|}{\int |\mathbf{1}_{S_i} - h|}, \quad (11)$$

for  $i = 1, \dots, l$ . If the features have decreasing scales and, in addition, the following holds

$$\lambda_1^{\min} \leq \lambda_1^{\max} < \lambda_2^{\min} \leq \lambda_2^{\max} < \dots < \lambda_l^{\min} \leq \lambda_l^{\max}, \quad (12)$$

then feature  $i$ , for  $i = 1, \dots, l$ , can be precisely retrieved as  $u_{\lambda_i^{\max} + \epsilon} - u_{\lambda_i^{\min} - \epsilon}$  (here  $\epsilon$  is a small scalar that forces unique solutions because  $\lambda_i^{\min} = \lambda_i^{\max}$  is allowed).

This proposition holds since for  $\lambda = \lambda_i^{\min} - \epsilon$ , feature  $i$  completely vanishes in  $u_\lambda$ , but for  $\lambda = \lambda_i^{\max} + \epsilon$ , feature  $i$  is fully contained in  $u_\lambda$  while there is no change to any other features.

To extract a feature represented by set  $S$  in real-world applications, one can computer  $G(\partial|\mathbf{1}_S|)$  off-line and use  $\lambda$  slightly greater than  $1/G(\partial|\mathbf{1}_S|)$ . The intensity and the position of the feature in  $f$  are not required as priors.

Next, we present a corollary of Theorem 3 to finish this section.

**Corollary 2. [Morphological invariance]** For any strictly increasing function  $g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $u_\lambda(g \circ f) = g \circ u_\lambda(f)$ .

## 5 Second-order cone programming formulations

In this section, we briefly show how to formulate the discrete versions of the TV- $L^1$  model (3) as a second-order program (SOCP).

In an SOCP the vector of variables  $\mathbf{x} \in \mathbb{R}^n$  is composed of subvectors  $\mathbf{x}_i \in \mathbb{R}^{n_i}$  – i.e.,  $\mathbf{x} \equiv (\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_r)$  – where  $n = n_1 + n_2 + \dots + n_r$  and each subvector  $\mathbf{x}_i$  must lie either in an elementary *second-order cone* of dimension  $n_i$

$$\mathcal{K}^{n_i} \equiv \{\mathbf{x}_i = (x_i^0; \bar{\mathbf{x}}_i) \in \mathbb{R} \times \mathbb{R}^{n_i-1} \mid \|\bar{\mathbf{x}}_i\| \leq x_i^0\},$$

or an  $n_i$ -dimensional *rotated second-order cone*

$$\mathcal{Q}^{n_i} \equiv \{\mathbf{x}_i \in \mathbb{R}^{n_i} \mid \mathbf{x}_i = \bar{\mathbf{x}}, 2\bar{x}_1\bar{x}_2 \geq \sum_{i=3}^{n_i} \bar{x}_i^2, \bar{x}_1, \bar{x}_2 \geq 0\},$$

which is an elementary second-order cone under a linear transformation.

With these definitions an SOCP can be written in the following form [1]:

$$\begin{aligned} \min \quad & \mathbf{c}_1^\top \mathbf{x}_1 + \dots + \mathbf{c}_r^\top \mathbf{x}_r \\ \text{s.t.} \quad & A_1 \mathbf{x}_1 + \dots + A_r \mathbf{x}_r = \mathbf{b} \\ & \mathbf{x}_i \in \mathcal{K}^{n_i} \text{ or } \mathcal{Q}^{n_i}, \quad \text{for } i = 1, \dots, r, \end{aligned} \quad (13)$$

where  $\mathbf{c}_i \in \mathbb{R}^{n_i}$  and  $A_i \in \mathbb{R}^{m \times n_i}$ , for any  $i$ , and  $\mathbf{b} \in \mathbb{R}^m$ . As is the case for linear programs, SOCPs can be solved in polynomial time by interior point methods.

