COMMUN. MATH. SCI.
Vol. 6, No. 2, pp. 281–307
© 2008 International Press

.

MULTISCALE HIERARCHICAL DECOMPOSITION OF IMAGES WITH APPLICATIONS TO DEBLURRING, DENOISING AND SEGMENTATION*

EITAN TADMOR†, SUZANNE NEZZAR‡, AND LUMINITA VESE§

Abstract. We extend the ideas introduced in [33] for hierarchical multiscale decompositions of images. Viewed as a function $f \in L^2(\Omega)$, a given image is hierarchically decomposed into the sum or product of simpler "atoms" u_k, where u_k extracts more refined information from the previous scale u_{k-1}. To this end, the u_k's are obtained as dyadically scaled minimizers of standard functionals arising in image analysis. Thus, starting with $v_{-1} := f$ and letting v_k denote the residual at a given dyadic scale, $\lambda_k \sim 2^k$, the recursive step $[u_k, v_k] = \operatorname{arginf} \mathcal{Q}_T(v_{k-1}, \lambda_k)$ leads to the desired hierarchical decomposition, $f \sim \sum T u_k$; here T is a blurring operator. We characterize such $\mathcal{Q}_T$-minimizers (by duality) and expand our previous energy estimates of the data f in terms of $\|u_k\|$. Numerical results illustrate applications of the new hierarchical multiscale decomposition for blurry images, images with additive and multiplicative noise and image segmentation.

Key words. natural images, multiscale expansion, total variation, segmentation, image decomposition, image deblurring.

AMS subject classifications. 26B30, 65C20, 68U10

1. Introduction — hierarchical (X,Y) decompositions

We continue our study of the hierarchical image decomposition method introduced by the authors in [33] (hereafter abbreviated TNV). We extend the hierarchical decomposition method to the case of functionals arising in image deblurring, in multiplicative image denoising, and in image segmentation. Convergence results and energy estimates are given, together with experimental results on real images.

The starting point is a pair of normed function spaces, $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, and their associated $\mathcal{Q}$-functional,

$$\mathcal{Q}(f,\lambda) := \mathcal{Q}(f,\lambda;X,Y) := \inf_{u\in X} \Big\{ \|u\|_X + \lambda \|f-u\|_Y^2 \Big\}, \quad \lambda > 0. \tag{1.1}$$

We will talk about functions in the smaller space X being "smoother" than those in Y, so that the condition $\|u\|_X < \infty$ can be viewed as a regularizing constraint. The use of regularized $\mathcal{Q}$-like functionals has a long history starting with the classical Tikhonov-type regularizations; consult [34, 21, 22]. They can be found in a variety of applications; we mention here two: the work on *support vector regression*, e.g., [35], and the early works in the context of *image processing* [14, 30, 1, 10, 15]. In these works, λ is treated as a fixed *threshold parameter*. The $\mathcal{Q}$-functional (1.1) is also closely related to the standard K-functional which arises in *interpolation theory*, e.g., [4, 3, 12, 17],

$$K(f,\lambda;X,Y) := \inf_{u\in X} \Big\{ \|u\|_X + \lambda \|f-u\|_Y \Big\}, \quad \lambda > 0.$$

*Received: November 5, 2007; accepted (in revised version): February 13, 2008. Communicated by Shi Jin.

†Department of Mathematics, Center for Scientific Computation and Mathematical Modeling (CSCAMM) and Institute for Physical Science and Technology (IPST), University of Maryland, College Park, MD 20742, USA (tadmor@cscamm.umd.edu). http://www.cscamm.umd.edu/~tadmor.

‡Richard Stockton College of New Jersey, Pomona, NJ 08240 (nezzars@stockton.edu).

§Department of Mathematics, UCLA, Los Angeles, CA 90095, USA (lvese@math.ucla.edu). http://www.math.ucla.edu/~lvese/

The relation between the $\mathcal{Q}$- and the K-functionals is summarized in

$$\mathcal{Q}(f,\mu_\lambda;X,Y)\approx K(f,\lambda;X,Y),\quad \mu_\lambda:=\frac{\lambda^2}{K(f,\lambda)}.$$

In this context of K-functionals, however, λ does not just serve as a threshold parameter, but in fact is treated as a *variable*: the collection of f's with prescribed behavior of $K(f,\lambda)$ as $\lambda\uparrow\infty$ forms intermediate interpolation smoothness spaces[1]. This is the point of view we adopt in the hierarchical decomposition of $\mathcal{Q}$-functionals described below, with λ being treated as a *scaling variable*.

We assume that the minimization problem has a solution $u:=u_\lambda$, and we let v_λ denote the residual, $v_\lambda:=f-u_\lambda$. This will be expressed as

$$f=u_\lambda+v_\lambda,\quad [u_\lambda,v_\lambda]=\arg\inf_{u+v=f}\mathcal{Q}(f,\lambda;X,Y). \tag{1.2a}$$

In general, $\|u\|_X$ will be a regularizing term; thus u_λ will contain only the "larger" features of f, while the residual v_λ will contain the "smaller" features. Of course, the distinction between these two components is scale-dependent – whatever is interpreted as 'small' features at a given λ-scale, may contain significant features when viewed under a refined scale, say 2λ

$$v_\lambda=u_{2\lambda}+v_{2\lambda},\quad [u_{2\lambda},v_{2\lambda}]=\arg\inf_{u+v=v_\lambda}\mathcal{Q}(v_\lambda,2\lambda). \tag{1.2b}$$

By combining (1.2a) with (1.2b) we arrive at a better two-scale representation of f given by $f\approx u_\lambda+u_{2\lambda}$. Features below scale $1/2\lambda$ remain unresolved in $v_{2\lambda}$, but the process (1.2b) can be continued. Starting with an initial scale $\lambda=\lambda_0$,

$$f=u_0+v_0,\quad [u_0,v_0]=\arg\inf_{u+v=f}\mathcal{Q}(f,\lambda_0), \tag{1.3a}$$

a more refined decomposition of f into simpler "atoms" is obtained by successive application of the dyadic refinement step (1.2b),

$$v_j=u_{j+1}+v_{j+1},\quad [u_{j+1},v_{j+1}]:=\arg\inf_{u+v=v_j}\mathcal{Q}(v_j,\lambda_0 2^{j+1};X,Y),\quad j=0,1,.... \tag{1.3b}$$

After k such steps, we end up with the following *hierarchical decomposition* of f:

$$\begin{aligned} f&=u_0+v_0\\ &=u_0+u_1+v_1\\ &=.....\\ &=u_0+u_1+\cdots+u_k+v_k. \end{aligned} \tag{1.4}$$

The above multiscale expansion provides a new hierarchical representation of the data $f\in Y$,

$$f\cong\sum_{j\geq 0}u_j,\qquad u_j\in X, \tag{1.5}$$

[1]Observe that the definition of the K-functional here exchanges the usual ordering between X and Y, so that it scales with *increasing* λ, in order to keep the compatability with the usual use of the scaling of $\mathcal{Q}$-functional in image processing.

where the approximate equality, $\cong$ in (1.5), should be interpreted as the convergence of the residual v_k's in (1.4), to be made precise below. The partial sum, $\sum_{j=0}^{k} u_j$, provides a multi-layered description of f which lies in an intermediate scale of spaces, in between X and Y, though the precise regularity may vary, depending on the scales which are present in f.

Multi-layered representations of images are not new. We mention here those based on wavelet expansions, e.g., [19], [7], and the TV based expansion suggested by Rudin and Caselles in [26]. As a prototype example, we mention the hierarchical (BV,L^2) decomposition introduced in TNV. A special case of Tikhonov regularization with $(X,Y)=(BV(\Omega),L^2(\Omega))$ is the Total Variation (TV) functional[1] introduced in [10], $\mathcal{Q}(f,\lambda,BV(\Omega),L^2(\Omega))$, to recover a sharp image u from its noisy version, $f=u+v$, corresponding to the TV constrained minimization of Rudin-Osher-Fatemi [30]. Here, $\lambda>0$ is a threshold parameter. The (BV,L^2)-hierarchical decomposition corresponding to (1.5) was introduced by the authors in [33],

$$f\cong\sum_j u_j,\qquad u_j\in BV(\Omega). \tag{1.6}$$

This decomposition is independent of a priori parameters and is *essentially nonlinear* in the sense that its dyadic blocks, u_j, depend on the data itself, $u_j=u_j(f)$. The dyadic blocks capture different layers of scales of the original image. Their precise multiscale nature is quantified in terms of the energy decomposition [33, corollary 2.3]

$$\|f\|^2_{L^2(\Omega)}=\sum_{j=0}^{\infty}\left[\frac{1}{\lambda_j}|u_j|_{BV(\Omega)}+\|u_j\|^2_{L^2(\Omega)}\right].$$

The hierarchical, multi-layered (BV,L^2)-decompositions (1.6) were found to be effective tools in image processing [33, 8] and image registration [18, 27]. They apply to image denoising in the presence of additive or multiplicative noise, to image deblurring and to image segmentation. These extensions of hierarchical decompositions will be discussed in the sections below.

