THE TOTAL VARIATION REGULARIZED L^1 MODEL FOR MULTISCALE DECOMPOSITION

WOTAO YIN*, DONALD GOLDFARB†, AND STANLEY OSHER‡

Abstract. This paper studies the total variation regularization model with an L^1 fidelity term (TV-L^1) for decomposing an image into features of different scales. We first show that the images produced by this model can be formed from the minimizers of a sequence of decoupled geometry subproblems. Using this result we show that the TV-L^1 model is able to separate image features according to their scales, where the scale is analytically defined by the G-value. A number of other properties including the geometric and morphological invariance of the TV-L^1 model are also proved and their applications discussed.

Key words. multiscale image decomposition, total variation, L^1 distance, feature selection, geometry optimization.

AMS subject classifications. 49Q10, 65K10, 65J22, 68U10, 94A08.

1. Introduction. Let a grey-scale n-dimensional image be represented by a function f on a domain of $\mathbb{R}^n$. In this paper, we restrict our discussion to typical 2-dimensional open domains (typically, $\Omega = \mathbb{R}^2$ or $(0,1)^2$). The TV-L^1 model obtains a decomposition of f by solving the following model:

$$\inf_{u\in BV} TV(u) + \lambda\|f-u\|_{L^1}, \tag{1.1}$$

for the minimizer u^*_λ and $v^*_\lambda = f - u^*_\lambda$, where BV is the space of functions of bounded variation, $TV(u)$ is the total variation of u, and $f \in L^1(\Omega)$. The latter is needed for technical reasons given below. Previous work on this model for image/signal processing includes Alliney's pioneering study [2, 3, 4] of the discrete version of (1.1), Nikolova's [31, 32, 33] discovery of the usefulness of this model for removing impulsive noise, Chan and Esedoglu's [16] further analysis of this model, and a series of applications of this model in computer vision by Chen et al. [21, 19, 20] and in biomedical imaging Yin et al. [40].

In this paper we extend the existing analysis of the TV-L^1 model. In particular, we show its equivalence to the non-convex geometry problem:

$$\min_U \mathbf{Per}(U) + \lambda\,|S\triangle U|, \tag{1.2}$$

where $\mathbf{Per}(U)$ is the perimeter of the set U, $S\triangle U \equiv (S\backslash U)\cup(U\backslash S)$ is the symmetric difference of the sets S and U, and $|S\triangle U|$ is its area. We use this equivalence to obtain results on TV-L^1's scale-based feature selection properties. We first give a brief introduction to the framework of total variation-based image processing models and the space of functions of bounded variation.

*Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005 (wotao.yin@rice.edu). Research supported by NSF Grant DMS-01-04282, ONR Grant N00014-03-1-0514, and DOE Grant GE-FG01-92ER-25126.

†Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027 (goldfarb@columbia.edu). Research supported by NSF Grants DMS-01-04282 and DMS-06-06712, ONR Grant N00014-03-1-0514, and DOE Grant GE-FG01-92ER-25126.

‡Department of Mathematics, UCLA, Los Angeles, CA 90095 (sjo@math.ucla.edu). Research supported by NSF Grants ITR ACI-0321917 and DMS-0312222, and NIH Grant P20 MH65166.

1.1. Notation. We use following notation throughout this paper (the TV and **Per** functionals are defined in the next section):

1. The TV-L^1 energy: $\mathbf{E}(u;\lambda,f) = TV(u) + \lambda\|u-f\|_{L^1}$;
2. The minimizer of $\mathbf{E}(u;\lambda,f)$: u^*, which may also be subscripted by λ;
3. The TV-L^1 operator: $T_t \cdot f$ denotes a minimizer u^* of $\mathbf{E}(u;1/t,f)$ (if there are multiple minimizers, $T_t \cdot f$ gives a specific u^*);
4. Composite operator: $T_s \circ T_t \cdot f = T_s \cdot (T_t \cdot f)$ denotes a minimizer of $\mathbf{E}(u;1/s,(T_s \cdot f))$;
5. The super level set: $L(f,\mu) = \{x \in \mathrm{dom}(f) : f(x) > \mu\}$;
6. The geometry energy: $\mathbf{E_G}(U;\lambda,S) = \mathbf{Per}(U) + \lambda\,|S \triangle U|$;
7. The minimizer of $\mathbf{E_G}(U;\lambda,S)$: U^*, which may also be subscripted by the penalty parameter λ or the level parameter μ, when $S = L(f,\mu)$, to emphasize the dependence on λ or μ, respectively;
8. For two sets S and T, $S+T := S \cup T$ when $S \cap T = \emptyset$ is (implicitly) assumed;
9. For two sets S and T, $S-T := S \setminus T$ when $S \supseteq T$ is (implicitly) assumed.

1.2. Total variation (TV) models. A general framework for obtaining a decomposition of an image f into a regular part u and an irregular part v is to solve the problem

$$\inf_u \{\|s(u)\|_A \;\mid\; \|t(u,f)\|_B \le \sigma\}, \tag{1.3}$$

for u, where $s(\cdot)$ and $t(\cdot,\cdot)$ are two functionals on appropriate spaces and $\|\cdot\|_A$ and $\|\cdot\|_B$ are norms (or semi-norms). $\|\cdot\|_A$ and $s(\cdot)$ should be chosen so that $\|s(u)\|_A$ is small for regular signals u but much bigger for irregular noise v. Then, minimizing $\|s(u)\|_A$ is equivalent to regularizing u according to the measure $\|s(u)\|_A$. A typical choice for $\|s(u)\|_A$ is $\int |Du|^p$, where $u \in BV$, the space of functions of bounded variation (see Def. 1.1), and Du denotes the generalized derivative of u. For $p > 1$, minimizing $\int |Du|^p$ tends to produce rather smooth functions. In particular, $p = 2$ gives Tihoknov regularization. Therefore, to keep edges like object boundaries in u (i.e. to allow discontinuities in u), one should use $p = 1$. An adaptive combination of these semi-norms can be used to keep sharp edges while avoiding staircase effects in regions where the image varies smoothly. The fidelity term $\|t(u,f)\|_B \le \sigma$ forces u to be close to f. $t(u,f)$ is often chosen to be $f-u \equiv v$. The choice of a particular norm depends on the application. In image denoising, a common choice (known as the ROF model) is $\|t(u,f)\|_B = \|f-u\|_{L^2}$, which is small if $f-u$ is noise. The ROF model [36] by Rudin, Osher, and Fatemi was the first use of total variation regularization in image processing. In deblurring with denoising, for example, $\|t(u,f)\|_B = \|f-Au\|_{L^2}$ is commonly used, where A is a blurring operator. The pioneering ROF model led the way to a rich area of total variation-based image processing. See [17] for a survey.

If $\|s(u)\|_A$ and $\|t(u,f)\|_B$ are convex in u, the constrained minimization problem $\min\{\|s(u)\|_A$ s.t. $\|t(u)\|_B \le \sigma\}$ is equivalent to its Lagrangian form $\min\{\|s(u)\|_A + \lambda\|t(u,f)\|_B\}$, where λ is the *Lagrange multiplier* for the constraint $\|t(u)\|_B \le \sigma$. The two problems have the same solution if λ is chosen equal to the optimal value of the dual variable corresponding to the constraint in the constrained problem. Given σ or λ, one can calculate the other value by solving the corresponding problem.

1.3. The BV and G spaces and norms. We now formally define the Banach space BV of functions with bounded variation and the Banach space G, which is dual to a subspace of BV, and norms defined on these spaces. We provide these for the

completeness. They are relevant to some of the results in Section 7. Readers familiar with the theoretical foundation of total variation [6] can skip this subsection.

DEFINITION 1.1. *[42] Let $u \in L^1$, and define*

$$TV(u) := \sup\left\{ \int u \operatorname{div}(\vec{g})\, dx \ : \ \vec{g} \in C_0^1(\mathbb{R}^n;\mathbb{R}^n), |\vec{g}(x)|_{l^2} \le 1 \text{ for all } x \in \mathbb{R}^n \right\},$$

and $\|u\|_{BV} := \|u\|_{L^1} + TV(u)$, 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

$$BV = \left\{ u \in L^1 : \|u\|_{BV} < \infty \right\},$$

and is equipped with the $\|\cdot\|_{BV}$-norm. Moreover, $TV(u)$ is a semi-norm. $BV(\Omega)$ with Ω being a bounded open domain is defined analogously to BV with L^1 and $C_0^1(\mathbb{R}^n;\mathbb{R}^n)$ replaced by $L^1(\Omega)$ and $C_0^1(\Omega;\mathbb{R}^n)$, respectively.

If u is in the smaller Sobolev space $W^{1,1}(\mathbb{R}^n)$ and Du is the generalized derivative of u, then we have from the definition above, using integration by parts:

$$u \in BV \cap W^{1,1}(\mathbb{R}^n) \Leftrightarrow \sup_{\vec{g}\in C_0(\mathbb{R}^n;\mathbb{R}^n)\ |\vec{g}(x)|_{l^2}\le 1} \int Du \cdot \vec{g} \le \infty. \tag{1.4}$$

We can also see from (1.4) that each u defines a bounded linear functional $L_u(g)$ on $C_0(\mathbb{R}^n;\mathbb{R}^n)$ [11]. Using the Riesz representation theorem (also referred to as the Riesz-Markov theorem) on the isomorphism between the dual of $C_0(\mathbb{R}^n;\mathbb{R}^n)$ and the set of bounded vector Radon measures, we immediately have the following equivalent and often used definition:

$$BV = \{u : Du \text{ is a bounded Radon vector measure on } \mathbb{R}^n\}.$$

When Du is considered as a measure, $TV(u)$ over a set $\Omega \subseteq \mathbb{R}^n$ equals the total variation of Du as the Borel positive measure over Ω. This is given by

$$\|Du\|(\Omega) = \sup\left\{ \sum_{i=1}^{n} \|Du(E_i)\| : \bigcup_{i=1}^{n} E_i \subseteq \Omega,\ E_i\text{'s are disjoint Borel sets} \right\},$$

where the Borel sets are the σ-algebra generated by the open sets in $\mathbb{R}^n$. This is true because each $\vec{g} \in C_0(\mathbb{R}^n;\mathbb{R}^n)$ such that $\|\vec{g}\|_{l^2} \le 1$ is the limit of a series of $[-1,1]$-valued vector functions that are piecewise constant on Borel sets.

In the dual space of $C_0(\mathbb{R}^n;\mathbb{R}^n)$ we define weak-* convergence of Du_n to Dn as

$$\lim_{n\to\infty} \int_\Omega Du_n \cdot \vec{g} = \int_\Omega Du \cdot \vec{g},$$

for all $\vec{g} \in C_0(\mathbb{R}^n;\mathbb{R}^n)$.

Sets in $\mathbb{R}^n$ with finite perimeter are often referred to as BV sets. The perimeter of a set S is defined as follows:

$$\mathbf{Per}(S) := TV(\mathbf{1}_S) = \sup\left\{ \int_S \operatorname{div}(\vec{g}) dx : \vec{g} \in C_0^1(\mathbb{R}^n;\mathbb{R}^n), |\vec{g}(x)|_{l^2} \le 1, \forall\ x \in \mathbb{R}^n \right\}, \tag{1.5}$$

where $\mathbf{1}_S$ is the indicator function of S.