We assume that images are represented as 2-dimensional  $n \times n$  matrices, whose elements give the “grey” values of corresponding pixels, i.e.,  $f_{i,j} = u_{i,j} + v_{i,j}$ , for  $i, j = 1, \dots, n$ .

First, as the total variation of  $u$  is defined discretely by forward finite differences as  $\int |\nabla u| := \sum_{i,j} [((\partial_x^+ u)_{i,j})^2 + ((\partial_y^+ u)_{i,j})^2]^{1/2}$ , by introducing new variables  $t_{i,j}$ , we can express  $\min\{\int |\nabla u|\}$  as  $\min\{\sum_{i,j} t_{i,j}\}$  subject to the 3-dimensional second-order cones  $(t_{i,j}; (\partial_x^+ u)_{i,j}, (\partial_y^+ u)_{i,j}) \in \mathcal{K}^3$ . Second, minimizing the fidelity term  $\int |f - u|$  is equivalent to minimizing  $s$  subject to  $\sum_{i,j} (f_{i,j} - u_{i,j}) \leq s$  and  $\sum_{i,j} (u_{i,j} - f_{i,j}) \leq s$ . Therefore, the SOCP formulation of the TV- $L^1$  model is

$$\begin{aligned} \min_{s,t,u,\partial_x^+ u, \partial_y^+ u} \quad & \sum_{1 \leq i,j \leq n} t_{i,j} + \lambda s \\ \text{s.t.} \quad & (\partial_x^+ u)_{i,j} = u_{i+1,j} - u_{i,j} \quad \forall i, j = 1, \dots, n, \\ & (\partial_y^+ u)_{i,j} = u_{i,j+1} - u_{i,j} \quad \forall i, j = 1, \dots, n, \\ & \sum_{1 \leq i,j \leq n} (f_{i,j} - u_{i,j}) \leq s, \\ & \sum_{1 \leq i,j \leq n} (u_{i,j} - f_{i,j}) \leq s, \\ & (t_{i,j}; (\partial_x^+ u)_{i,j}, (\partial_y^+ u)_{i,j}) \in \mathcal{K}^3 \quad \forall i, j = 1, \dots, n. \end{aligned} \quad (14)$$

Finally, we note that both  $G(\partial|f|)$  and  $\sup_{h \in BV} \frac{\|Df\| - \|Dh\|}{\int |f-h|}$ , after homogenizing the objective function of the latter, can be easily developed based on the SOCP formulation of the total variation term  $\int |Dh|$ .

## 6 Numerical results

### 6.1 Comparison among three decomposition models

In this subsection, we present numerical results of the TV- $L^1$  model and compare them with the results of the Meyer [13] and the Vese-Osher (VO) [21] models, below.

$$\text{The Meyer model: } \min_{u \in BV} \left\{ \int |\nabla u| : \|v\|_G \leq \sigma, f = u + v \right\}.$$

$$\text{The Vese-Osher model: } \min_{u \in BV} \int |\nabla u| + \lambda \int |f - u - \text{div}(\mathbf{g})|^2 + \mu \int |\mathbf{g}|.$$

We also formulated these two models as SOCPs, in which no regularization or approximation is used (refer to [9] for details). We used the commercial package Mosek as our SOCP solver. In the first set of results, we applied the models to relatively noise-free images.