REMARK 1.1. (The homogeneity of the hierarchical decomposition). We note here the anomaly of the $\mathcal{Q}$-functional (1.1): when an image f with minimizer $[u,v]$ doubles its intensity, $2f$, its minimizer does not scale accordingly since the quadratic-based $\mathcal{Q}$ is not homogeneous. This anomaly of the $\mathcal{Q}$-functionals is fixed by their hierarchical decompositions. To this end, we observe that if $[u_0,v_0]$ is the minimizing pair of $\mathcal{Q}(f,\lambda)$ then $[2u_0,2v_0]$ is the minimizer of $\mathcal{Q}(2f,\lambda/2)$. Consequently, if an image f has the hierarchical description (1.3b), $f\cong\sum_{j=0}^{\infty}u_j$, then we find recursively that

$$2v_j=2u_{j+1}+2v_{j+1},\quad [2u_{j+1},2v_{j+1}]:=\arg\inf_{u+v=2v_j}\mathcal{Q}(v_j,\lambda_0 2^j;X,Y),\quad j=0,1,....$$

We conclude that the hierarchical decomposition is homogeneous of degree one: when doubling the intensity, $2f$ has the corresponding hierarchical decomposition $2f\cong\sum_{j=0}^{\infty}2u_j$.

[1] TV should not be confused with TNV.

2. Hierarchical decomposition of blurry and noisy images

We are given a blurred image, represented by $f \in L^2(\Omega)$; blurring will be modeled by a linear, continuous blurring operator, $T : L^2(\Omega) \to L^2(\Omega)$ (such as a convolution with a Gaussian kernel). We consider a decomposition of f provided by the following $\mathcal{Q}_T(f,\lambda)$ minimization in the presence of blur,

$$\mathcal{Q}_T(f,\lambda;X(\Omega),L^2(\Omega)) := \inf_{u\in X(\Omega)} \Big\{ \|u\| + \lambda \|f - Tu\|^2_{L^2(\Omega)} \Big\}. \tag{2.1a}$$

Here, the *regularization functional* $\|\cdot\| : X \subset L^2(\Omega) \to [0,\infty]$ is a semi-norm which takes the general form:

$$\|u\| \equiv \|u\|_p := \int_\Omega \phi(D^p u), \qquad p \geq 1. \tag{2.1b}$$

A few examples for such regularizing functionals are in order. If $\|u\| = |u|_{BV(\Omega)} = \int_\Omega |Du|$ is the total variation of u, then (2.1a) becomes the denoising-deblurring model introduced in [9, 10] following the BV-type constrained minimization of [30, 31]. A more general BV-type model is provided by $\|u\| = \int_\Omega \phi(|Du|)$ with a proper norm, ϕ, defined on the space of measures; the corresponding $\mathcal{Q}_T$ in(2.1a) then becomes the generalized BV model studied in [11, 36, 13]. Other examples, defined on dense subspaces of BV, are provided by $\|u\| = \|Du\|_{L^2(\Omega)} : H^1(\Omega) \to [0,\infty)$, or the second-order $\|u\|_2 = \int_\Omega |D^2 u| : BH(\Omega) \to [0,\infty)$ defined on BH, the space of functions with bounded Hessian.

Let $v_\lambda := f - Tu_\lambda$ denote the 'texture' at scale λ associated with the blurring model (2.1a),

$$[u_\lambda, v_\lambda] = \arg \inf_{Tu+v=f} \mathcal{Q}_T(f,\lambda).$$

Starting with $\lambda = \lambda_0$ in (2.1a),

$$[u_0, v_0] = \arg \inf_{Tu+v=f} \mathcal{Q}_T(f,\lambda_0), \tag{2.2a}$$

we proceed by iterating at the dyadic scales $\lambda_j := \lambda_0 2^j$:

$$[u_{j+1}, v_{j+1}] = \arg \inf_{Tu_{j+1}+v_{j+1}=v_j} \mathcal{Q}(v_j, \lambda_{j+1}), \qquad \lambda_j := \lambda_0 2^j, j = 0,1,\ldots. \tag{2.2b}$$

Thus we have $v_j = Tu_{j+1} + v_{j+1}$ where $v_{-1} := f$. Summing the last recursive relation, we end up with a hierarchical representation of the blurred image f,

$$f = Tu_0 + Tu_1 + \cdots + Tu_{k-1} + Tu_k + v_k, \tag{2.3}$$

which in turn paves the way for a hierarchical, multiscale denoised-deblurred expansion

$$u \cong \sum_{j\geq 0} u_j.$$

REMARK 2.1. (When to start and when to stop the hierarchical decomposition). How should one choose the initial scale λ_0? According to Corollary 2.5 below, if $\lambda_0 \|T^* f\|_* \leq 1/2$, then the minimizer of (1.3a) is $[u_0, v_0] = [0, f]$. Thus, if λ_0 is chosen too small relative to the amount of texture/noise/blurring present in f, then its dyadic

multiples, $2^j\lambda_0,\ j=0,1,\ldots$, will have zero contribution to the hierarchical expansion until they reach the critical value of $2^{j_0}\lambda_0$ so that

$$\frac{1}{2}<2^{j_0}\lambda_0\|T^*f\|_*\le 1.$$

This then dictates the precise initial step of the hierarchical decomposition,

$$u\cong\sum_{j\ge j_0}^{m}u_j. \tag{2.4}$$

We note that the last expansion *has to* be truncated at appropriately chosen finite m in order to avoid the ill-posedness which must occur as we accumulate infinitely many terms, approaching the "inversion" of the ill-conditioned T. When should one truncate the expansion? A systematic study can be found in, e.g., [25] which studies hierarchical decompositions based on inverse scale iterations.

REMARK 2.2. (Dyadic scales). How important is the use of dyadic scales, $\lambda_j\sim 2^j$? As noted in TNV, one can use hierarchical decomposition with *any* sequence of scales, $\lambda_0<\lambda_1<\ldots<\lambda_m$. In particular, $\lambda_j\sim\lambda^j$ with $\lambda>1$. The dyadic choice is a default choice (think wavelets), in the absence of any other systematic policy. Ideally, such a policy will dictate a judicious, data-dependent choice of how the next scale, λ_{k+1}, is to be determined by the information available from previous layers, $\sum_{j\ge j_0}^{k}u_j$.

2.1. Hierarchical decomposition using $\mathcal{Q}_T$-minimizers. To study the hierarchical expansions (2.4), we first characterize the minimizers of the $\mathcal{Q}_T$-functionals (2.1). The characterization summarized in the theorem below extends Meyer's result [20, Theorem 4] (and we also refer to [5, Chapter1] for related characterization of minimizers involving dual functionals).

We recall that the regularizing functional $\|f\|$ in (2.1b) is a semi-norm and we define its *dual* with respect to the $L^2(\Omega)$ scalar product $\langle\cdot,\cdot\rangle$,

$$\|f\|_*:=\sup_{\|\varphi\|\ne 0}\frac{\langle f,\varphi\rangle}{\|\varphi\|}, \tag{2.5}$$

so that the usual duality holds

$$\langle f,\varphi\rangle\le\|\varphi\|\|f\|_*. \tag{2.6}$$

We say that (f,φ) is an *extremal pair* if equality holds above. The theorem below characterizes u as minimizer of the $\mathcal{Q}_T$-functional if and only if u and $T^*(f-Tu)$ form an extremal pair.

THEOREM 2.3. *Let $T:L^2(\Omega)\to L^2(\Omega)$ be a linear continuous blurring operator with adjoint T^* and let $\mathcal{Q}_T$ denote the associated functional (2.1a).*

(i) The variational problem (2.1) admits a minimizer u. Moreover, if $\|\cdot\|$ is strictly convex, then a minimizer u with $\|u\|\ne 0$ is unique.

(ii) u is a minimizer of (2.1) if and only if

$$\big\langle u,T^*(f-Tu)\big\rangle=\|u\|\cdot\|T^*f-T^*Tu\|_*=\frac{\|u\|}{2\lambda}. \tag{2.7}$$

The detailed proof of the theorem, whose second part was sketched in [20]) is postponed to the end of this section. The next two remarks lead to refinement of Theorem 2.3, depending on the size of $\|T^*f\|_*$.

REMARK 2.4. (The trivial minimizer). More can be said in case f consists mostly of texture, in the sense that

$$\|T^*f\|_* \le \frac{1}{2\lambda}. \tag{2.8}$$

Indeed, if (2.8) holds, then the characterization of the minimizer u in (2.7) implies

$$\frac{\|u\|}{2\lambda} = \langle u, T^*f\rangle - \langle u, T^*Tu\rangle \le \frac{\|u\|}{2\lambda} - \|Tu\|_{L^2}^2,$$

and hence $Tu=0$. But then $\mathcal{Q}_T(f,\lambda) = \|u\| + \lambda\|f\|_{L^2}^2$ implies that $u\equiv 0$ is a minimizer of (2.1), with $\mathcal{Q}_T(f,\lambda) = \lambda\|f\|_{L^2}^2$. The converse of this assertion also holds. We summarize with the following corollary.