Next, we define the space G [29].

DEFINITION 1.2. *Let G denote the Banach space consisting of all generalized functions $v(x)$ defined on $\mathbb{R}^n$ that can be written as*

$$v = \operatorname{div}(\vec{g}), \quad \vec{g} = [g_i]_{i=1,\dots,n} \in L^\infty(\mathbb{R}^n;\mathbb{R}^n), \tag{1.6}$$

and equipped with the norm $\|v\|_G$ defined as the infimum of all L^∞ norms of the functions $|\vec{g}(x)|_{l^2}$ over all decompositions (1.6) of v. In short, $\|v\|_G = \inf\{\| \, |\vec{g}(x)|_{l^2} \, \|_{L^\infty} : v = \operatorname{div}(\vec{g})\}$. G is the dual of the closed subspace $\mathcal{BV}$ of BV, where $\mathcal{BV} := \{u \in BV : |Du| \in L^1\}$ [29]. We note that finite difference approximations to functions in BV and in $\mathcal{BV}$ are the same. For the definition and properties of $G(\Omega)$, see [10].

An immediate consequence of Definition 1.2 is that

$$\int u\, v = \int u \nabla \cdot \vec{g} = -\int Du \cdot \vec{g} \le \|Du\| \|v\|_G, \tag{1.7}$$

holds for any $u \in \mathcal{BV}$ with compact support and $v \in G$. We say (u, v) is an *extremal pair* if (1.7) holds with equality. This result was recently used by Kindermann, Osher and Xu [27] to recover f from an ROF generated v, which may have applications in image encryption.

The rest of the paper is organized as follows. In Section 2 we present a simple example to introduce some preliminaries and existing results on the TV-L^1 model. Section 3 is devoted to the monoticity property of the geometry problem $\min_{U\subset\mathbb{R}^2} \mathbf{Per}(U) + \lambda|S \triangle U|$, which serves as a basis for the rest of the paper. In Section 4 we use the results in Section 3 to construct the solution of the TV-L^1 model and discuss computational methods based on this construction. Sections 5 and 6 discuss feature selection and geometric and morphological invariance properties of the TV-L^1 model. In Section 7 we take a different analytic approach to establish the relationship between an approximate and the exact TV-L^1 model. Numerical results illustrating properties of the model are given in Section 8. Some technical results are given in Appendices A and B.

2. Preliminary properties of the TV-L^1 model. Let us first see how the TV-L^1 model (1.1) acts on the simple 2-dimensional image

$$f = c\mathbf{1}_{B_r(0)},$$

which has intensity $c > 0$ at all points (x_1, x_2) inside the disk $B_r(0) \equiv \{(x_1, x_2) : x_1^2 + x_2^2 \le r^2\}$ and intensity 0 everywhere else.

Chan and Esedoglu [16] showed that for this f, solving (1.1) gives

$$u_\lambda^* = \begin{cases} 0, & \text{if } 0 < \lambda < \frac{2}{r}, \\ c'f \text{ for any } c' \in [0,1], & \text{if } \lambda = \frac{2}{r}, \\ f, & \text{if } \lambda > \frac{2}{r}. \end{cases}$$

In general the minimizer of the TV-L^1 is nonunique. In the above disk example, if $\lambda = 2/r$, problem (1.1) has an infinite number of minimizers.

Two surprising points about this example are worth mentioning. First, the signal is never corrupted by the TV-L^1 decomposition: for all values of λ except $2/r$, where nonuniqueness occurs, the entire signal of f is either completely contained in u_λ^* or completely contained in v_λ^*. This is not case in the ROF (TV-L^2) decomposition.

Strong and Chan [39] showed that, for this f, the ROF model gives $u^*_\lambda = c'_\lambda f$, where c'_λ is a constant scalar lying in $[0, 1)$, never reaching 1 exactly. In other words, even if the input image f is completely noiseless, there does not exist a value of λ that gives an uncorrupted output in the ROF model. Meyer [29] characterized this phenomenon using the G-norm: if $\|f\|_G \geq \frac{1}{2\lambda}$, then the v^*_λ of the ROF model satisfies $\|v^*_\lambda\|_G = \frac{1}{2\lambda}$; otherwise, if $\|f\|_G < \frac{1}{2\lambda}$, the decomposition is meaningless since $u^*_\lambda \equiv 0$ and $v^*_\lambda = f$. Since v^*_λ corresponds to signal noise in the ROF model, Meyer's analysis indicates that the ROF model corrupts a noiseless input. However, it is desirable that noise-free images or image portions be kept invariant by a signal-noise decomposition. To achieve this, a remedy proposed by Osher et al in [34] for the ROF model iterates the decomposition using $f_k = f + v_{k-1}$. In contrast to the ROF model, the TV-L^1 model keeps the integrity of the signal in its decomposition.

The second surprising property of the TV-L^1 model exposed by this example is that the value of λ at which u^*_λ switches from 0 to f depends on the size of signal (i.e., the radius of the disk), and not on its intensity c. This is very surprising because the fidelity term $\|f - u\|_{L^1}$ clearly penalizes the intensity difference between f and u^*. For the aforementioned disk signal [16] or inputs like an annulus or a set of concentric annuli, and more general inputs with level sets contained inside a convex outer curve, one may derive analytic solutions using a symmetry argument (e.g., see Appendix A) or total variation flow (see [5, 12, 13, 8, 9] for extensive studies of the minimizers of $\min_{U \subset S} \mathbf{Per}(U) - \lambda|U|$ when S is convex or has an external boundary that is a convex curve using total variation flow); but how should one choose λ for an image that may contain signals of different scales and arbitrary shapes so as to isolate in u^*_λ certain of these signals? In Sections 5 and 6, we give a rigorous proof of this intensity-independent property of the TV-L^1 model for general image inputs. This property was observed by Chan and Esedoglu and mentioned in their work [16].

Even though Alliney and Nikolova did not explicitly draw these conclusions in an analytic way in their papers [2, 3, 4, 31, 32], they made related observations, and their successful attempts of applying the 1D/2D TV-L^1 to signal processing were based on these properties. Alliney studied the 1D and discrete version of the TV-L^1 energy and proved that his recursive median filter can construct u^*_λ directly. Many of his 1D results were later extended to 2D or higher dimensional spaces in [16]. Nikolova focused on the minimization of non-differentiable data fidelity terms, including the L^1 fidelity term, and presented impressive and successful applications of the TV-L^1 model to impulsive noise removal and outlier identification. She observed that u^*, the reconstructed image, was exact at some pixels and related this finding to the property of contrast-invariance. Later, Chan and Esedoglu [16] compared the continuum of ROF and TV-L^1 energies and studied the geometric properties of u^*_λ. Some of their results are quoted below.

PROPOSITION 2.1. *The TV-L^1 energy of a function f can be expressed as an integral of the geometry energies of the super level sets of f; i.e.,*

$$\mathbf{E}(u; \lambda, f) = \int_{-\infty}^{\infty} \mathbf{E_G}\left(L(u, \mu); \lambda, L(f, \mu)\right) \mathrm{d}\mu. \tag{2.1}$$

Formula (2.1), which is called the *Layer Cake* formula in [16], can be obtained by combining the co-area formula [24] $\int |\nabla u| = \int_{-\infty}^{\infty} \mathbf{Per}(L(u, \mu))\mathrm{d}\mu$ with the formula $\int |u - f|\mathrm{d}x = \int_{-\infty}^{\infty} |L(u, \mu) \triangle L(f, \mu)|\mathrm{d}\mu$.

PROPOSITION 2.2. *If the observed image is a characteristic function of a set S, i.e., $f = \mathbf{1}_S$, and $u^*_\lambda = \min_u \mathbf{E}(u; \lambda, \mathbf{1}_S)$, then $\mathbf{1}_{L(u^*_\lambda, \mu)}$ for any $\mu \in [0, 1)$ also*

minimizes $\mathbf{E}(u;\lambda,\mathbf{1}_S)$.

An equivalent of the above proposition is

PROPOSITION 2.3. *Any* $[0,1)$ *level set of the minimizer* u^*_λ *of* $\mathbf{E}(u;\lambda,\mathbf{1}_S)$ *solves the geometry problem*

$$\min_{U\subseteq\mathbb{R}^2} \mathbf{E_G}(U;\lambda,S). \tag{2.2}$$

Propositions 2.2 and 2.3 state that the non-convex geometry problem (2.2) can be converted into the TV-L^1 problem (1.1), which is a convex problem. This geometry problem finds applications in removing isolated binary noise and enhancing binary fax images. Given a binary signal S, one can solve the TV-L^1 problem with input $f=\mathbf{1}_S$. If the solution u^* is not binary, then one should examine its level sets.

In the rest of the paper, we exploit the converse of this process: given an observed image f, one can solve a series of geometry problems (2.2) and use the series of solutions, which are sets, to construct u^* explicitly.

3. The TV-L^1 geometry problem. In [16] Chan and Esedoglu raised the following question about the geometry problem (2.2):

> If $S_1\subset S_2$ and U_1^* and U_2^* are minimizers of the geometry problem (2.2) with inputs $S=S_1$ and $S=S_2$, respectively, is $U_1^*\subset U_2^*$?

While the absolute answer is "no", the answer is affirmative for a variant of the above question:

THEOREM 3.1. *Suppose that* $S_1\subset S_2$ *and* U_1^* *and* U_2^* *are minimizers of the geometry problem (2.2) with inputs* $S=S_1$ *and* $S=S_2$,

1. *if either* U_1^* *or* U_2^* *is a unique minimizer, then* $U_1^*\subseteq U_2^*$;
2. *otherwise, i.e., both problems with input* $S=S_1$ *and* $S=S_2$ *have multiple solutions, there exists a solution pair* $(\bar{U}_1^*,\bar{U}_2^*)$ *such that* $\bar{U}_1^*\subseteq\bar{U}_2^*$.

This section focuses on proving Theorem 3.1, which is used in the next section to construct a solution u^* of the TV-L^1 problem from a series of U_λ^*'s from input sets $S=L(f,\mu)$ for all values of μ.

Before we present this proof, let us see why the answer to Chan and Esedoglu's question is negative:

Example Let $\lambda=2/r$ and the input sets $S_1=B_r(0)$ and $S_2=B_r(0)\cup B_r(x)$ where x is a point distant from the origin 0. Clearly, $S_1\subset S_2$ strictly. However, there exist solutions

$$U_1^*=B_r(0)\quad\text{and}\quad U_2^*=\emptyset$$

of the geometry problem (2.2) for $S=S_1$ and $S=S_2$, respectively, where $U_1^*\supset U_2^*$ strictly. According to the results in [12, 13], both problems have multiple solutions. For $S=S_1$, the set of minimizers is $\{\emptyset,B_r(0)\}$ and, for $S=S_2$, this set is $\{\emptyset,B_r(0),B_r(x),B_r(0)\cup B_r(x)\}$.