We tested textile texture decomposition by applying the three models to a part (Fig. 1 (b)) of the image ‘‘Barbara’’ (Fig. 1 (a)). Ideally, only the table texture and the strips on Barbara’s clothes should be extracted. Surprisingly, Meyer’s method did not give good results in this test as the texture  $v$  output clearly contains inhomogeneous background. To illustrate this effect, we used a very conservative parameter - namely, a small  $\sigma$  - in Meyer’s model. The outputs are depicted in Fig. 1 (d). As  $\sigma$  is small, some table cloth and clothes textures remain in the cartoon  $u$  part. One can imagine that by increasing  $\sigma$  we can get a result with less texture left in the  $u$  part, but with more inhomogeneous background left in the  $v$  part. While Meyer’s method gave unsatisfactory results, the other two models gave very good results in this test as little background is shown in Figures 1 (e) and (f). The Vese-Osher model was originally proposed as an approximation of Meyer’s model in which the  $L^\infty$ -norm of  $|\mathbf{g}|$  is approximated by the  $L^1$ -norm of  $|\mathbf{g}|$ . We guess that the use of the  $L^1$ -norm allows  $\mathbf{g}$  to capture more texture signal while the original  $L^\infty$ -norm in Meyer’s model makes  $\mathbf{g}$  to capture only the oscillatory pattern of the texture signal. Whether the texture or only the oscillatory pattern is more preferable depends on applications. For example, the latter is more desirable in analyzing fingerprint images. Compared to the Vese-Osher model, the TV- $L^1$  model generated a little sharper cartoon in this test. The biggest difference, however, is that the TV- $L^1$  model kept most brightness changes in the texture part while the other two kept them in the cartoon part. In the top right regions of the output images, the wrinkles of Barbara’s clothes are shown in the  $u$  part of Fig. 1 (e) but in the  $v$  part of (f). This shows that the texture extracted by TV-L1 has a wider dynamic range.

In the second set of results, we applied the three models to the image ‘‘Barbara’’ after adding a substantial amount of Gaussian noise (standard deviation equal to 20). The resulting noisy image is depicted in Fig. 1 (c). All the three models removed the noise together with the texture from  $f$ , but noticeably, the cartoon parts  $u$  in these results (Fig. 1 (g)-(i)) exhibit a staircase effect to different extents. We tested different parameters and conclude that none of the three decomposition models is able to separate image texture and noise.

## 6.2 Feature selection using the TV- $L^1$ model

Component	$\bar{S}_1$	$\bar{S}_2$	$\bar{S}_3$	$\bar{S}_4$	$\bar{S}_5$
$G$ -value	19.39390	13.39629	7.958856	4.570322	2.345214
$\lambda^{\min}$	0.0515626	0.0746475	0.125646	0.218803	0.426400
$\lambda_1 =$	$\lambda_2 =$	$\lambda_3 =$	$\lambda_4 =$	$\lambda_5 =$	$\lambda_6 =$
0.0515	0.0746	0.1256	0.2188	0.4263	0.6000