COROLLARY 2.5. *[The case $\|T^*f\|_* \le 1/2\lambda$]. Let $T: L^2(\Omega)\to L^2(\Omega)$ be a linear continuous blurring operator with adjoint T^* and let $\mathcal{Q}_T$ denote the associated functional (2.1a). Then $\|T^*f\|_* \le \frac{1}{2\lambda}$ if and only if $u\equiv 0$ is a minimizer of (2.1).*

Proof. Assume $\|T^*f\|_* \le 1/2\lambda$. We have already seen that $u\equiv 0$ is a minimizer. One can also argue directly that since $2\lambda\langle T^*f,\varphi\rangle \le \|\varphi\|$ for all $\varphi\in BV(\Omega)$,

$$\begin{aligned}\|\varphi\| + \lambda\|f - T\varphi\|_{L^2(\Omega)}^2 &= \|\varphi\| + \lambda\|f\|_{L^2(\Omega)}^2 - 2\lambda\langle f, T\varphi\rangle + \lambda\|T\varphi\|_{L^2(\Omega)}^2\\ &= \|\varphi\| - 2\lambda\langle T^*f,\varphi\rangle + \lambda\|f\|_{L^2(\Omega)}^2 + \lambda\|T\varphi\|_{L^2(\Omega)}^2\\ &\ge \lambda\|f\|_{L^2(\Omega)}^2 + \lambda\|T\varphi\|_{L^2(\Omega)}^2 \ge \lambda\|f - T0\|_{L^2(\Omega)}^2 + \|0\|,\end{aligned}$$

and therefore $u\equiv 0$ is a minimizer of (2.1a). It remains to verify the "if" part, namely, if $u\equiv 0$ is a minimizer of (2.1), then for all $\varphi\in BV(\Omega)$ we have

$$\lambda\|f\|_{L^2(\Omega)}^2 \le \lambda\|T\varphi - f\|_{L^2(\Omega)}^2 + \|\varphi\| \tag{2.9}$$

or

$$2\lambda\langle f, T\varphi\rangle \le \lambda\|T\varphi\|_{L^2(\Omega)}^2 + \|\varphi\|.$$

Rescaling $\varphi\to\epsilon\varphi$, we obtain

$$2\lambda\epsilon\langle f, T\varphi\rangle \le \lambda\epsilon^2\|T\varphi\|_{L^2(\Omega)}^2 + \epsilon\|\varphi\|.$$

Dividing by ϵ and letting $\epsilon\to 0_+$ yields $2\lambda\langle T^*f,\varphi\rangle = 2\lambda\langle f, T\varphi\rangle \le \|\varphi\|$ for all $h\in BV(\Omega)$, and we conclude $\|T^*f\|_* \le 1/2\lambda$. ∎

REMARK 2.6. (Equivalence classes). Consider the pth order semi-norm in (2.1b), $\|u\| \equiv \|u\|_p := \int_\Omega \phi(D^p u)$, and assume

$$\|T^*f\|_* < \infty.$$

We note that $\|\cdot\|_* \equiv \big(\|\cdot\|_p\big)_*$ should be considered on the complement of appropriate equivalence classes of "modulo polynomials of degree p". Indeed, since $\|\varphi + P\|_p =$

$\|\varphi\|_p$ for any polynomial $P=P(x)$ of degree $\le p-1$, we have for arbitrary constant c that

$$\frac{\langle f,\varphi+cP\rangle}{\|\varphi+cP\|}=\frac{\langle f,\varphi\rangle+\langle f,cP\rangle}{\|\varphi\|}=\frac{\langle f,\varphi\rangle}{\|\varphi\|}+c\frac{\langle f,P\rangle}{\|\varphi\|}.$$

Thus, $\|T^*f\|_*<\infty$ implies $\langle T^*f,P\rangle=0$ for all $\deg P\le p-1$.

In particular, if we assume that

$$\frac{1}{2\lambda}<\|T^*f\|_*<\infty,$$

then a minimizer u of $\mathcal{Q}_T(f)$ does not vanish, i.e., $\|u\|\neq 0$. Otherwise, if $\|u\|=0$, then u would be a polynomial of degree $\le p-1$; but by the preceding argument, the polynomial u should be orthogonal to T^*f and hence

$$\mathcal{Q}_T(f,\lambda)=\lambda\|f-Tu\|_{L^2}^2=\lambda\big(\|f\|_{L^2}^2+\|Tu\|_{L^2}^2\big),$$

which is minimized when $Tu=0$. Given that $\|u\|=Tu=0$, one can follow the proof of Corollary 2.5, starting with (2.9) and concluding that $\|T^*f\|_*\le 1/2\lambda$, which contradicts our assumption. We can summarize this case in the following corollary.

COROLLARY 2.7. (The case $\|T^*f\|_*>1/2\lambda$). *Let $T:L^2(\Omega)\to L^2(\Omega)$ be a linear continuous blurring operator with adjoint T^* and let $\mathcal{Q}_T$ denote the associated functional (2.1a) with $\|\cdot\|=\|\cdot\|_p$. Assume that*

$$\frac{1}{2\lambda}<\|T^*f\|_*<\infty.$$

Then u is a minimizer of (2.1) if and only if u and $T^*(f-Tu)$ is an extremal pair and

$$\|T^*f-T^*Tu\|_*=\frac{1}{2\lambda}. \tag{2.10}$$

Moreover, if $\|\cdot\|$ is strictly convex then the minimizer u is unique.

Proof. We can now divide the equality on the right of (2.7) by $\|u\|=\|u\|_p\neq 0$. Moreover, since $\|T^*f\|_*>1/2\lambda$, it follows by Remark 2.6 that $\|u\|\neq 0$, and uniqueness follows from Theorem 2.3(i). ∎

Equipped with Theorem 2.3, we can extend the (BV,L^2)-hierarchical decompositions introduced in TNV to general $\mathcal{Q}_T$-functionals.

THEOREM 2.8. (**i**)(Hierarchical expansion). *Let $f\in L^2(\Omega)$ and consider the dyadically-based $\mathcal{Q}_T$ decomposition (2.2). Then f admits the following hierarchical expansion*

$$f\cong\sum_{j=0}^{\infty}Tu_j; \tag{2.11a}$$

here the $\cong$ should be interpreted as the convergence $\sum_{j=0}^{k}Tu_j\to f$ in the weak $\|\cdot\|$-sense,

$$\|T^*(f-\sum_{j=0}^{k}Tu_j)\|_*=\frac{1}{\lambda_0 2^{k+1}}. \tag{2.11b}$$

(**ii**) (Energy decomposition). *The following energy estimate holds:*

$$\sum_{j=0}^{\infty}\Big[\frac{1}{\lambda_j}\|u_j\|+\|Tu_j\|^2_{L^2(\Omega)}\Big]\le\|f\|^2_{L^2(\Omega)},\quad \lambda_j:=\lambda_0 2^j. \tag{2.12a}$$

Moreover, if $f\in BV(\Omega)$ then equality holds in (2.12a),

$$\sum_{j=0}^{\infty}\Big[\frac{1}{\lambda_j}\|u_j\|+\|Tu_j\|^2_{L^2(\Omega)}\Big]=\|f\|^2_{L^2(\Omega)},\quad f\in BV(\Omega). \tag{2.12b}$$

We note in passing that the BV regularity assumption can be relaxed for the energy decomposition*, (2.12b), to hold; consult [33, Corollary 2.3].*

Proof. If $\|T^*f\|_*\le 1/2\lambda$ then by Corollary 2.5 the minimizer of (2.1a), $[u_\lambda,v_\lambda]=[0,f]$; otherwise

$$\|T^*v_\lambda\|_*=\frac{1}{2\lambda},\quad \langle Tu_\lambda,v_\lambda\rangle=\frac{1}{2\lambda}\|u_\lambda\|. \tag{2.13}$$

The first statement (2.11a) then follows from the basic hierarchical expansion, $f=\sum_0^k Tu_j+v_k$, while noting that $\|T^*v_k\|_*=1/2\lambda_k$. For the second statement, (2.12), we begin by squaring the basic refinement step, $Tu_{j+1}+v_{j+1}=v_j$,

$$\|v_{j+1}\|^2_{L^2(\Omega)}+\|Tu_{j+1}\|^2_{L^2(\Omega)}+2\langle Tu_{j+1},v_{j+1}\rangle=\|v_j\|^2_{L^2(\Omega)},\quad j=-1,0,1,\dots. \tag{2.14}$$

Observe that the last equality holds for $j=-1$ with v_{-1} interpreted as $v_{-1}:=f$. We recall that $[u_{j+1},v_{j+1}]$ is a minimizing pair for $\mathcal{Q}_T(v_j,\lambda_{j+1})$, and hence, by (2.13),

$$2\langle Tu_{j+1},v_{j+1}\rangle=\frac{1}{\lambda_{j+1}}\|u_{j+1}\|,$$

yielding $\frac{1}{\lambda_{j+1}}\|u_{j+1}\|+\|Tu_{j+1}\|^2_{L^2(\Omega)}=\|v_j\|^2_{L^2(\Omega)}-\|v_{j+1}\|^2_{L^2(\Omega)}$. We sum up obtaining (2.12a):

$$\begin{aligned}\sum_{j=0}^{k}\Big[\frac{1}{\lambda_j}\|u_j\|+\|Tu_j\|^2_{L^2(\Omega)}\Big]&=\sum_{j=-1}^{k-1}\Big[\frac{1}{\lambda_{j+1}}\|u_{j+1}\|+\|Tu_{j+1}\|^2_{L^2(\Omega)}\Big]\\&=\|v_{-1}\|^2_{L^2(\Omega)}-\|v_k\|^2_{L^2(\Omega)}=\|f\|^2_{L^2(\Omega)}-\|v_k\|^2_{L^2(\Omega)}\le\|f\|^2_{L^2(\Omega)}.\end{aligned}$$

Given that f has BV regularity, one can follow the argument in [33, Theorem 2.2] to conclude the equality (2.12b). ∎

We conclude this section with the promised proof:

Proof. (of Theorem 2.3).