Assumption A: In the rest of this section, we assumer U_1^* and U_2^* are minimizers ($U_1^*\subseteq U_2^*$ may not hold) of $\mathbf{E_G}(U;\lambda,S)$ with $S=S_1$ and $S=S_2$ for a fixed λ, respectively, and $S_1\subseteq S_2$. Since λ is fixed, we omit λ and write $\mathbf{E_G}(U;\lambda,S)$ as $\mathbf{E_G}(U;S)$. Moreover, we define $U_\wedge=U_1^*\cap U_2^*$ and $U_\vee=U_1^*\cup U_2^*$.

LEMMA 3.2 (Proposition 3.38 in [6]). *For two arbitrary sets* U_1 *and* U_2 *with finite perimeters, we have*

$$\mathbf{Per}(U_1)+\mathbf{Per}(U_2)\geq\mathbf{Per}(U_1\cap U_2)+\mathbf{Per}(U_1\cup U_2) \tag{3.1}$$

This property is also called the submodularity of **Per** functional. The following example shows that the above inequality can hold strictly.

Example Let U_1 be a square with opposite corners at $(0,0),(-1,1)$ and U_2 be another square with opposite corners at $(0,0),(1,1)$. U_1 has its entire right edge touching the entire left edge of U_2. According to the definition, $\mathbf{Per}(U_1) = \mathbf{Per}(U_2) = 4$, $\mathbf{Per}(U_1 \cap U_2) = 0$, and $\mathbf{Per}(U_1 \cup U_2) = 6$, where the third equation holds because $U_1 \cap U_2$ has measure 0 in $\mathbb{R}^2$ and hence is untestable by continuous functions.

In general, if two sets U_1 and U_2 share opposite edges for a strictly positive reduced length (a concept in geometric measure theory, see [6]), then this length is excluded from $\mathbf{Per}(U_1 \cap U_2)$ so (3.1) holds strictly. Lemma 3.2 is used in the proof of Lemma 3.4 below.

LEMMA 3.3. *Under Assumption A, the following inequalities hold:*

$$0 \geq \mathbf{E_G}(U_1^*; S_1) - \mathbf{E_G}(U_\wedge; S_1) \geq \mathbf{E_G}(U_1^*; S_2) - \mathbf{E_G}(U_\wedge; S_2). \tag{3.2}$$

Proof. The first inequality follows from the optimality of U_1^*. To prove the second inequality in (3.2), we expand its left-hand and right-hand sides:

$$\begin{aligned}
&\mathbf{E_G}(U_1^*; S_1) - \mathbf{E_G}(U_\wedge; S_1) \\
&= \mathbf{Per}(U_1^*) - \mathbf{Per}(U_1^* \cap U_2^*) && (3.3)\\
&+ \lambda(|U_1^* \backslash S_1| - |(U_1^* \cap U_2^*) \backslash S_1|) && (3.4)\\
&+ \lambda(|S_1 \backslash U_1^*| - |S_1 \backslash (U_1^* \cap U_2^*)|); && (3.5)\\
&\mathbf{E_G}(U_1^*; S_2) - \mathbf{E_G}(U_\wedge; S_2) \\
&= \mathbf{Per}(U_1^*) - \mathbf{Per}(U_1^* \cap U_2^*) && (3.6)\\
&+ \lambda(|U_1^* \backslash S_2| - |(U_1^* \cap U_2^*) \backslash S_2|) && (3.7)\\
&+ \lambda(|S_2 \backslash U_1^*| - |S_2 \backslash (U_1^* \cap U_2^*)|). && (3.8)
\end{aligned}$$

As (3.3) is identical to (3.6), we only need to prove (3.4) $\geq$ (3.7) and (3.5) $\geq$ (3.8). In fact,

$$\begin{aligned}
&|U_1^* \backslash S_1| - |(U_1^* \cap U_2^*) \backslash S_1| \\
&= |(U_1^* \backslash (U_1^* \cap U_2^*)) \backslash S_1| \\
&= |(U_1^* \setminus U_2^*) \backslash S_1| \\
&\geq |(U_1^* \setminus U_2^*) \backslash S_2| \quad (\because S_1 \subseteq S_2) \\
&= |(U_1^* \backslash (U_1^* \cap U_2^*)) \backslash S_2| \\
&= |U_1^* \backslash S_2| - |(U_1^* \cap U_2^*) \backslash S_2|,
\end{aligned}$$

and

$$\begin{aligned}
&|S_1 \backslash U_1^*| - |S_1 \backslash (U_1^* \cap U_2^*)| \\
&= |S_1 \cap \overline{U_1^*}| - |S_1 \cap \overline{(U_1^* \cap U_2^*)}| \\
&= -|S_1 \cap (\overline{(U_1^* \cap U_2^*)} - \overline{U_1^*})| \quad (\because \overline{U_1^*} \subseteq \overline{(U_1^* \cap U_2^*)}) \\
&\geq -|S_2 \cap (\overline{(U_1^* \cap U_2^*)} - \overline{U_1^*})| \quad (\because S_1 \subseteq S_2) \\
&= |S_2 \cap \overline{U_1^*}| - |S_2 \cap \overline{(U_1^* \cap U_2^*)}| \\
&= |S_2 \backslash U_1^*| - |S_2 \backslash (U_1^* \cap U_2^*)|.
\end{aligned}$$

□

LEMMA 3.4. *Under Assumption A, the following inequalities hold:*

$$\mathbf{E_G}(U_1^*;S_2)-\mathbf{E_G}(U_\wedge;S_2)\geq \mathbf{E_G}(U_\vee;S_2)-\mathbf{E_G}(U_2^*;S_2)\geq 0. \tag{3.9}$$

Proof. The second inequality follows directly from the optimality of U_2^* with respect to $S=S_2$. To prove the first inequality, let us expand $\mathbf{E_G}(U_\vee;S_2)-\mathbf{E_G}(U_2^*;S_2)$. (The other term $\mathbf{E_G}(U_1^*;S_2)-\mathbf{E_G}(U_\wedge;S_2)$ was expanded above in the proof of Lemma 3.3.)

$$\begin{aligned}
&\mathbf{E_G}(U_\vee^*;S_2)-\mathbf{E_G}(U_2^*;S_2)\\
&=\mathbf{Per}(U_1^*\cup U_2^*)-\mathbf{Per}(U_2^*) &(3.10)\\
&+\lambda(|(U_1^*\cup U_2^*)\backslash S_2|-|U_2^*\backslash S_2|) &(3.11)\\
&+\lambda(|S_2\backslash(U_1^*\cup U_2^*)|-|S_2\backslash U_2^*|). &(3.12)
\end{aligned}$$

We need to prove that (3.6), (3.7), and (3.8) are no less than (3.10), (3.11), and (3.12), respectively. Lemma 3.2 gives (3.6)≥(3.10). Moreover, we have (3.7) = (3.11) and (3.8) = (3.12) as proved below:

$$\begin{aligned}
&|U_1^*\backslash S_2|-|(U_1^*\cap U_2^*)\backslash S_2|=|((U_1^*\cap U_2^*)+(U_1^*\backslash U_2^*))\backslash S_2|-|(U_1^*\cap U_2^*)\backslash S_2|\\
&=|(U_1^*\backslash U_2^*)\backslash S_2|+|(U_1^*\cap U_2^*)\backslash S_2|-|(U_1^*\cap U_2^*)\backslash S_2| \quad (\because (U_1^*\cap U_2^*)\cap(U_1^*\backslash U_2^*)=\emptyset)\\
&=|(U_1^*\backslash U_2^*)\backslash S_2|+|U_2^*\backslash S_2|-|U_2^*\backslash S_2|\\
&=|(U_2^*+(U_1^*\backslash U_2^*))\backslash S_2|-|U_2^*\backslash S_2|=|(U_1^*\cup U_2^*)\backslash S_2|-|U_2^*\backslash S_2| \quad (\because (U_1^*\backslash U_2^*)\cap U_2^*=\emptyset).
\end{aligned}$$

and

$$\begin{aligned}
&|S_2\backslash U_1^*|-|S_2\backslash(U_1^*\cap U_2^*)|=|S_2\cap\overline{U_1^*}|-|S_2\cap\overline{(U_1^*\cap U_2^*)}|\\
&=-|S_2\cap(\overline{(U_1^*\cap U_2^*)}-\overline{U_1^*})|; \quad (\because \overline{U_1^*}\subseteq\overline{(U_1^*\cap U_2^*)})\\
&=-|S_2\cap(U_1^*\backslash U_2^*)|; \quad (\because U_1^*\backslash U_2^*=U_1^*-U_1^*\cap U_2^*=\overline{(U_1^*\cap U_2^*)}-\overline{U_1^*})\\
&=-|S_2\cap(\overline{U_2^*}\backslash\overline{U_1^*})|\\
&=-|S_2\cap(\overline{U_2^*}-(\overline{U_2^*}\cap\overline{U_1^*})|\\
&=|S_2\cap\overline{(U_2^*\cup U_1^*)}|-|S_2\cap\overline{U_2^*}|; \quad (\because \overline{U_2^*}\supseteq\overline{U_2^*}\cap\overline{U_1^*}=\overline{(U_2^*\cup U_1^*)})\\
&=|S_2\cap\overline{(U_2^*+U_1^*\backslash U_2^*)}|-|S_2\cap\overline{U_2^*}|\\
&=|S_2\backslash(U_2^*+U_1^*\backslash U_2^*)|-|S_2\backslash U_2^*|=|S_2\backslash(U_1^*\cup U_2^*)|-|S_2\backslash U_2^*|.
\end{aligned}$$

□

Proof. (of Theorem 3.1) Concatenating the inequalities in Lemmas 3.3 and 3.4 gives us

$$\begin{aligned}
0\geq\ &\mathbf{E_G}(U_1^*;S_1)-\mathbf{E_G}(U_\wedge;S_1)\\
\geq\ &\mathbf{E_G}(U_1^*;S_2)-\mathbf{E_G}(U_\wedge;S_2)\\
\geq\ &\mathbf{E_G}(U_\vee;S_2)-\mathbf{E_G}(U_2^*;S_2)\\
\geq\ &0.
\end{aligned} \tag{3.13}$$

Therefore, all inequalities above hold as equalities. It follows that $U_\wedge$ and $U_\vee$ minimize $\mathbf{E_G}(U;S_1)$ and $\mathbf{E_G}(U;S_2)$, respectively. We have found a solution pair $(\bar{U}_1^*,\bar{U}_2^*):=(U_\wedge,U_\vee)$ such that $\bar{U}_1^*\subseteq\bar{U}_2^*$ □

COROLLARY 3.5. *Under Assumption A, $U_\wedge$ minimizes $\mathbf{E_G}(U;S_1)$; $U_\vee$ minimizes $\mathbf{E_G}(U;S_2)$.*

The following two corollaries extend the above geometric results to the minimizers of the TV-L^1 energy $\mathbf{E}$.