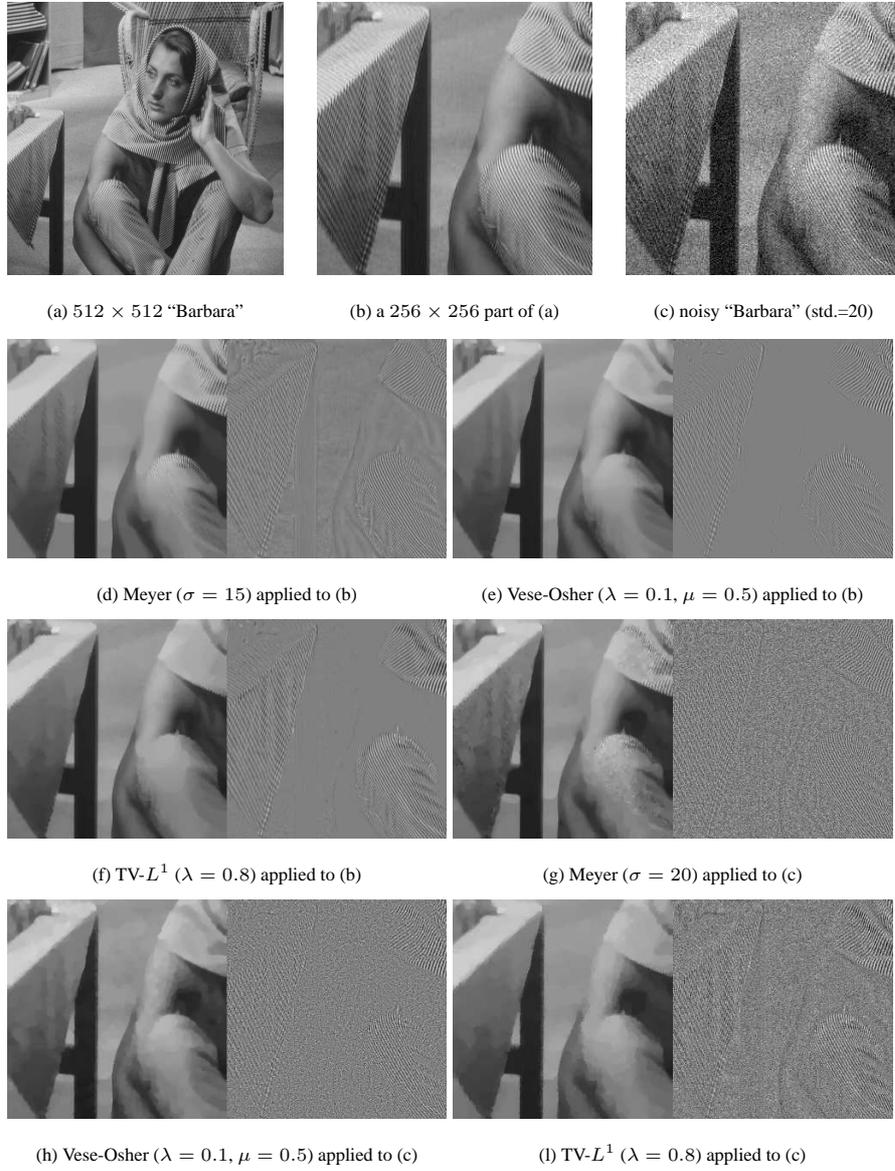
Table 1.

We applied the TV- $L^1$  model with different  $\lambda$ 's to the composite input image (Fig. 2 ( $f$ )). Each of the five components in this composite image is depicted in Fig. 2 ( $S_1$ )-( $S_5$ ). We name the components by  $S_1, \dots, S_5$  in the order they are depicted in Fig. 2. They are decreasing in scale. This is further shown by the decreasing  $G$ -values of their mask sets  $\bar{S}_1, \dots, \bar{S}_5$ , and hence, their increasing  $\lambda^{\min}$  values (see (11)), which are given in Table 1. We note that  $\lambda_1^{\max}, \dots, \lambda_6^{\max}$  are large since the components do not possess smooth edges in the pixelized images. This means that the property (12) does not hold for these components, so using the lambda values  $\lambda_1, \dots, \lambda_6$  given in Table 1 does not necessarily give entire feature signal in the output  $u$ . We can see from the numerical results depicted in Fig. 2 that we are able to produce output  $u$  that contains only those features with scales larger than  $1/\lambda_i$  and that leaves, in  $v$ , only a small amount of the signal of these features near non-smooth edges. For example, we can see the white boundary of  $S_2$  in  $v_3$  and four white pixels corresponding to the four corners of  $S_3$  in  $v_4$  and  $v_5$ . This is due to the nonsmoothness of the boundary and the use of finite difference. However, the numerical results closely match the analytic results given in Subsection 4.1. By forming differences between the outputs  $u_1, \dots, u_6$ , we extracted individual features  $S_1, \dots, S_5$  from input  $f$ . These results are depicted in the fourth row of images in Fig. 2.