(i) The existence of a minimizer for the $\mathcal{Q}_T$-functional follows from standard arguments which we omit; consult [1, 10, 36] or [16, Section 8.6]. We address the issue of uniqueness. Assume u_1 and u_2 are minimizers

$$\|u_i\|+\lambda\|f-Tu_i\|^2_{L^2}=j_{min},\quad i=1,2$$

then by a standard line of argument we consider the average $u_3 := (u_1+u_2)/2$ to find that

$$\begin{aligned} j_{min} &\le \|u_3\| + \lambda \|f - Tu_3\|_{L^2}^2 \\ &\le \frac{1}{2}\Big(\|u_1\| + \lambda \|f-Tu_1\|_{L^2}^2 + \|u_2\| + \lambda \|f-Tu_2\|_{L^2}^2 \Big) - \frac{1}{4}\|T(u_1-u_2)\|_{L^2}^2 \\ &\le j_{min} - \frac{1}{4}\|T(u_1-u_2)\|_{L^2}^2. \end{aligned}$$

Thus, u_1-u_2 belongs to the kernel of T. We then end up with the one-parameter family of minimizers, $u_\theta := u_1 + \theta(u_2-u_1)$, $\theta \in [0,1]$,

$$\begin{aligned} j_{min} &\le \|u_\theta\| + \lambda \|f - Tu_\theta\|_{L^2}^2 \\ &\le \theta \|u_2\| + (1-\theta)\|u_1\| + \theta\lambda \|f-Tu_2\|_{L^2}^2 + (1-\theta)\lambda \|f-Tu_1\|_{L^2}^2 = j_{min}. \end{aligned}$$

Clearly, the two minimizers satisfy $\|u_1\| = \|u_2\|$, and we conclude that the ball $\|u\| = \|u_1\| \neq 0$ contains the segment $\{u_\theta,\ \theta \in [0,1]\}$, which, by strict convexity, must be the trivial segment, i.e., $u_2 = u_1$.

(ii) If u is a minimizer of (2.1a), namely, if for any $\varphi \in BV(\Omega)$ we have

$$\lambda \|f - T(u+\epsilon\varphi)\|_{L^2(\Omega)}^2 + \|u+\epsilon\varphi\| \ge \lambda \|f-Tu\|_{L^2(\Omega)}^2 + \|u\|,$$

then

$$\lambda \epsilon^2 \|T\varphi\|_{L^2(\Omega)}^2 - 2\lambda\epsilon \langle T\varphi, f-Tu\rangle + \|u+\epsilon\varphi\| \ge \|u\|. \tag{2.15}$$

Since $\|\cdot\|$ is sublinear, the last inequality yields

$$\lambda \epsilon^2 \|T\varphi\|_{L^2(\Omega)}^2 - 2\lambda\epsilon \langle T\varphi, f-Tu\rangle + \|u\| + |\epsilon|\|\varphi\| \ge \|u\|,$$

or, after division by $\epsilon > 0$,

$$\|\varphi\| + \lambda\epsilon \|T\varphi\|_{L^2(\Omega)}^2 \ge 2\lambda \langle T\varphi, f-Tu\rangle.$$

Letting $\epsilon \downarrow 0_+$, we obtain $\|\varphi\| \ge 2\lambda\langle T\varphi, f-Tu\rangle$ for any $\varphi \in BV(\Omega)$ and hence

$$\frac{1}{2\lambda} \ge \sup_{\|\varphi\|\neq 0} \frac{\langle \varphi, T^*(f-Tu)\rangle}{\|\varphi\|} = \|T^*f - T^*Tu\|_*. \tag{2.16}$$

To confirm (2.7) it remains to verify the reverse inequality in (2.16). To this end we set $\varphi = u$ and $-1 < \epsilon < 0$ in (2.15), yielding

$$\lambda\epsilon^2 \|Tu\|_{L^2(\Omega)}^2 + (1+\epsilon)\|u\| \ge \|u\| + 2\lambda\epsilon \langle u, T^*(f-Tu)\rangle,$$

or

$$\lambda\epsilon^2 \|Tu\|_{L^2(\Omega)}^2 + \epsilon\|u\| \ge 2\lambda\epsilon \langle u, T^*(f-Tu)\rangle.$$

Dividing by ϵ and letting $\epsilon \uparrow 0_-$, we obtain $\frac{1}{2\lambda}\|u\| \le \langle u, T^*(f-Tu)\rangle$. This, together with (2.16), implies that

$$\frac{1}{2\lambda}\|u\| \le \langle u, T^*(f-Tu)\rangle \le \|u\| \|T^*f - T^*Tu\|_* \le \frac{1}{2\lambda}\|u\|,$$

so that the last inequalities become equalities and (2.7) follows.

Conversely, we show that if u satisfies (2.7) then it is the desired minimizer. To this end, we rewrite

$$\begin{aligned}
&\|f-T(u+\varphi)\|^2_{L^2(\Omega)}\\
&\equiv\big\langle (f-Tu)-T\varphi, f-T(u+\varphi)\big\rangle\\
&\equiv\big\langle f-Tu, f+T(u+\varphi)\big\rangle-2\big\langle f-Tu, T(u+\varphi)\big\rangle-\big\langle T\varphi, f-Tu\big\rangle+\|T\varphi\|^2_{L^2(\Omega)}\\
&\equiv\|f-Tu\|^2_{L^2(\Omega)}-2\overbrace{\big\langle f-Tu, T(u+\varphi)\big\rangle}^{\#1}+2\overbrace{\big\langle Tu, f-Tu\big\rangle}^{\#2}+\|T\varphi\|^2_{L^2(\Omega)}.
\end{aligned}$$

Now, Assumption (2.7), $\|T^*(f-Tu)\|_*=\frac{1}{2\lambda}$, implies $2\lambda(\#1)\le\|u+\varphi\|$ and $2\lambda(\#2)=\|u\|$. We conclude that for any $h\in BV(\Omega)$,

$$\begin{aligned}
&\|u+\varphi\|+\lambda\|f-T(u+\varphi)\|^2_{L^2(\Omega)}\\
&=\|u+\varphi\|-2\lambda(\#1)+2\lambda(\#2)+\lambda\|f-Tu\|^2_{L^2(\Omega)}+\lambda\|T\varphi\|^2_{L^2(\Omega)}\\
&\ge\|u\|+\lambda\|f-Tu\|^2_{L^2(\Omega)}.
\end{aligned}$$

Thus, u is a minimizer of (2.1). ∎

We remark that a lack of uniqueness is demonstrated in an example of [20, pp. 40], using the ℓ^∞-unit ball, which in turn lacks strict convexity. Thus, strict convexity is necessary and sufficient for uniqueness.

2.2. Discretization of Euler-Lagrange equation and numerical results. We consider for illustration the case of the total variation [31], therefore $\|u\|=\int_\Omega|Du|$. In practice, we simplify the formulation by working only on $W^{1,1}(\Omega)$ and we write $\|u\|=\int_\Omega|\nabla u|dxdy$. In order to construct the hierarchical decomposition of f, we use the associated Euler-Lagrange equation of $\mathcal{Q}(f,\lambda)$

$$T^*Tu_\lambda=T^*f+\frac{1}{2\lambda}\ \mathrm{div}\left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\right).$$

When working on a bounded domain Ω, we augment the Euler-Lagrange equations by the following Neumann boundary condition:

$$\left.\frac{\partial u_\lambda}{\partial n}\right|_{\partial\Omega}=0.$$

The hierarchical decomposition, $f\sim\sum_{j=0}^k Tu_j=T\sum_{j=0}^k u_j$, is obtained. Note that we are really interested in the deblurred image $u=\sum_{j=0}^k u_j$, from which the u_j's are constructed as (approximate) solutions of the recursive relation governed by the following PDE:

$$T^*Tu_{j+1}-\frac{1}{2\lambda_{j+1}}\ \mathrm{div}\left(\frac{\nabla u_{j+1}}{|\nabla u_{j+1}|}\right)=\overbrace{-\frac{1}{2\lambda_j}\ \mathrm{div}\left(\frac{\nabla u_j}{|\nabla u_j|}\right)}^{T^*v_j}.$$

We implement our algorithm for this type of image in essentially the same way as for the case without blurring (see [33]). The only difference is that we have to deal

with the blurring operator T, a Gaussian kernel in our experiments, and in this we follow the method of discretization in [6] and [36]. The first step is to remove the singularity when $|\nabla u_\lambda| = 0$ by replacing $\mathcal{Q}_T(f,\lambda)$ with

$$\mathcal{Q}_T^\epsilon(f,\lambda) := \inf_{u\in BV}\Big\{\int_\Omega \sqrt{\epsilon^2+|\nabla u|^2}dxdy+\lambda\|f-Tu\|^2_{L^2(\Omega)}\Big\}.$$

We find the minimizer $u_\lambda \equiv u_{\lambda,\epsilon}$ of the regularized functional associated with $\mathcal{Q}_T^\epsilon$ at each step of our hierarchical decomposition. The associated Euler-Lagrange equations are

$$T^*Tu_\lambda = T^*f + \frac{1}{2\lambda}\ \text{div}\ \Big(\frac{\nabla u_\lambda}{\sqrt{\epsilon^2+|\nabla u_\lambda|^2}}\Big)\ \text{in}\ \Omega, \tag{2.17a}$$

$$\frac{\partial u_\lambda}{\partial n} = 0\ \text{on}\ \partial\Omega, \tag{2.17b}$$

which we solve by a dynamic gradient descent scheme $(x,y,t)\mapsto u(x,y,t)$,

$$\begin{aligned}&\frac{\partial u_\lambda}{\partial t}+T^*Tu_\lambda = T^*f+\frac{1}{2\lambda}\ \text{div}\ \Big(\frac{\nabla u_\lambda}{\sqrt{\epsilon^2+|\nabla u_\lambda|^2}}\Big),\quad (x,y,t)\in\Omega\times[0,+\infty),\\ &\qquad u(x,y,0)=f(x,y),\quad (x,y)\in\Omega.\end{aligned} \tag{2.18}$$

As in TNV, we use a computational grid, $(x_i := ih, y_j := jh,\ t^n := n\Delta t)$, to cover the domain Ω for $t\geq 0$, where h is the cell size. Let D_+, D_- and $D_0 := (D_+ + D_-)/2$ denote the usual forward, backward and centered divided difference, e.g., $D_+^x u_{i,j} = (u_{i+1,j}-u_{i,j})/h$, and let

$$|\nabla u|_{i,j} := \sqrt{\epsilon^2+(D_+^x u^n_{i,j})^2+(D_+^y u^n_{i,j})^2} \tag{2.19}$$

denote the "numerical gradient" at (x_i,y_j), with a safeguard $\epsilon \ll 1$ chosen below the relevant scales, just to secure $|\nabla u|_{i,j}\neq 0$.