COROLLARY 3.6. *Let u_1^* and u_2^* minimizes the TV-L^1 energies $\mathbf{E}(u;\lambda,f_1)$ and $\mathbf{E}(u;\lambda,f_2)$, respectively, where $f_1 \geq f_2$ point-wise. We have*

1. *if either u_1^* or u_2^* is a unique minimizer, then $u_1^* \leq u_2^*$;*
2. *otherwise, i.e., both problems with inputs $f=f_1$ and $f=f_2$ have nonunique solutions, then there exists a solution pair $(\bar{u}_1^*, \bar{u}_2^*)$ such that $\bar{u}_1^* \leq \bar{u}_2^*$.*

COROLLARY 3.7. *Under the same assumption as in the above corollary, $\min(u_1^*,u_2^*)$ and $\max(u_1^*,u_2^*)$ minimize $\mathbf{E}(u;\lambda,f_1)$ and $\mathbf{E}(u;\lambda,f_2)$, respectively.*

The above two corollaries can be proved by first using the Layer Cake formula (2.1) to express the TV-L^1 energy as an integral of geometry energies over all super level sets and then applying the results of Theorem 3.1 and Corollary 3.5 to the geometry energy at each level. We leave these proofs to the reader.

4. Constructing u^* from U_μ^*. In this section we show how Theorem 3.1 can be applied to construct a minimizer u^* of $\mathbf{E}(u;\lambda,f)$ from minimizers of $\mathbf{E_G}(U;\lambda,S)$. Given an observed image f, we let the inputs S be the level sets of f, i.e., $S=L(f,\mu)$, which are contained inside one another as μ is increased, i.e., $L(f,\mu_2)\subseteq L(f,\mu_1)$ if $\mu_2>\mu_1$. From Theorem 3.1, the following collections of sequences of minimizers U_μ^* of $\mathbf{E_G}(U;\lambda,L(f,\mu))$ is well-defined and nonempty:

$$\mathbb{U}^*(f) := \left\{ \{U_\mu^*\}_{\mu\in\mathbb{R}} : \begin{array}{l} \forall \mu\in\mathbb{R},\ U_\mu^* \text{ is a minimizer of } \mathbf{E_G}(U;\lambda,L(f,\mu)), \\ \text{and } U_{\mu_2}^* \subseteq U_{\mu_1}^* \text{ if } \mu_2>\mu_1. \end{array} \right\} \tag{4.1}$$

Let $\{U_\mu^*\}_{\mu\in\mathbb{R}} \in \mathbb{U}^*(f)$, i.e., $\{U_\mu^*\}_{\mu\in\mathbb{R}}$ is a specific sequence of sets in the collection $\mathbb{U}^*(f)$, in which the minimizers U_μ^* of $\mathbf{E_G}(U;\lambda,L(f,\mu))$ are contained inside one another as μ is increased. We define $\bar{u}$ pointwise by

$$\bar{u}(x) = \sup\{\mu : x\in U_\mu^*\}. \tag{4.2}$$

A geometrically clearer version of the above construction is the following: let

$$\begin{array}{l} f = f^+ - f^-, \quad (f^+, f^- \geq 0) \\ \{U_\mu^+\}_{\mu\in\mathbb{R}} \in \mathbb{U}^*(f^+) \text{ and } \{U_\mu^-\}_{\mu\in\mathbb{R}} \in \mathbb{U}^*(f^-), \end{array}$$

then we have

$$\bar{u}(x) = \underbrace{\int_0^\infty \mathbf{1}_{U_\mu^+}\mathrm{d}\mu}_{u^+} - \underbrace{\int_0^\infty \mathbf{1}_{U_\mu^-}\mathrm{d}\mu}_{u^-}, \tag{4.3}$$

where u^+ and u^- are built up by stacking the shrinking (and nested) U_μ^+'s and U_μ^-'s. One can easily check that (4.2) and (4.3) are equivalent. Darbon and Sigelle [22] independently obtained a similar result for their discretized version of a variant of the TV-L^1 problem, which is a linear program.

Below is the main result of this section based on the above construction.

THEOREM 4.1. *The function $\bar{u}(x)$ given by (4.2) minimizes $\mathbf{E}(u;\lambda,f)$ where $\{U_\mu^*\}_{\mu\in\mathbb{R}}$ is any sequence from the collection of sequences $\mathbb{U}^*(f)$ given by (4.1).*

Proof. By the construction formula (4.2), the level set of $\bar{u}$

$$L(\bar{u},\bar{\mu}) = \{x : \bar{u}(x)\geq\bar{\mu}\} = \{x : \exists \mu, \ni x\in U_\mu^* \text{ and } \mu\geq\bar{\mu}\} \tag{4.4}$$

Since $U^*_\mu \subseteq U^*_{\bar{\mu}}$ if $\mu > \bar{\mu}$, the condition of the existence of a μ that gives both $x \in U^*_\mu$ and $\mu \geq \bar{\mu}$ in the right-hand side of (4.4) holds if and only if $x \in U^*_{\bar{\mu}}$. It follows that

$$L(\bar{u}, \mu) = U^*_\mu. \tag{4.5}$$

Since each $L(\bar{u}, \mu)$ minimizes the integrand of the integral

$$\int_{-\infty}^{\infty} \mathbf{E_G}(U; \lambda, L(f, \mu)) \mathrm{d}\mu \tag{4.6}$$

over all levels μ, the whole integral is minimized and is equivalent to $\mathbf{E}(\bar{u}; \lambda, f)$ by the Layer Cake formula (2.1). □

In Section 5 below, Theorem 4.1 is used to characterize those image features that appear in u^* and those that end up in $v^* = f - u^*$.

In the light of this theorem, one may wonder whether this equivalence between the TV-L^1 model (1.1) and a series of geometry problems makes it easier (or at least provides a more geometric way) to find a solution. There is both bad and good news. In theory, the number of different geometry problems we need to solve is infinite while, in practice, when processing computerized images this number is finite and is limited by the number of grey-scale levels (typically 256 or 2^{16}). However, obtaining a solution of (2.2) is non-trivial because it is nonconvex and has nonunique solutions in general. Let us again examine the disk example $S = B_r(0)$ with radius r for an illustration of solution nonuniqueness: if $S = B_r(0)$ and $\lambda = 2/r$, then both $U = B_r(0)$ and $U = \emptyset$ minimize the geometry energy while any other sets, especially those $U' = B_{\bar{r}}(0)$ satisfying $0 < \bar{r} < r$, are not minimizers; if $0 < \lambda < 2/r$, then $U = B_r(0)$ is a local, non-global, minimizer in then sense that $U = B_{r\pm\epsilon}(0)$ for any ϵ small gives higher energies (so $U = B_r(0)$ is locally minimal) but $U = \emptyset$ is the unique global minimum. This suggests that a global minimization algorithm for solving (2.2) may have to examine a large number of sets before restricting its search locally. The recent algorithm by Darbon and Sigelle [22] based on sampling a 256-level Markov Random Field on grey-scale images falls into this category. They used the Layer Cake formula and associated a Markov random field with each level set of an image with 256 grey-scale levels. They were able to reformulate and decompose the geometry energy (2.2) as conditional posterior energies at each pixel (i.e., at each site in the Markov random field) and thus defined a Markov random field. To optimize this Markov random field, i.e., to find a lowest energy configuration, they used a min-cut algorithm [28], which finds the global minimizer in polynomial time. This is in fact not surprising despite the fact that the geometry problem may have strictly local solutions. Though the min-cut algorithm is an 0-1 combinatorial program, its dual is the max flow problem, which has long been known to have polynomial-time algorithms and give a min-cut, as the dual solution, upon termination. This link between a TV variational problem and a discrete network algorithm is very interesting.

After our work was first submitted, an anonymous referee brought to our attention the recent work [15] by A. Chambolle, which exploits the connection between the ROF model and a series of decoupled geometry problems to solve the ROF model using a graph-cut based algorithm [28] that is essentially identical to the one proposed in [22].

5. Feature selections. In the previous sections, the parameter λ in the TV-L^1 model (1.1) was fixed. In this section, however, we vary λ and relate it to the scales of the features in u^* and v^*. It is well know that Meyer's G-norm (Def. 1.2) is a good

measure of the oscillation of functions [29]. Using G-value [37] defined below, which is an extension of Meyer's G-norm proposed by Scherzer et al, we are able to fully characterize v^* for a given parameter λ. To emphasize the role of λ in determining u^*, v^*, U^*_μ, and V^*_μ, we add λ as a subscript to these quantities (i.e., we write u^*_λ, v^*_λ, $U^*_{\lambda,\mu}$, and $V^*_{\lambda,\mu}$) in this section.

In general, the TV-L^1 model, using a particular value of the parameter λ, returns an image combining many features. In certain applications one is interested in extracting small and/or large-scale features in an image. Therefore, we are interested in determining a λ that gives all targeted features with the least number of unwanted features in the output. Below we show how to choose an appropriate λ that will allow us to extract geometric features of a given scale, measured by the G-values of their level sets.

DEFINITION 5.1. ***(G-value)****[37] Let $\Psi : \mathbb{R}^2 \to 2^{\mathbb{R}}$ be a set-valued function (also called a multifunction and a set-valued map) 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 between Ψ being a set-valued function and a set of measurable (single-valued) functions, and let*

$$\Psi := \{\psi \,:\, \psi : \mathbb{R}^2 \to \mathbb{R} \text{ is measurable and } \psi(x) \in \Psi(x), \forall\, x\}.$$

The G-value of Ψ is defined as follows:

$$G(\Psi) := \sup_{h \in C_0^\infty : \int |\nabla h| = 1} \inf_{\psi \in \Psi} \int \psi(x) h(x) \mathrm{d}x. \tag{5.1}$$

An easy way to understand the above definition is to compare the definitions of G-value and G-norm. Since the G-norm of a function ψ can be defined as

$$G(\{\psi\}) = \sup_{h \in C_0^\infty : \int |\nabla h| = 1} \int \psi(x) h(x) \mathrm{d}x = \|\psi\|_G, \tag{5.2}$$

where a single-function set $\{\psi\}$ replaces Ψ in (5.1), the G-value can be viewed as an extension of the G-norm to set-valued functions. In [37] Scherzer et al applied the G-value to the sub-differential of the absolute value of f, $\partial|f|$, and the *Slope* from [7] of f to determine when u^*_λ or v^*_λ vanishes:

THEOREM 5.2. *[37] Let $\partial|f|$ denote the set-valued sub-differential of $|f|$, i.e.,*

$$\partial|f(x)| = \begin{cases} \{\mathrm{sign}(f(x))\} & f(x) \neq 0 \\ [-1, 1] & f(x) = 0. \end{cases} \tag{5.3}$$

Then, for the TV-L^1 problem (1.1)

1. *$u^*_\lambda = 0$ ($v^*_\lambda = f$) is an optimal solution if and only if*

$$\lambda \leq \frac{1}{G(\partial|f|)};$$

2. *$u^*_\lambda = f$ ($v^*_\lambda = 0$) is an optimal solution if and only if*

$$\lambda \geq \sup_{h \in BV} \frac{\|Df\| - \|Dh\|}{\int |f - h|}.$$

Instead of directly applying Theorem 5.2 to the input f, we apply it to the characteristic functions of the level sets of f. We easily have the following results as a corollary of Theorem 5.2:

COROLLARY 5.3. *For the geometric problem (2.2) with a given λ,*

1. *$U_\lambda^* = \emptyset$ ($V_\lambda^* = S$) is an optimal solution if and only if*

$$\lambda \leq \frac{1}{G(\partial|\mathbf{1}_S|)};$$

2. *$U_\lambda^* = S$ ($V_\lambda^* = \emptyset$) is an optimal solution if and only if*

$$\lambda \geq \sup_{h\in BV} \frac{\|D\mathbf{1}_S\| - \|Dh\|}{\int |\mathbf{1}_S - h|}.$$

The above corollary characterizes $U_{\lambda,\mu}^*$ in (4.2) for given level μ and scalar λ. Suppose that the set S of a geometric feature coincides with $L(f,\mu)$ for $\mu \in [\mu_0, \mu_1)$. Then, for any $\lambda < 1/G(\partial|\mathbf{1}_S|)$, S is not observable in u_λ^*. This is because $1/G(\partial|1_{L(f,\mu)}|)$ is increasing in μ and therefore, for $\mu \geq \mu_0$, $U_{\lambda,\mu}^*$ vanishes. On the other hand, once $\lambda \geq 1/G(\partial|\mathbf{1}_S|)$, according the above corollary, $\mathbf{1}_{U_{\lambda,\mu}^*} \not\equiv 0$ for $\mu \in [\mu_0, \mu_1)$, which implies that at least some part of S can be observed in u_λ^*. If λ is increased further such that $\lambda \geq \sup_{h\in BV}(\|D\mathbf{1}_S\| - \|Dh\|)/\int |\mathbf{1}_S - h|$, we get $U_{\lambda,\mu}^* = L(f,\mu) = S$ for all $\mu \in [\mu_0, \mu_1)$ and therefore, the feature is fully contained in u_λ^*, which is given by (4.2). 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 set in $L(f,\mu)$, for some μ, which also contains sets representing other features. Consequently, u_λ^* that contains a targeted feature may also contain many other features. However, from Theorem 4.1 and Corollary 5.3, we can easily see that the arguments for the case $S = L(f,\mu)$ still hold for the case $S \subset L(f,\mu)$.