We further illustrate the feature selection capacity of the TV- $L^1$  model by presenting two real-world applications. The first application is background correction for cDNA microarray images, in which the mRNA-cDNA gene spots are often plagued with the inhomogeneous background that should be removed. Since the gene spots have similar small scales, an appropriate  $\lambda$  can be easily derived from Proposition 1. The results are depicted in Fig. 2 (c)-(f). The second application is illumination removal for face recognition. Fig. 2 (i)-(iii) depicts three face images in which the first two images belong to the same face but were taken under different lighting conditions, and the third image belongs to another face. We decomposed their logarithm using the TV- $L^1$  model (i.e.,  $f \xrightarrow{\log} f' \xrightarrow{\text{TV-}L^1} u' + v'$ ) with  $\lambda = 0.8$  and obtained the images ( $v'$ ) depicted in Fig. 2 (iv)-(vi). Clearly, the first two images (Fig. 2 (iv) and (v)) are more correlated than their originals while they are very less correlated to the third. The role of the TV- $L^1$  model in this application is to extract the small-scale facial objects like the mouth edges, eyes, and eyebrows that are nearly illumination invariant. The processed images shall make the subsequent computerized face comparison and recognition easier.

## References

1. F. ALIZADEH AND D. GOLDFARB, *Second-order cone programming*, Mathematical Programming, Series B, 95(1), 3–51, 2003.
2. S. ALLINEY, *Digital filters as absolute norm regularizers*, IEEE Trans. on Signal Processing, 40:6, 1548–1562, 1992.
3. S. ALLINEY, *Recursive median filters of increasing order: a variational approach*, IEEE Trans. on Signal Processing, 44:6, 1346–1354, 1996.
4. S. ALLINEY, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Trans. on Signal Processing, 45:4, 913–917, 1997.
5. L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient flows, in metric spaces and in the space of probability measures*, Birkhäuser, 2005.
6. G. AUBERT AND J.F. AUJOL, *Modeling very oscillating signals. Application to image processing*, Applied Mathematics and Optimization, 51(2), March 2005.
7. M. BURGER, S. OSHER, J. XU, AND G. GILBOA, *Nonlinear inverse scale space methods for image restoration*, UCLA CAM Report, 05-34, 2005.
8. T.F. CHAN AND S. ESEDOGLU, *Aspects of total variation regularized  $L^1$  functions approximation*, UCLA CAM Report 04-07, to appear in SIAM J. Appl. Math.
9. D. GOLDFARB AND W. YIN, *Second-order cone programming methods for total variation-based image restoration*, Columbia University CORC Report TR-2004-05.
10. A. HADDAD AND Y. MEYER, *Variational methods in image processing*, UCLA CAM Report 04-52.
11. T. LE AND L. VESE, *Image decomposition using the total variation and  $\text{div}(BMO)$* , UCLA CAM Report 04-36.
12. L. LIEU AND L. VESE, *Image restoration and decomposition via bounded total variation and negative Hilbert-Sobolev spaces*, UCLA CAM Report 05-33.
13. Y. MEYER, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, University Lecture Series Volume 22, AMS, 2002.
14. M. NIKOLOVA, *Minimizers of cost-functions involving nonsmooth data-fidelity terms*, SIAM J. Numer. Anal., 40:3, 965–994, 2002.
15. M. NIKOLOVA, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, 20:1-2, 99–120, 2004.
16. M. NIKOLOVA, *Weakly constrained minimization. Application to the estimation of images and signals involving constant regions*, Journal of Mathematical Imaging and Vision 21:2, 155–175, 2004.
17. S. OSHER, M. BURGER, D. GOLDFARB, J. XU, AND W. YIN, *An iterative regularization method for total variation-based image restoration*, SIAM J. on Multiscale Modeling and Simulation 4(2), 460–489, 2005.
18. S. OSHER, A. SOLE, AND L.A. VESE, *Image decomposition and restoration using total variation minimization and the  $H^{-1}$  norm*, UCLA C.A.M. Report 02-57, (Oct. 2002).
19. L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D, 60, 259–268, 1992.
20. O. SCHERZER, W. YIN, AND S. OSHER, *Slope and G-set characterization of set-Valued functions and applications to non-Differentiable optimization problems*, UCLA CAM Report 05-35.
21. L. VESE AND S. OSHER, *Modelling textures with total variation minimization and oscillating patterns in image processing*, UCLA CAM Report 02-19, (May 2002).
22. W. YIN, D. GOLDFARB, AND S. OSHER, *Total variation-based image cartoon-texture decomposition*, Columbia University CORC Report TR-2005-01, UCLA CAM Report 05-27, 2005.