We discretize (2.18) with the Gauss-Seidel iteration scheme, using the most recent values of the available $u_{i,j}$'s,

$$\begin{aligned}u^{n+1}_{i,j} &= u^n_{i,j}+\Delta t T^*Tu^n_{i,j}+\Delta t T^*f_{i,j}\\ &+\ \frac{\Delta t}{2\lambda}\left[\frac{u^n_{i+1,j}-u^{n+1}_{i,j}}{h|\nabla u^n|_{i,j}}-\frac{u^{n+1}_{i,j}-u^n_{i-1,j}}{h|\nabla u^n|_{i-1,j}}\right]+\frac{1}{2\lambda}\left[\frac{u^n_{i,j+1}-u^{n+1}_{i,j}}{h|\nabla u^n|_{i,j}}-\frac{u^{n+1}_{i,j}-u^n_{i,j-1}}{h|\nabla u^n|_{i,j-1}}\right].\end{aligned} \tag{2.20}$$

REMARK 2.9. (Semi-implicit scheme). The resulting semi-implicit scheme (2.20) is not more computational intensive than the explicit scheme. Yet, it offers a larger stability region and in fact is unconditionally L^∞-stable for $T=I$, and it converges in fewer iterations than the explicit one.

In Figure 2.1, we show a test image, its blurry version (with Gaussian blur, no noise), the results obtained using the Rudin-Osher model [31], and the results obtained

using the hierarchical model. Following [6, 36], we work with convolution-type T's which are realized by symmetric matrices, $(C_{\alpha\beta})_{\alpha,\beta=1,d}$,

$$Tu_{i,j} := \sum_{\alpha,\beta=1}^{d} C_{\alpha\beta} u_{i+d/2-\alpha,j+d/2-\beta}, \qquad \sum_{\alpha,\beta=1}^{d} C_{\alpha\beta} = 1.$$

Thus $T^*Tu = T^2 u_{i,j} = \sum_{\alpha,\beta=1}^{d} \sum_{\gamma,\delta=1}^{d} C_{\alpha\beta} C_{\gamma\delta} u_{i+d-\alpha-\gamma,j+d-\beta-\delta}$. The proposed hierarchical model gives improved results over the standard RO model in terms of the

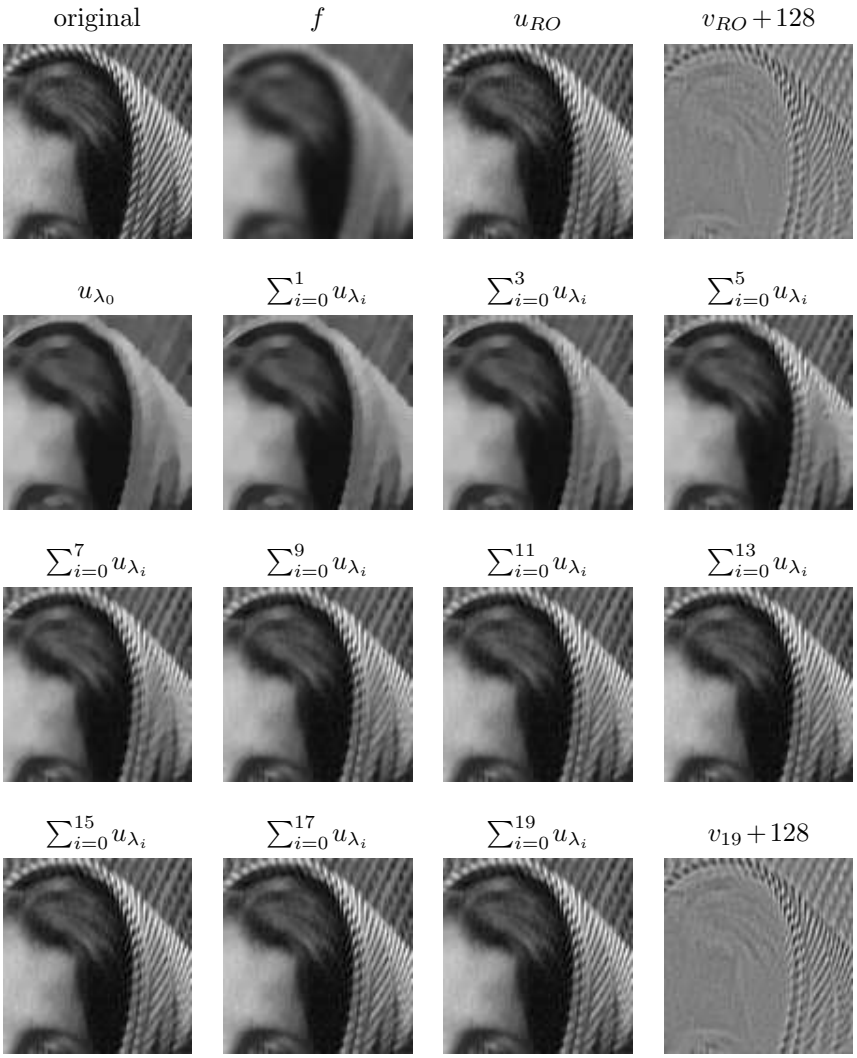

FIG. 2.1. *1st row (from left to right): original test image, its blurry version f, and Rudin-Osher restoration u_{RO}, $v_{RO} = f - u_{RO} + 128$, rmse=0.1066. RO parameters: $\lambda = 2000$, $h = 1$, $\triangle t = 0.1$. 2nd to last rows: hierarchical recovery from the same blurry initial image f using $k = 19$ steps, shown every other one, and residual at step 19. Parameters: $\lambda_0 = 1$, $\triangle t = 0.1$, $h = 1$, and $\lambda_k = 2^k \lambda_0$. Final rmse= 0.0922.*

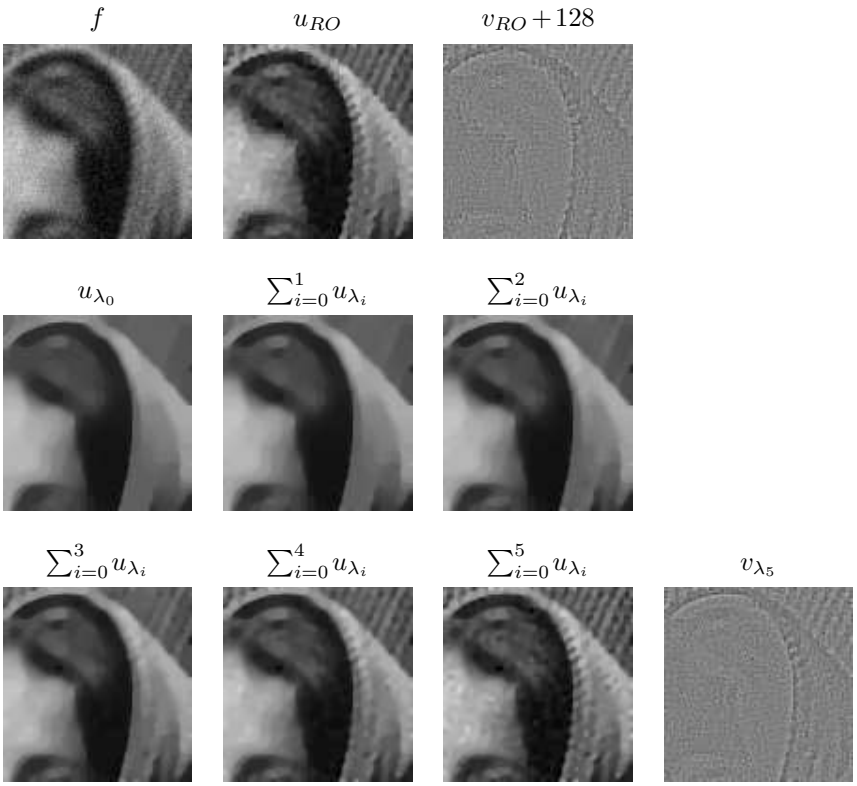

FIG. 2.2. *1st row, from left to right: blurry-noisy version f, and Rudin-Osher restoration u_{RO}, $v_{RO}=f-u_{RO}+128$, rmse =0.2028. RO parameters $\lambda=0.5$, $h=1$, $\triangle t=0.025$. The other rows, left to right: the hierarchical recovery of u from the same blurry-noisy initial image f using 6 steps. Parameters: $\lambda_0=0.02$, $\triangle t=0.025$, $h=1$, and $\lambda_k=2^k\lambda_0$, rmse= 0.2011.*

root mean square error,

$$\text{rmse} := \frac{\sqrt{\sum_{i,j}(u_{\text{orig},i,j}-u_{\text{restored},i,j})^2}}{\sum_{i,j}1}.$$

Similarly Figure 2.2 illustrates how the hierarchical decomposition works for noisy blurred images, again as a (small) improvement over the Rudin-Osher model. Note that the result obtained with the hierarchical model suffers less from the staircasing effect.

3. The hierarchical (SBV, L^2) decomposition

We want to construct the hierarchical decomposition based on the Mumford-Shah functional [23]. To this end we consider its elliptic approximation of Ambrosio and Tortorelli [2],

$$\mathcal{AT}_\rho(f,\lambda) := \inf_{\{u,v,w \mid u+v=f\}} \left\{ \mu \sqrt{\int_\Omega (w^2+\rho\epsilon_\rho)|\nabla u|^2 dx} + \rho\|\nabla w\|^2_{L^2(\Omega)} + \frac{\|w-1\|^2_{L^2(\Omega)}}{4\rho} + \lambda\|v\|^2_{L^2(\Omega)} \right\},$$

where $\epsilon_\rho \to 0$ as $\rho \downarrow 0$, and λ,μ are positive weight parameters.