Suppose there is a sequence of features in f that are represented by sets $S_1, S_2, \ldots, S_l$ and have distinct intensity values. Let

$$\lambda_i^{\min} := \frac{1}{G(\partial|\mathbf{1}_{S_i}|)},\ \lambda_i^{\max} := \sup_{h\in BV} \frac{\|D\mathbf{1}_{S_i}\| - \|Dh\|}{\int |\mathbf{1}_{S_i} - h|}, \tag{5.4}$$

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

$$\lambda_1^{\min} \leq \lambda_1^{\max} < \lambda_2^{\min} \leq \lambda_2^{\max} < \ldots < \lambda_l^{\min} \leq \lambda_l^{\max}, \tag{5.5}$$

then feature i, for $i = 1, \ldots, l$, can be precisely retrieved as $u_{\lambda_i^{\max}+\epsilon}^* - u_{\lambda_i^{\min}-\epsilon}^*$ (here ϵ is a small scalar that forces unique solutions because $\lambda_i^{\min} = \lambda_i^{\max}$ is allowed). This is true since for $\lambda = \lambda_i^{\min} - \epsilon$, feature i completely vanishes in u_λ^*, but for $\lambda = \lambda_i^{\max} + \epsilon$, feature i is fully contained in u_λ^* while there is no change to any other features.

6. Geometric and morphological invariance. Based on the results and discussion in the previous sections, we now show that the TV-L^1 model is invariant under certain geometric and morphological transformations.

PROPOSITION 6.1. *(Geometric invariance) The TV-L^1 decomposition, defined as the operator T_t, is invariant under the following geometric transformations:*

1. *Translation τ (shifting the coordinates by a constant): $T_t \circ \tau = \tau \circ T_t$;*
2. *Isometry transformation R: $T_t \circ R = R \circ T_t$;*
3. *Scaling $\rho > 0$: $T_{t\rho} \circ \rho = \rho \circ T_t$, where $\rho \cdot x = \rho x$.*

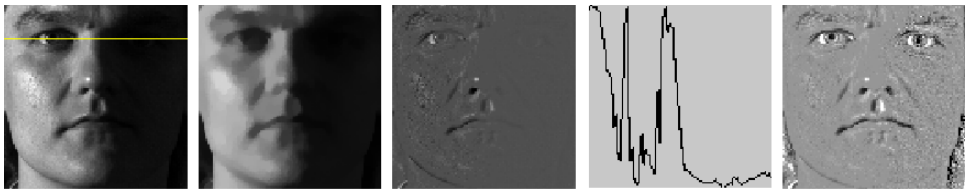

FIG. 6.1. *The LTV model. From left to right: f, u^*, v^*, $f(20,:)$, $v' = f/u^*$. $f(20,:)$ is the signal along the line depicted in f.*

Proof. The proof is simple since it is sufficient to provide a proof for the geometry problem (2.2). The geometry formulation (2.2) has the straightforward properties of translational and isometric (e.g., rotational) invariance. If the n-dimensional geometry sets U and S in (2.2) are both uniformly scaled by a constant ρ, then $\mathbf{Per}(U^\rho)$ and $|U^\rho \triangle S^\rho|$ equal $\rho^{n-1}\mathbf{Per}(U)$ and $\rho^n|U \triangle S|$, respectively. One can scale λ in (2.2) by $1/\rho$ and obtain an energy homogeneously scaled by ρ^{n-1}:

$$\rho^{n-1}\left(\mathbf{Per}(U) + \lambda|U \triangle S|\right) = \mathbf{Per}(U^\rho) + \lambda/\rho|U^\rho \triangle S^\rho|.$$

By noticing $1/(t\rho) = \lambda/\rho$, therefore, $T_{t\rho} \circ \rho \cdot f = (u^*)^\rho = \rho \circ T_t \cdot f$ for any f. □

PROPOSITION 6.2. *(Morphological invariance) Let C be a constant scalar and g be an increasing real function. Then,*

$$T_t \cdot (f + C) = (T_t \cdot f) + C, \quad T_t \circ g = g \circ T_t. \tag{6.1}$$

Proof. These results are simple consequences of Theorem 4.1, which states that minimizing $\mathbf{E}(u; \lambda, f)$ can be decoupled into independent minimizations of $\mathbf{E_G}(U; \lambda, L(F, \mu))$ over a range of values of μ. □

Since log is an increasing function, we have the following from Proposition 6.2:
Example Suppose $f > 1$ strictly, then $T_t \cdot \log f = \log(T_t \cdot f) = \log u^*$. Let $\bar{f} = \log f$ and $\bar{u} = T_t \cdot \bar{f}$, then

$$\frac{f}{u^*} = e^{\log f - \log u^*} = e^{\bar{f} - \bar{u}}. \tag{6.2}$$

This means that the following two processes are equivalent:
(1) Take f, obtain u^* by applying the TV-L^1 decomposition to f, and get $v' = \frac{f}{u^*}$;
(2) Take the logarithm $\bar{f}$ of f, obtain $\bar{v} = \bar{f} - \bar{u}$ by applying the TV-L^1 decomposition to $\bar{f}$, and get the same $v' = \exp(\bar{v})$.
This example was explored by Chen et al [20] who proposed the Logarithm TV (LTV) model for preprocessing face images to correct varying illuminations prior to automated face recognition. In a face image, one half of the face can look brighter than the other half if the light shines on the face from one side. An extreme case occurs when a point light source is located exactly to the left of the face. In such a case the right half of the face only receives a very little amount of ambient light, resulting in face images with very unbalanced brightness and contrast. In Figure 6.1 we present a face f to illustrate this issue. Geometric information intrinsic to the face must be extracted for use by distance-based algorithms for comparing an inquiry image with reference images. It is well known in signal processing that the logarithm function, which is steep near 1 and flat near ∞, can be applied to a grey-scale image f to enhance

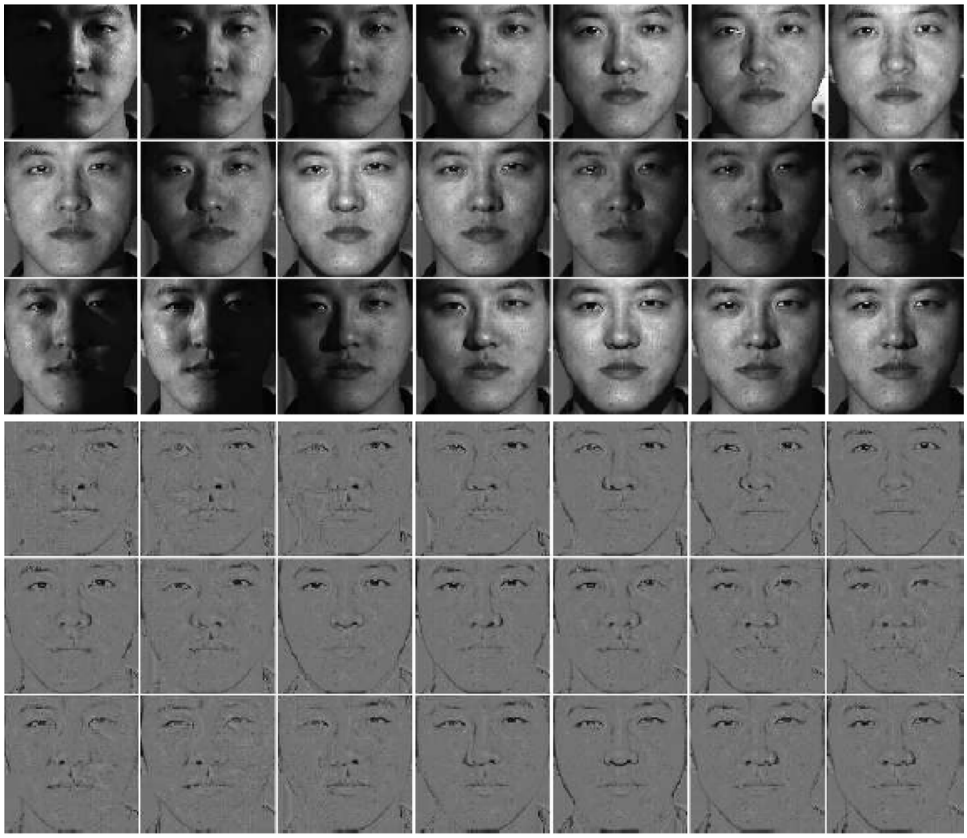

FIG. 6.2. *The LTV model. First 3 rows: input face images f; Last 3 rows: illumination-free output images $v' = f/u^*$ for automated face recognition.*

the contrast of its low-light range signal, corresponding to the range of small intensity values. Therefore, the authors applied the TV-L^1 model to $\log f$ and examined the small-scale output $\bar{v}$ and its restored signal $\exp(\bar{v})$. They found that $v' = \exp(\bar{v})$ (Figure 6.2) contains signal of small-scale facial features that does not vary too much among the images of the same face under different illuminations. Their experiments based on angle testing and principal component analysis (PCA) then proved that $v' = \exp(\bar{v})$ has illumination-free signal that is an ideal element for distance measure. Specifically, they treat two $\exp(\bar{v})$'s as two vectors w_1 and w_2 and measure their distance by the cosine of the angle between them (i.e., $\langle w_1, w_2\rangle/(\|w_1\|\,\|w_2\|)$). Their analysis in [20] provides an explanation for LTV's excellent performance that is based on the relationship between $v' = f/u^*$ and a multiplicative light reflection model.