**Fig. 1.** Cartoon-texture decomposition and denoising results by the three models.

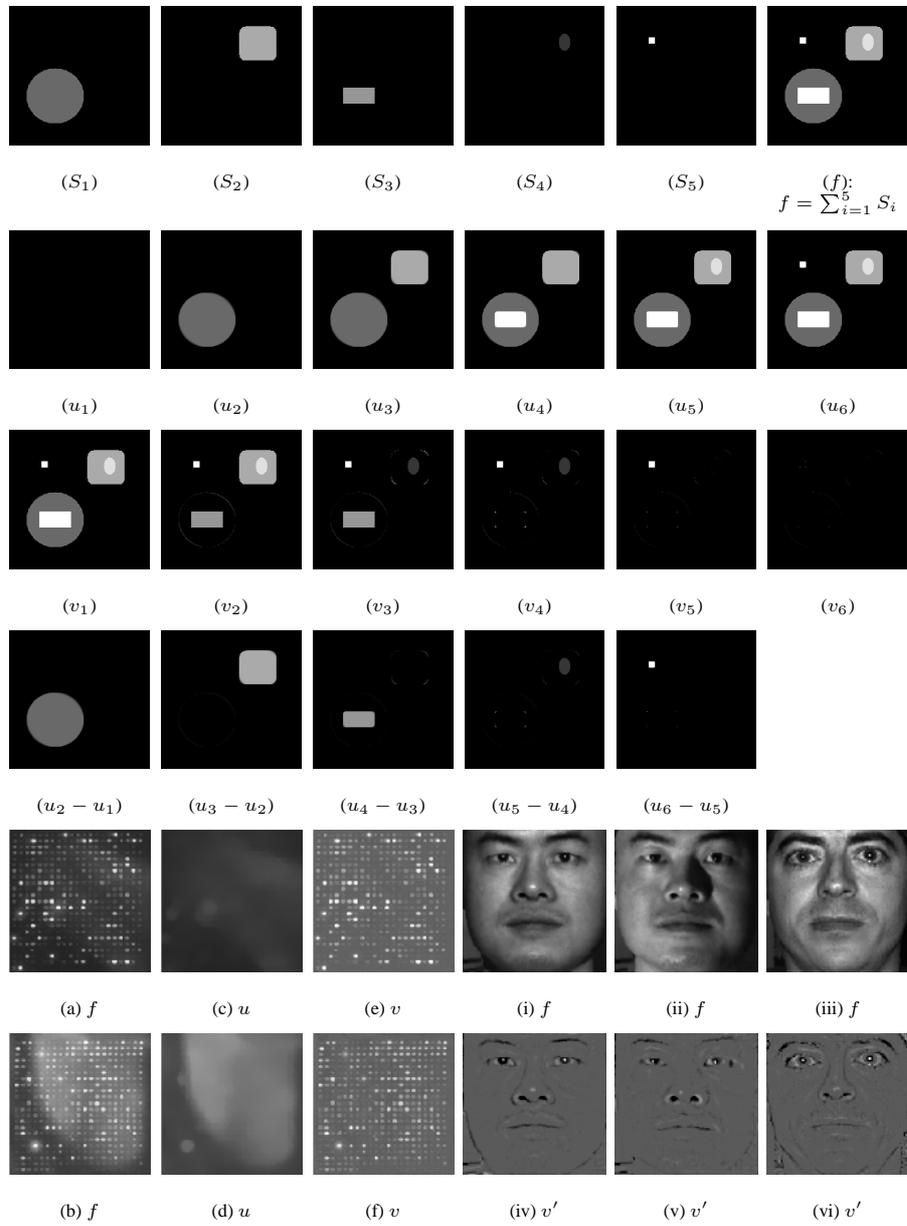


Fig. 2. Feature selection using the TV- $L^1$  model.