REMARK 3.1. Note that we modified the $\mathcal{AT}_\rho$-functional, where the first square-root term on the right replaces the original term, $\int_\Omega (w^2+\rho\epsilon_\rho)|\nabla u|^2 dx$, appearing in the work of Ambrosio-Tortorelli [2]. Our modified $\mathcal{AT}_\rho$-functional does not affect the main properties of the segmentation model, however; it is introduced here in order to enable the characterization of $\mathcal{AT}_\rho$-minimizers in Section 3.1 below. Our numerical calculations will then utilize the original formulation of the $\mathcal{AT}_\rho$-functional.

Let $[u_\lambda, v_\lambda]$ be the minimizer of $\mathcal{AT}_\rho(f,\lambda)$ (depending on w). Here $f \in L^\infty(\Omega)$ and u_λ is restricted to the smaller SBV space (a special subclass of BV space consisting of measure gradients free of the Cantor component [2]), while the texture v_λ lives in L^2. We proceed to construct the hierarchical (SBV,L^2) decomposition of f in the same manner as before, letting $[u_{j+1}, v_{j+1}]$ be the AT minimizer

$$[u_{j+1}, v_{j+1}] = \arg \inf_{u,v,w,u+v=v_j} \mathcal{AT}_\rho(v_j, \lambda_j), \quad \lambda_j = \lambda_0 2^j.$$

We end up with the hierarchical decomposition

$$f = u_0 + u_1 + \cdots + u_k + v_k.$$

Here, at each hierarchical step, we also obtain the *edge detectors* $1-w_j = 1-w_{\lambda_j}$, which are (essentially) supported along the boundaries of objects enclosed by edges identified by u_j.

3.1. Characterization of (SBV,L^2)-minimizers. We proceed along the lines of our analysis of general $\mathcal{Q}$-functionals in Section 2.1. We begin with a general characterization of $\mathcal{AT}_\rho$-minimizers as extremal pairs. To this end, we introduce the *weighted spaces* for given $w \in H^1(\Omega)$,

$$|\varphi|_{H^1_w(\Omega)} := \sqrt{\int_\Omega (w^2+\rho\epsilon_\rho)|\nabla\varphi|^2 dx},$$

and we let

$$\|f\|_{H^{-1}_w(\Omega)} := \sup_{\varphi \in H^1_w(\Omega)} \langle f,\varphi\rangle / |\varphi|_{H^1_w(\Omega)}$$

denote the dual norm. We have now the following characterization of the minimizers u,w of the $\mathcal{AT}_\rho$-energy.

THEOREM 3.2. *If $u,w \in [0,1]$ are minimizers of the $\mathcal{AT}_\rho$-energy, then*

$$\|f-u\|_{H^{-1}_w(\Omega)} = \frac{\mu}{2\lambda}, \quad \textit{and} \quad \langle f-u, u\rangle = \frac{\mu}{2\lambda}|u|_{H^1_w(\Omega)}. \tag{3.1}$$

Proof. Let $[u,w]$ be a minimizing pair. Considering the variation of $\mathcal{AT}_\rho$ only with respect to u, we find that for any $h\in H^1(\Omega)$, we have

$$\mu|u|_{H^1_w(\Omega)}+\lambda\|f-u\|^2_{L^2(\Omega)}\le\mu|u+\epsilon\varphi|_{H^1_w(\Omega)}+\lambda\|f-u-\epsilon\varphi\|^2_{L^2(\Omega)}.$$

The triangle inequality yields the "first variation" (or, more precisely, sub-differential)

$$2\lambda\epsilon\langle f-u,\varphi\rangle\le\lambda\epsilon^2\|\varphi\|^2_{L^2(\Omega)}+\mu\epsilon|\varphi|_{H^1_w(\Omega)}. \tag{3.2}$$

For $\epsilon>0$ this gives $2\lambda\langle f-u,\varphi\rangle\le\lambda\epsilon\|\varphi\|^2_{L^2(\Omega)}+\mu|\varphi|_{H^1_w(\Omega)}$, and, as $\epsilon\to 0_+$, we deduce that for all $\varphi\in H^1_w(\Omega)$,

$$\frac{\langle f-u,\varphi\rangle}{|\varphi|_{H^1_w(\Omega)}}\le\frac{\mu}{2\lambda},$$

or

$$\|f-u\|_{H^{-1}_w(\Omega)}\le\frac{\mu}{2\lambda}. \tag{3.3}$$

Now, if we set $\varphi=u$ in (3.2), we find that

$$2\lambda\epsilon\langle f-u,u\rangle\le\lambda\epsilon^2\|u\|^2_{L^2(\Omega)}+\mu\epsilon|u|_{H^1_w(\Omega)}.$$

Again, first dividing by $\epsilon<0$ and letting $\epsilon\uparrow 0_-$, we obtain

$$\langle f-u,u\rangle\ge\frac{\mu}{2\lambda}|u|_{H^1_w(\Omega)}. \tag{3.4}$$

Combining (3.3) and (3.4) we find that

$$\frac{\mu}{2\lambda}|u|_{H^1_w(\Omega)}\le\langle f-u,u\rangle\le\|f-u\|_{H^{-1}_w(\Omega)}|u|_{H^1_w(\Omega)}\le\frac{\mu}{2\lambda}|u|_{H^1_w(\Omega)},$$

confirming that the last inequalities are in fact equalities and thus concluding the proof. ∎

As before, consult Corollary 2.5, the fact that the image f contains too much texture is linked to a trivial $\mathcal{AT}_\rho$-minimizer. One part of this link is the content of the following theorem.

THEOREM 3.3. *If $u\equiv 0$ and $w\equiv 1$ are minimizers of the $\mathcal{AT}_\rho$-energy, then*

$$\|f\|_{H^{-1}(\Omega)}\le\frac{\mu\sqrt{1+\rho\epsilon_\rho}}{2\lambda}. \tag{3.5}$$

Proof. If $[u,w]\equiv[0,1]$ is a $\mathcal{AT}_\rho$-minimizer then for any $\varphi\in H^1(\Omega)$,

$$\lambda\|f\|^2_{L^2(\Omega)}\le\epsilon\mu\sqrt{\int_\Omega(1+\rho\epsilon_\rho)|\nabla\varphi|^2dx}+\lambda\|f-\epsilon\varphi\|^2_{L^2(\Omega)},$$

or, after expanding terms,

$$2\epsilon\lambda\langle f,\varphi\rangle\le\epsilon^2\lambda\|\varphi\|^2_{L^2(\Omega)}+\epsilon\mu\sqrt{\int_\Omega(1+\rho\epsilon_\rho)|\nabla\varphi|^2dx}.$$

As $\epsilon \downarrow 0_+$, we deduce that

$$2\lambda\langle f,\varphi\rangle \le \mu\sqrt{\int_\Omega (1+\rho\epsilon_\rho)|\nabla\varphi|^2 dx},$$

and (3.5) follows. ∎

Equipped with the characterization of $\mathcal{AT}_\rho$-minimizers, we turn to analyze the corresponding hierarchical decomposition.

THEOREM 3.4. (**i**)(Hierarchical decomposition). *Consider $f\in L^2(\Omega)$. Then f admits the following hierarchical decomposition:*

$$f \cong \sum_{j=0}^{\infty} u_j, \tag{3.6a}$$

where $\cong$ is interpreted as weak H_w^{-1}-convergence of the residuals

$$\|f-\sum_{j=0}^{k} u_j\|_{H^{-1}_{w_k}(\Omega)} = \frac{\mu}{\lambda_0 2^{k+1}}. \tag{3.6b}$$

Here, w_k is computed recursively as the weighting minimizer of $\mathcal{AT}_\rho(f-\sum_{j=0}^{k} u_j, \lambda_0 2^j)$.

(**ii**)(Energy decomposition). *The following 'energy' estimate holds:*

$$\sum_{j=0}^{\infty}\Big[\frac{1}{\lambda_j}|u_j|_{H^1_{w_j}(\Omega)} + \|u_j\|_{L^2}^2\Big] \le \|f\|_{L^2}^2, \quad \lambda_j := \lambda_0 2^j. \tag{3.7}$$

Moreover, if f is sufficiently smooth then equality holds in (3.7).