7. Smooth approximation of the TV-L^1 model. In contrast with the previous sections where we followed a geometric approach, in this section we analytically study the approximate TV-L^1 energy

$$\mathbf{E}_\varepsilon(u;\lambda,f) = \int_\Omega |\nabla u| + \lambda \int_\Omega \sqrt{(f-u)^2+\varepsilon}, \tag{7.1}$$

defined in a bounded convex open set Ω with a C^1 boundary (typically rectangular). We let u_ε^* denote the unique minimizer of $\mathbf{E}_\varepsilon(u;\lambda,f)$ and $v_\varepsilon^* = f - u_\varepsilon^*$. Because the

L^1 energy is convex but nonsmooth, PDE-based iterative methods [36, 14] inevitably employ a smoothing regularization like (7.1). For large-scale and nondecomposible problems, such as those arise in processing 3D and 4D medical images, this type of method is the only one that does not exceed memory limits. Therefore, it is important to understand the behavior of the approximation (7.1). Below we characterize the G-norm of the minimizers of (7.1), which allows us to compare (7.1) with the ROF model [36] and view the results in the previous sections from a different perspective. We note that the TV term is also not smooth and its approximation is discussed in [1].

In our proof below of Theorem 7.5, which states the convergence of the solutions of the perturbed TV-L^1 model (7.1) to the solution of the TV-L^1 model, we need the following three known results.

LEMMA 7.1 (General Minimax Theorem [23, 38]). *Let K be a compact convex subset of a Hausdorff topological vector space X, C be a convex subset of a vector space Y, and f be a real-valued functional defined on $K \times C$ which is (1) convex and lower-semicontinuous in x for each y, and (2) concave in y for each x. Then*

$$\inf_{x\in K}\sup_{y\in C} f(x,y) = \sup_{y\in C}\inf_{x\in K} f(x,y).$$

LEMMA 7.2 (See [35, 42] for related results). *If a sequence $\{u_i\}_{i\in\mathbb{N}}$ defined in BV satisfies $\sup_i \|Du_i\| < +\infty$, then it has a subsequence that weakly converges in both $L^{n/(n-1)}$ and BV to $u \in BV$. Moreover, weak lower semi-continuity holds for this sequence:*

$$\|Du\| \le \liminf_i \|Du_i\|.$$

This lemma shows that the BV space has the so-called *relatively weakly compact* property. For L^p $(1 \le p < \frac{n}{n-1})$, the lemma below gives a stronger result:

LEMMA 7.3 ((See [1, 24])). *Let $\mathcal{S}$ be a BV-bounded set of functions defined on a bounded open domain Ω. Then $\mathcal{S}$ is relatively compact in L^p for $1 \le p < n/(n-1)$.*

We also need the following technical lemma.

LEMMA 7.4. *The sets $G_0 := \{v : v = \mathrm{div}(\vec{g}),\ \vec{g} \in C_0^1(\Omega;\mathbb{R}^n),\ \| \, |\vec{g}(x)|_{l^2} \, \|_{L^\infty} \le 1 \}$ and $BV_0 := \{u \in L^1(\Omega) : \|Du\| \le R, \|u\|_{L^1} \le \|f\|_{L^1}\} \subset BV(\Omega)$, where R is given, are convex. Moreover, BV_0 is compact in L^1.*

Proof. Suppose v_g and v_h are in G_0. There exist $\vec{g}, \vec{h} \in C_0^1(\Omega;\mathbb{R}^n)$ satisfying

$$v_g = \mathrm{div}(\vec{g}), \quad v_h = \mathrm{div}(\vec{h}), \quad \| \, |\vec{g}|_{l^2} \, \|_{L^\infty} \le 1, \quad \| \, |\vec{h}|_{l^2} \, \|_{L^\infty} \le 1.$$

For any $\lambda \in [0,1]$, we have (Minkowski inequality)

$$\begin{aligned} \| \, |\lambda\vec{g} + (1-\lambda)\vec{h}|_{l^2} \, \|_{L^\infty} &\le \| \, \lambda|\vec{g}|_{l^2} + (1-\lambda)|\vec{h}|_{l^2} \, \|_{L^\infty} \\ &\le \lambda \| \, |\vec{g}|_{l^2} \, \|_{L^\infty} + (1-\lambda)\| \, |\vec{h}|_{l^2} \, \|_{L^\infty} \\ &\le \lambda + (1-\lambda) = 1. \end{aligned}$$

This means $\|\lambda v_g + (1-\lambda)v_h\|_G \le 1$; consequently, $\lambda v_g + (1-\lambda)v_h \in G_0$, i.e., G_0 is convex. The convexity of BV_0 can be proved analogously from its definition. The compactness of BV_0 in L^1 is a direct result of Lemma 7.3. □

THEOREM 7.5. *The solution $u_\varepsilon^* \in BV(\Omega)$ of the approximate TV-L^1 model (7.1) satisfies*

$$\|\mathrm{sign}_\varepsilon(v_\varepsilon^*)\|_G \le 1/\lambda,$$

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

A proof for more general cases can be found in [35]. We give a short proof below based on Lemma 7.1. A similar approach is also used in [26] to derive the G-norm related properties for the ROF model [36].

Proof. Let R (in the definition of BV_0 in Lemma 7.4) be large enough, and consider the functional $L : BV_0 \times G_0 \to \mathbb{R}$:

$$L_\varepsilon(u,w) = \int_\Omega uw + \lambda\sqrt{(f-u)^2+\varepsilon}.$$

Define $P_\varepsilon(u) = \sup_{w\in G_0} L_\varepsilon(u,w)$ and $D_\varepsilon(w) = \inf_{u\in BV_0} L_\varepsilon(u,w)$. $L_\varepsilon(u,w)$ is convex and lower semi-continuous in u, and is linear (hence concave) in w.

Since G_0 is complete w.r.t. $\|\cdot\|_G$, there exists an optimal $w_\varepsilon^*(u)$ satisfying $P_\varepsilon(u) = L_\varepsilon(u, w_\varepsilon^*(u))$ for each $u \in BV_0$. On the other hand, by applying Lemmas 7.2 and 7.3, we have an optimal $u_\varepsilon^* \in BV_0$ that minimizes $P_\varepsilon(u)$.

The obtainability of optimizers and Lemma 7.4 allow us to apply Lemma 7.1 to $L_\varepsilon(u,w)$: there exists an optimal solution pair $(u_\varepsilon^*, w_\varepsilon^*) \in BV_0 \times G_0$ such that

$$D_\varepsilon(w_\varepsilon^*) = L_\varepsilon(u_\varepsilon^*, w_\varepsilon^*) = P_\varepsilon(u_\varepsilon^*). \tag{7.2}$$

The first equation in (7.2) indicates $\partial L_\varepsilon(u, w_\varepsilon^*)/\partial u|_{u=u_\varepsilon^*} = 0$, and this gives

$$w_\varepsilon^* = \lambda \frac{v_\varepsilon^*}{\sqrt{{v_\varepsilon^*}^2+\varepsilon}}, \tag{7.3}$$

where $v_\varepsilon^* = f - u_\varepsilon^*$. Therefore, $\|\text{sign}_\varepsilon(v_\varepsilon^*)\|_G \le 1/\lambda$. □

COROLLARY 7.6. *If* $\|\text{sign}_\varepsilon(f)\|_G \le 1/\lambda$, $u_{\lambda,\varepsilon} \equiv 0$ *minimizes* $\mathbf{E}_\varepsilon(u;\lambda,f)$.

Proof. Let $w_\varepsilon^* \equiv \lambda\,\text{sign}_\varepsilon(f)$ and $v_\varepsilon^* \equiv f$. Noting that $u_\varepsilon^* \equiv 0$, we have

$$D_\varepsilon(w_\varepsilon^*) = L_\varepsilon(u_\varepsilon^*, w_\varepsilon^*) = P_\varepsilon(u_\varepsilon^*).$$

The corollary then follows from the optimality of the saddle point $(u_\varepsilon^*, w_\varepsilon^*)$. □

COROLLARY 7.7. *If* $\|\text{sign}_\varepsilon(f)\|_G > 1/\lambda$, *then there exists an optimal solution* u_ε^* *satisfying*

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

Proof. Since $\|\text{sign}_\varepsilon(f)\|_G > 1/\lambda$ but $\|\text{sign}_\varepsilon(v_\varepsilon^*)\|_G \le 1/\lambda$ by Theorem 7.5, we must have $u_\varepsilon^* \not\equiv 0$. Then, we have $\|w_\varepsilon^*\|_G = 1$ from the second equation in (7.2). It follows from (7.3) that $\|\text{sign}_\varepsilon(v_\varepsilon^*)\|_G = 1/\lambda$. The second result of Corollary 7.7 follows from the equations $\int_\Omega u_\varepsilon^* w_\varepsilon^* = \sup_{w\in G_0}\int_\Omega u_\varepsilon^* w = \|Du_\varepsilon^*\|$ and $w_\varepsilon^* = \lambda\,\text{sign}_\varepsilon(v_\varepsilon^*)$. □

According to Theorem 7.5 and its corollaries, the approximate TV-L^1 model performs a soft thresholding on $\|\text{sign}_\varepsilon(f)\|_G$. If this value is bigger than $1/\lambda$, a part of the signal f, v_ϵ^* with $\|\text{sign}_\varepsilon(v_\varepsilon^*)\|_G = 1/\lambda$, is removed from f; if less or equal to $1/\lambda$, the whole signal $v_\varepsilon^* \equiv f$ is removed. This does not contradict the behavior of the exact TV-L^1 model. As $\varepsilon \to 0$, $\text{sign}_\varepsilon(f)$ converges to the characteristic function of the support of f. Therefore, the thresholding depends mainly on the shape of f rather than its value. The smaller that ε is, the less the value f affects the thresholding. This matches the results based on the G-value in Section 5. We point out that the need for the G-value in the analysis of the exact model, as an extension of the G-norm

to set-valued functions, is related to the fact that $\text{sign}_\varepsilon(\cdot)$ does not converge to the signum function $\text{sign}(\cdot)$ uniformly. Instead, we can think that $\text{sign}_\varepsilon(\cdot)$ "converges" to

$$\overline{\text{sign}} : \mathbb{R} \mapsto 2^{\mathbb{R}}, \quad \overline{\text{sign}}(x) = \begin{cases} \{\text{sign}(x)\}, & \text{if } x \neq 0; \\ [-1, 1], & \text{if } x = 0. \end{cases}$$

Finally, we show that the solution of the approximate model converges in L^1 to that of the exact model.