Proof. The first statement, (3.6), follows from the basic hierarchical expansion, $f=\sum_0^k u_j + v_k$, while noting that $\|v_k\|_{H^{-1}_{w_k}(\Omega)} = \mu/2\lambda_k$. For the second statement, (3.7), we begin by squaring the basic refinement step, $u_{j+1}+v_{j+1}=v_j$,

$$\|v_{j+1}\|_{L^2}^2 + \|u_{j+1}\|_{L^2}^2 + 2\langle u_{j+1}, v_{j+1}\rangle = \|v_j\|_{L^2}^2, \quad j=-1,0,1,\ldots. \tag{3.8}$$

Observe that the last equality holds for $j=-1$ with v_{-1} interpreted as $v_{-1}:=f$. We recall that $[u_{j+1},v_{j+1}]$ is a minimizing pair for $\mathcal{AT}_\rho(v_j,\lambda_{j+1})$ and hence, by (3.1),

$$2\langle u_{j+1}, v_{j+1}\rangle = \frac{\mu}{\lambda_{j+1}}|u_{j+1}|_{H^1_{w_{j+1}}(\Omega)},$$

yielding $\frac{\mu}{\lambda_{j+1}}|u_{j+1}|_{H^1_{w_{j+1}}(\Omega)} + \|u_{j+1}\|_{L^2}^2 = \|v_j\|_{L^2}^2 - \|v_{j+1}\|_{L^2}^2$. We sum up obtaining

$$\begin{aligned}\sum_{j=0}^{k}\Big[\frac{\mu}{\lambda_j}|u_j|_{H^1_{w_j}(\Omega)} + \|u_j\|_{L^2}^2\Big] &= \sum_{j=-1}^{k-1}\Big[\frac{\mu}{\lambda_{j+1}}|u_{j+1}|_{H^1_{w_{j+1}}(\Omega)} + \|u_{j+1}\|_{L^2}^2\Big]\\ &= \|v_{-1}\|_{L^2}^2 - \|v_k\|_{L^2}^2 = \|f\|_{L^2}^2 - \|v_k\|_{L^2}^2 \le \|f\|_2^2.\end{aligned}$$

∎

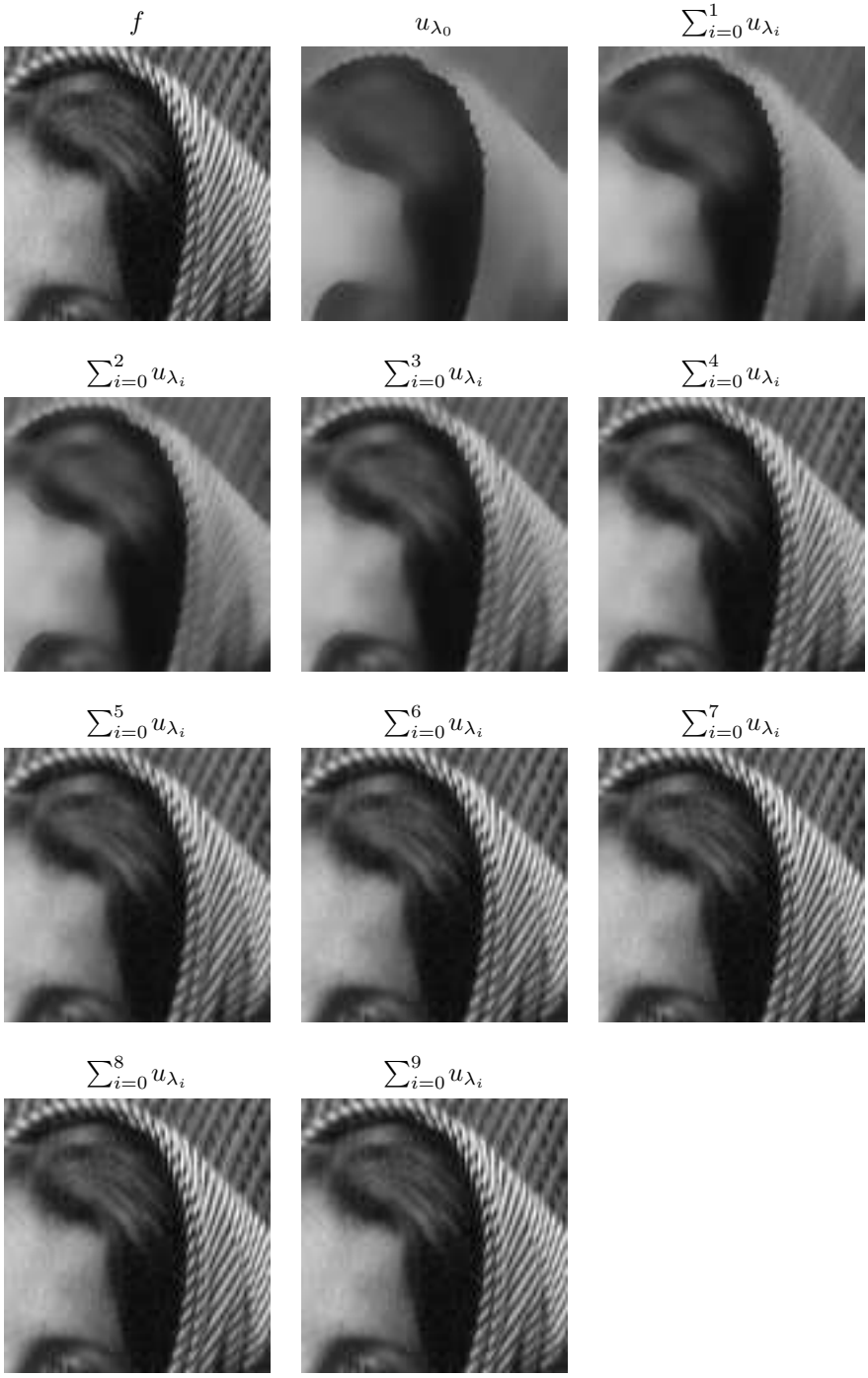

FIG. 3.1. *The sum of the u_i's using the Ambrosio-Tortorelli approximation of the image of a woman, using 10 steps. Parameters: $\lambda_0 = .25$, $\mu = 5$, $\rho = .0002$, and $\lambda_k = 2^k \lambda_0$.*

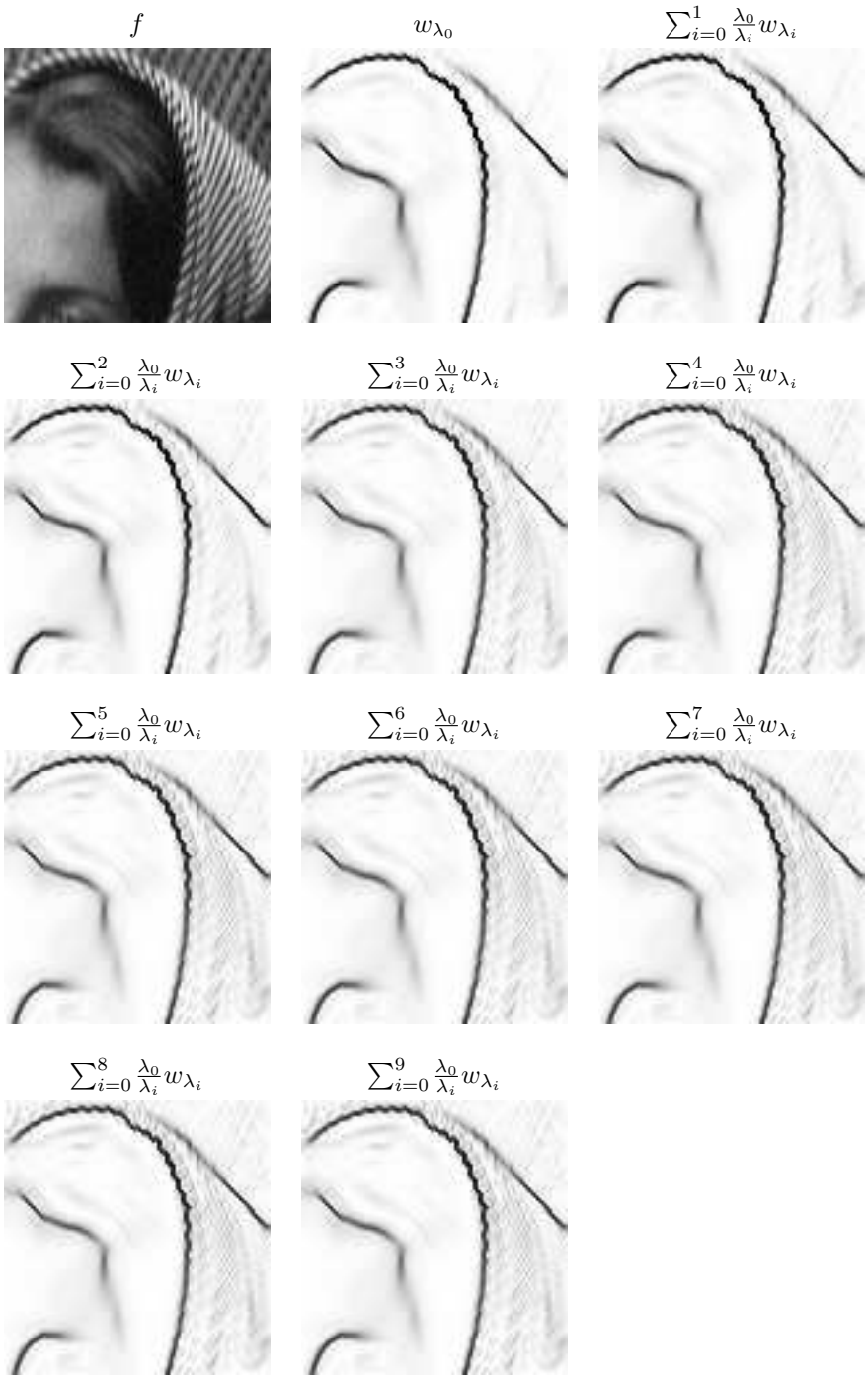

FIG. 3.2. *The weighted sum of the w_i's using the Ambrosio-Tortorelli approximation of the image of a woman, using 10 steps. Parameters: $\lambda_0 = .25$, $\mu = 5$, $\rho = .0002$, and $\lambda_k = 2^k \lambda_0$.*