THEOREM 7.8. *Assume u_ε^* minimizes $\mathbf{E}_\varepsilon(u; \lambda, f)$. There exists a minimizer $\bar{u}$ of $\mathbf{E}(u; \lambda, f)$ such that*

$$\lim_{\varepsilon \downarrow 0+} \|u_\varepsilon^* - \bar{u}\|_{L^1} = 0, \quad \lim_{\varepsilon \downarrow 0+} \|v_\varepsilon^* - \bar{v}\|_{L^1} = 0.$$

Proof. Noticing that $\sqrt{t+\varepsilon} \leq \sqrt{t} + \sqrt{\varepsilon}$, for $t, \varepsilon \geq 0$, we have, for all positive ε less than a given ε_0 and any minimizer u^* of $\mathbf{E}(u; \lambda, f)$,

$$\mathbf{E}_\varepsilon(u_\varepsilon^*; \lambda, f) \leq \mathbf{E}_\varepsilon(u^*; \lambda, f) \leq \mathbf{E}(u^*; \lambda, f) + \sqrt{\varepsilon}. \tag{7.4}$$

From this we conclude that $\mathbf{E}_\varepsilon(u_\varepsilon^*)$, and thus $\|Du_\varepsilon^*\|$, are bounded. Since Ω is bounded (f has compact support), Lemma 7.3 with $p = 1$ and $n = 2$ states there exists $\bar{u} \in BV(\Omega)$ such that $\lim_{i\to\infty} \|u_{\varepsilon_i}^* - \bar{u}\|_{L^1} = 0$ with $\lim_{i\to\infty} \varepsilon_i = 0$. The optimality of $\bar{u}$ follows from

$$\begin{aligned}
&\mathbf{E}(\bar{u}; \lambda, f) = \|D\bar{u}\| + \lambda \int |\bar{u} - f| \\
&= \|D\bar{u}\| + \lambda \lim_{i\to\infty} \int \sqrt{(u_{\varepsilon_i}^* - f)^2 + \varepsilon_i} \quad \text{(dominant convergence)} \\
&\leq \liminf_{i\to\infty} \|Du_{\varepsilon_i}^*\| + \lambda \int \sqrt{(u_{\varepsilon_i}^* - f)^2 + \varepsilon_i} \quad \text{(lower semi-continuity)} \\
&= \liminf_{i\to\infty} \mathbf{E}_{\varepsilon_i}(u_{\varepsilon_i}^*; \lambda, f) \\
&\leq \mathbf{E}(u^*; \lambda, f) \quad \text{(by (7.4))}.
\end{aligned}$$

Since $f \in L^1$ and hence $\bar{v} = f - \bar{u} \in L^1$, we also have $\lim_{\varepsilon\to 0+} \|v_\varepsilon^* - \bar{v}\|_{L^1} = 0$. ∎

Unfortunately, an L^1 error estimate independent of f is not available. We tested $\|u_\varepsilon^* - \bar{u}\|_{L^1}$ for different f's and obtained different orders of magnitudes of ε.

8. Numerical results. In this section, we present numerical results for the TV-L^1 model on multiscale feature selection. Since the second-order cone programming (SOCP) approach [25, 41] has proven to give very accurate solutions for solving TV-based image models, we formulated the TV-L^1 model (1.1) and the G-value formula (5.1) as SOCPs and solved them using the commercial optimization package Mosek [30].

The set of results depicted in Figure 8.2 were obtained by applying the TV-L^1 model with different λ's to the composite input image depicted in Figure 8.1 (f). Each of the five components in this composite image is depicted in Figure 8.1 (S_1)-(S_5). They are the image features that we are interested in extracting from f. We name the components by $S_1, \ldots, S_5$ in the order they are depicted in Figure 8.1. Clearly, they are decreasing in scale. This is further shown by their decreasing G-values (i.e., $G(|\partial \mathbf{1}_{S_i}|)$), and hence, their increasing $\lambda^{\min}$ values (see (5.4)), which are given in Table 8.1. We note that $\lambda_i^{\max}$, for $i = 1, \ldots, 6$, are large since the components do

TABLE 8.1
The G-values (i.e., $G(|\partial \mathbf{1}_{S_i}|)$) and $\lambda^{\min}$ of feature components $S_1, \ldots, S_5$; $\lambda_1, \ldots, \lambda_6$ used to obtain $u_1, \ldots, u_6$.

	S_1	S_2	S_3	S_4	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

not possess smooth edges in the pixelized images. This means that the property (5.5) does not hold for these components, so using the lambda values $\lambda_1, \ldots, \lambda_6$ given in Table 8.1 does not necessarily give the entire feature signals in the outputs u_i, $i = 1, \ldots, 6$. We can see from the numerical results depicted in Figure 8.2 that we were able to produce an output u_i that contains only those features with scales larger that $1/\lambda_i$, leaving in v_i 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, we can see that the numerical results closely match the analytic results given in Subsection 4.1. u_i's contain signal increasing in scale and v_i's contain the residual, which is decreasing in scale. Using the $\lambda_i^{\min}$ values, we were able to get the desired features in u and v. Moreover, by forming differences between the outputs $u_1, \ldots, u_6$, we extracted individual features $S_1, \ldots, S_5$ from input f. These results are depicted in the last two rows of images in Figure 8.2.

Besides multiscale feature selection demonstrated in the test above, the TV-L^1 decomposition can also be used to filter 1D signal [2], to remove impulsive (salt-n-pepper) noise [32], to extract textures from natural images [41], to remove varying illumination in face images for face recognition [21, 20], to decompose 2D/3D images for multiscale MR image registration [19], to assess damage from satellite imagery [18], and to remove inhomogeneous background from cDNA microarray and digital microscopic images [40]. These interesting results were obtained before their theoretical basis was proven above. We believe there exist broader and undiscovered applications of the TV-L^1 model or variants of it, and we hope that the insights into the TV-L^1 model provided here help in identifying such applications.

Acknowledgement. The authors want to express their appreciations to Terrence Chen (UIUC and Siemens Corporate Research), Kevin Vixie (Los Alamos National Lab), Bill Allard (Duke), Selim Esedoglu (U.Mich), Tony Chan (UCLA), and Otmar Scherzer (U.Innsbruck). Dr. Chen's early broad and successful applications of the TV-L^1 model encouraged us to explorer the model. Dr. Vixie's, Prof. Esedoglu's, and Prof. Allard's on-going work and many discussions with the first author have given us insight into the geometric properties of the model. Prof. Chan and Prof. Esedoglu introduced the TV-L^1 model to us, and their work [16] is the foundation of this paper. Our communications with Prof. Scherzer inspired us to associate the G-value with scale-based feature selection. Finally, we thank an anonymous referee for informing us of related work and helping us improve this paper.

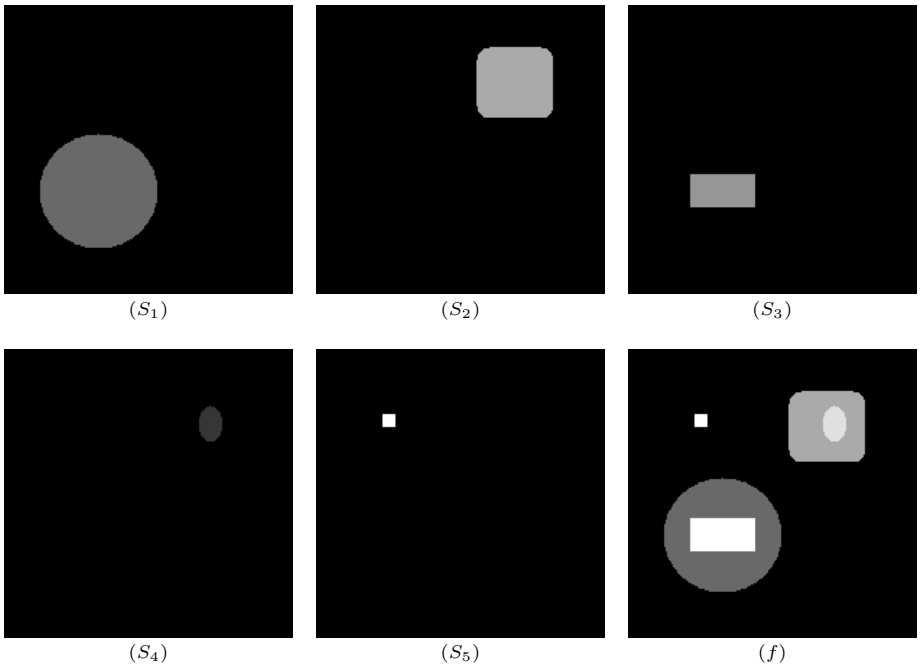

FIG. 8.1. *(S_1)-(S_5): individual feature components of composite image (f).*

Appendix A.

PROPOSITION A.1 (Annulus input). *Let the observed image be the characteristic function $f = c\mathbf{1}_{A_{r_1,r_2}(y)}, \quad 0 < r_2 < r_1$, which has intensity $c > 0$ at all points in the annulus $A_{r_1,r_2}(y) := \{(x_1, x_2) : r_2^2 \le (x_1 - y_1)^2 + (x_2 - y_2)^2 \le r_1^2\}$ and intensity 0 everywhere else. Let u^* denote the minimizer of the TV-L^1 energy $\mathbf{E}(u; \lambda, f)$.*

1. If $r_2 < \frac{r_1}{2}$, then

$$u^*_\lambda = \begin{cases} 0, & \text{if } \lambda < \frac{2r_1}{r_1^2 - 2r_2^2}, \\ \rho c\mathbf{1}_{B_{r_1}(y)}, \ \forall \rho \in [0,1], & \text{if } \lambda = \frac{2r_1}{r_1^2 - 2r_2^2}, \\ c\mathbf{1}_{B_{r_1}(y)}, & \text{if } \frac{2r_1}{r_1^2 - 2r_2^2} < \lambda < \frac{2}{r_2}, \\ \rho c\mathbf{1}_{B_{r_1}(y)} + (1-\rho)f, \ \forall \rho \in [0,1], & \text{if } \lambda = \frac{2}{r_2}, \\ f, & \text{if } \lambda > \frac{2}{r_2}. \end{cases}$$

2. If $r_2 = \frac{r_1}{2}$, then

$$u^*_\lambda = \begin{cases} 0, & \text{if } \lambda < \frac{2}{r_1 - r_2}, \\ \rho c\mathbf{1}_{B_{r_1}(y)} + \bar{\rho} f, \ \forall \rho, \bar{\rho} \in [0,1] \ \ni \rho + \bar{\rho} \le 1, & \text{if } \lambda = \frac{2}{r_1 - r_2}, \\ f, & \text{if } \lambda > \frac{2}{r_1 - r_2}. \end{cases}$$

3. If $\frac{r_1}{2} < r_2$, then

$$u^*_\lambda = \begin{cases} 0, & \text{if } \lambda < \frac{2}{r_1 - r_2}, \\ \rho f, \ \forall \rho \in [0,1], & \text{if } \lambda = \frac{2}{r_1 - r_2}, \\ f, & \text{if } \lambda > \frac{2}{r_1 - r_2}. \end{cases}$$

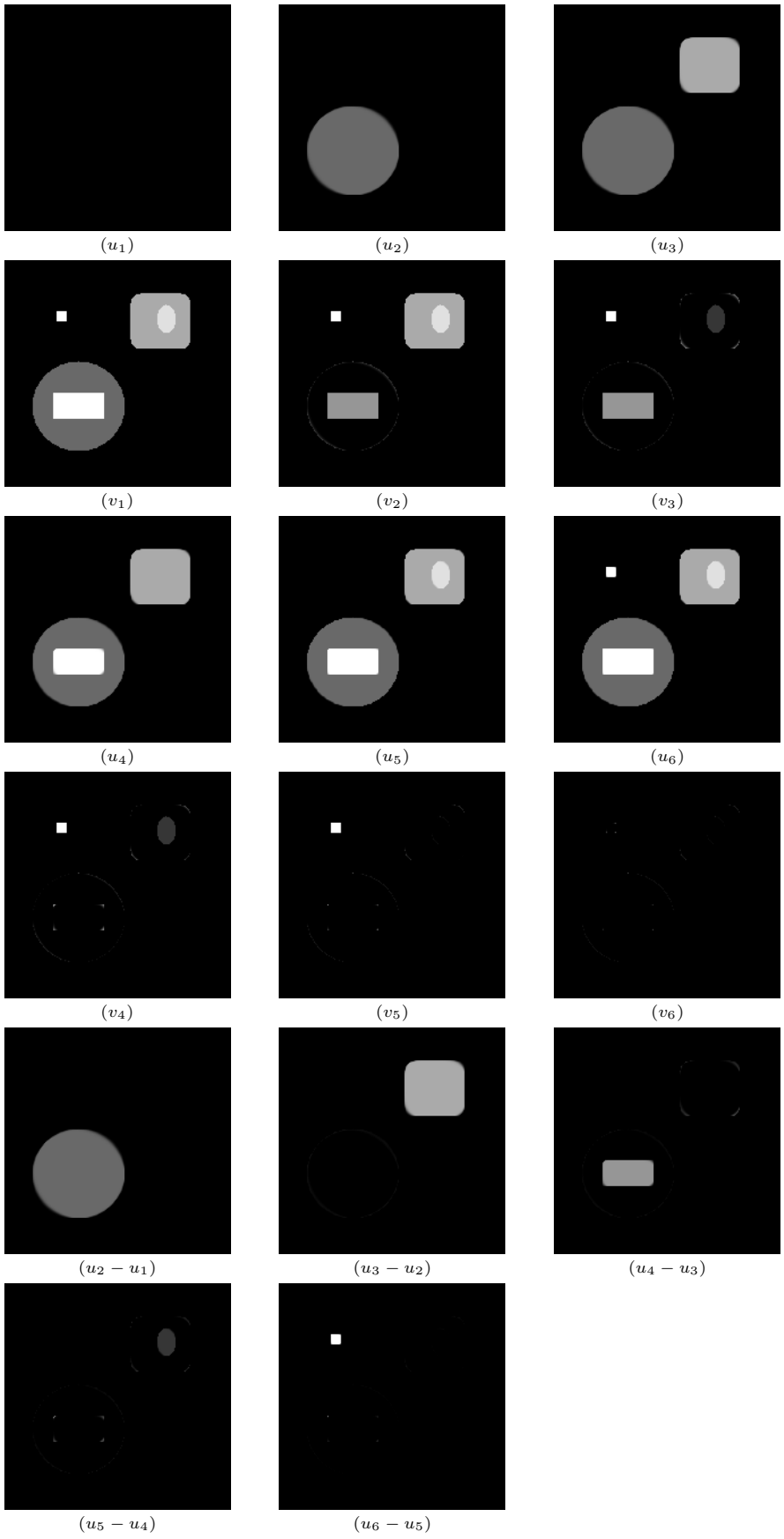

FIG. 8.2. *TV-L^1 decomposition outputs (u_i and v_i were obtained using λ_i).*

Proof. $L(f,\mu) \equiv A_{r_1,r_2}(y)$ for $\mu \in (0,1)$. It follows from $u^* = \text{argmin}_u \mathbf{E}(u;\lambda,f)$ and the Layer Cake formula (2.1) that $U_\mu := L(u^*,\mu)$ must minimize

$$\mathbf{Per}\,(U_\mu) + \lambda |U_\mu \triangle A_{r_1,r_2}(y)|. \tag{A.1}$$

We first show there exists a minimizer of (A.1) that is rotationally symmetric. Let $\bar{U}$ denote an arbitrary minimizer of (A.1) and suppose that $\bar{U}$ is not rotationally symmetric. Let $\bar{U}(\alpha)$ denote another minimizer of (A.1) obtained by rotating $\bar{U}$ clockwise around y for α radians. It follows from the Layer Cake formula (2.1) that

$$\bar{u} = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{1}_{\bar{U}(\alpha)} \mathrm{d}\alpha$$

minimizes $\mathbf{E}(u;\lambda,f)$. Therefore, $\tilde{U} := \{x : \bar{u} > 0\}$ is a minimizer of (A.1) that is rotationally symmetric. Using this result we can narrow the set of solutions to the empty set and rotationally symmetric (about y) sets. We now briefly outline the rest of the proof and leave the details to the reader.

First, we further limit the search for a solution U_λ^* to the empty set and rotationally symmetric sets with a single connected component, and this gives a U_λ^* exactly as the $L(u_\lambda^*, 0)$.

Then, allowing U to have more than one connected components, say $U = U_1 \cup \ldots \cup U_n$, and using the fact that minimizing (A.1) is equivalent to

$$\begin{aligned} &\min_U \mathbf{Per}(U) + \lambda |U \setminus A_{r_1,r_2}(y)| - \lambda |U \cap A_{r_1,r_2}(y)| \\ &= \min_U \sum_{i=1}^n \left(\mathbf{Per}(U_i) + \lambda |U_i \setminus A_{r_1,r_2}(y)| - \lambda |U_i \cap A_{r_1,r_2}(y)|\right), \end{aligned}$$

we conclude that each U_i must minimize (A.1) and hence equal to U_λ^*. Therefore, U_λ^* is a minimizer of the geometry problem (A.1), and the proposition follows from Theorem 4.1. □

REFERENCES

[1] R. Acar and C. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] S. Alliney, *Digital filters as absolute norm regularizers*, IEEE Transactions on Signal Processing, 40 (1992), pp. 1548–1562.

[3] ———, *Recursive median filters of increasing order: a variational approach*, IEEE Transactions on Signal Processing, 44 (1996), pp. 1346–1354.

[4] ———, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Transactions on Signal Processing, 45 (1997), pp. 913–917.

[5] F. Alter, V. Caselles, and A. Chambolle, *Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow*, Interfaces and Free Boundaries, 7 (2005), pp. 29–53.

[6] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, New York, 2000.

[7] L. Ambrosio, N. Gigli, and G. Savare, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Birkhauser, 2005.

[8] F. Andreu, V. Caselles, J. I. Diaz, and J. M. Mazon, *Some qualitative properties for the total variation flow*, Journal of Functional Analysis, 188 (2002), pp. 516–547.

[9] F. Andreu, V. Caselles, and J. M. Mazon, *Parabolic quasilinear equations minimizing linear growth functionals*, vol. 223 of Progress in Math, Birkhauser, 2004.

[10] G. Aubert and J. F. Aujol, *Modeling very oscillating signals, application to image processing*, Applied Mathematics and Optimization, 51 (2005), pp. 163–182.

[11] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equiations and the Calculus of Variations*, vol. 147 of Applied Mathematical Sciences, Springer, 2002.
[12] G. Bellettini, V. Caselles, and M. Novaga, *The total variation flow in R^N*, Journal of Differential Equations, 184 (2002), pp. 475–525.
[13] ———, *Explicit solutions of the eigenvalue problem $-\mathrm{div}\,(Du/|Du|) = u$*, SIAM Journal on Mathematical Analysis, 36 (2005), pp. 1095–1129.
[14] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97.
[15] ———, *Total variation minimization and a class of binary MRF models*, Tech. Report UMR CNRS 7641, Ecole Polytechnique, 2005.
[16] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM Journal on Applied Mathematics, 65 (2005), pp. 1817–1837.
[17] T. F. Chan and J. Shen, *Image Processing and Analysis*, SIAM Publisher, Philadelphia, 2005.
[18] T. Chen and T. Huang, *Non-rigid registration based on a new image hierarchy*, Tech. Report url: `http://www.ifp.uiuc.edu/~tchen5/satellite.html`, UIUC, 2006.
[19] T. Chen, T. Huang, W. Yin, and X. S. Zhou, *A new coarse-to-fine framework for 3D brain MR image registration*, in Computer Vision for Biomedical Image, vol. 3765 of Lecture Notes in Computer Science, Springer, 2005, pp. 114–124.
[20] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. Huang, *Total variation models for variable lighting face recognition*, To appear in IEEE Transactions of Pattern Analysis and Machine Intelligence, (2006).
[21] T. Chen, W. Yin, X. S. Zhou, D. Domaniciu, and T. Huang, *Illumination normalization for face recognition and uneven background correction using total variation based image models*, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, San Diego, 2005, pp. 532–539.
[22] Jerome Darbon and Marc Sigelle, *A fast and exact alogrithm for total variation minimization*, Tech. Report Technical Report, EPITA Research and Development Laboratory (LRDE), Jan 20, 2005 2005.
[23] K. Fan, *Minimax theorem*, Proceedings of the National Academy of Sciences, 9 (1953), pp. 42–47.
[24] E. Giusti, *Mnimal Surfaces and Functions of Bounded Variation*, Birkhäuser, 1984.
[25] D. Goldfarb and W. Yin, *Second-order cone programming methods for total variation based image restoration*, SIAM Journal on Scientific Computing, 27 (2005), pp. 622–645.
[26] A. Haddad and Y. Meyer, *Variational methods in image processing*, Tech. Report CAM Report 04-52, UCLA, 2004.
[27] S. Kindermann, S. Osher, and J. Xu, *Denoising by BV-duality*, Tech. Report CAM Report 05-21, UCLA, 2005.
[28] V. Kolmogorov and R. Zabih, *What energy functions can be minimized via graph cuts?*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26 (2004), pp. 147–159.
[29] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations*, vol. 22 of University Lecture Series, AMS, 2002.
[30] Mosek ApS Inc., *The Mosek optimization tools, ver 4.*, 2006.
[31] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms*, SIAM Journal on Numerical Analysis, 40 (2002), pp. 965–994.
[32] ———, *A variational approach to remove outliers and impluse noise*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 99–120.
[33] ———, *Weakly constrained minimization. application to the estimation of images and signals involving constant regions*, Journal of Mathematical Imaging and Vision, 21 (2004), pp. 155–175.
[34] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterated regularization method for total variation based image restoration*, SIAM Journal on Multiscale Modeling and Simulation, 4 (2005), pp. 460–489.
[35] S. Osher and O. Scherzer, *g-norm properties of bounded variation regularization*, Communications in Mathematical Sciences, 2 (2004), pp. 237–254.
[36] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[37] O. Scherzer, W. Yin, and S. Osher, *Slope and G-set characterization of set-valued functions and applications to non-differentiable optimization problems*, Communications in Mathematical Sciences, 3 (2005), pp. 479–492.
[38] M. Sion, *On general minimax theorems*, Pacific Journal of Mathematics, 8 (1958), pp. 171–176.
[39] D. Strong and T. F. Chan, *Edge-preserving and scale-dependent properties of total variation*

regularization, Inverse Problems, 19 (2003), pp. 165–187.

[40] W. Yin, T. Chen, X. S. Zhou, and A. Chakraborty, *Background correction for cDNA microarray image using the $TV + L^1$ model*, Bioinformatics, 21 (2005), pp. 2410–2416.

[41] W. Yin, D. Goldfarb, and S. Osher, *A comparison of three total variation-based texture extraction models*, Tech. Report CORC Report TR2006-04, Columbia University, 2006.

[42] William P. Ziemer, *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation*, Graduate Texts in Mathematics, Springer, 1989.