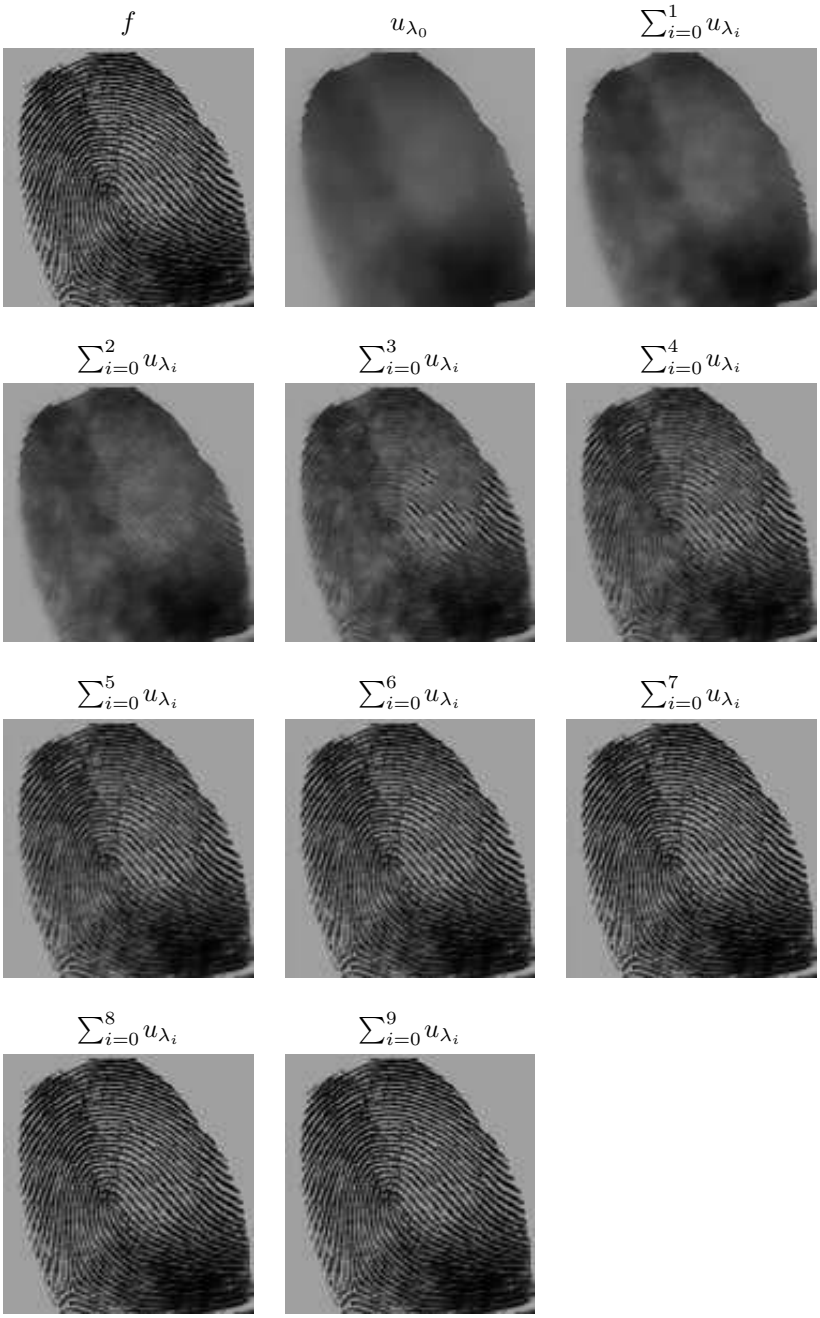

FIG. 3.3. *The sum of the u_i's using the Ambrosio-Tortorelli approximation on the image of a fingerprint. Parameters: $\mu = 5$, $\rho = .0002$, $\lambda_0 = .25$, $k = 10$, and $\lambda_k = 2^k \lambda_0$.*

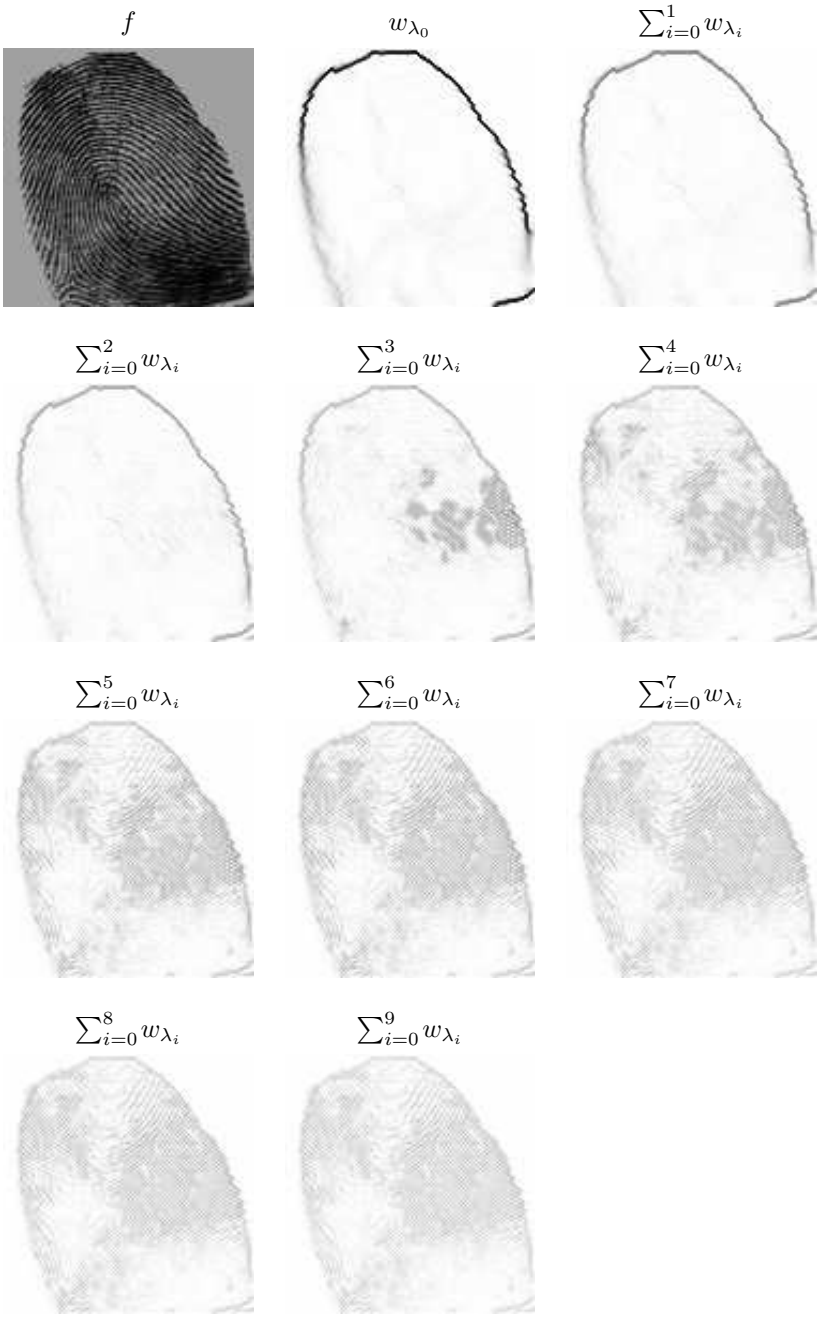

FIG. 3.4. *The sum of the w_i's using the Ambrosio-Tortorelli approximation on the image of a fingerprint. Parameters: $\mu = 5$, $\rho = .0002$, $\lambda_0 = .25$, $k = 10$, and $\lambda_k = 2^k \lambda_0$.*

3.2. Discretization of Euler-Lagrange equations and numerical results. We consider the (original) Ambrosio-Tortorelli functional [2] (neglecting the term $\rho\epsilon_\rho$):

$$\inf_{\{u,v,w \mid u+v=f\}} \left\{ \mu \int_\Omega w^2 |\nabla u|^2 dx + \rho \|\nabla w\|^2_{L^2(\Omega)} + \frac{\|w-1\|^2_{L^2(\Omega)}}{4\rho} + \lambda \|v\|^2_{L^2(\Omega)} \right\}.$$

The associated Euler-Lagrange equations are

$$\begin{cases} \lambda u_\lambda - \mu \nabla (w_\lambda^2 \nabla u_\lambda) = \lambda f, \\ \\ -\triangle w_\lambda + \dfrac{1+4\mu\rho|\nabla u_\lambda|^2}{4\rho^2} \left(w_\lambda - \dfrac{1}{1+4\mu\rho|\nabla u_\lambda|^2} \right) = 0. \end{cases}$$

We construct the hierarchical decomposition in the same manner as before, so $f \sim \sum_{j=0}^k u_j$ with the additional feature that the accumulated w_j's detect the set of edges of the image u. To this end, we discretize the Euler-Lagrange equations, yielding

$$\begin{cases} \lambda f_{i,j} = \lambda u_{i,j} - \mu D^x_- (w^2_{i,j} D^x_+ u_{i,j}) - \mu D^y_- (w^2_{i,j} D^y_+ u_{i,j}), \\ \\ D^x_- D^x_+ w_{i,j} + D^y_- D^y_+ w_{i,j} \\ \\ \quad = \dfrac{1+4\mu\rho[(D^x_0 u_{i,j})^2 + (D^y_0 u_{i,j})^2]}{4\rho^2} \left(w_{i,j} - \dfrac{1}{1+4\mu\rho[(D^x_0 u_{i,j})^2 + (D^y_0 u_{i,j})^2]} \right). \end{cases}$$

Using the notation

$$A_- := \lambda + \frac{\mu}{h^2}(2w^2_{i,j} + w^2_{i-1,j} + w^2_{i,j-1}) \tag{3.9a}$$

$$B := 1 + 4\mu\rho \sqrt{\left(\frac{u_{i+1,j} - u_{i-1,j}}{2h}\right)^2 + \left(\frac{u_{i,j+1} - u_{i,j-1}}{2h}\right)^2} + \frac{16\rho^2}{h^2}, \tag{3.9b}$$

we have

$$u_{i,j} = \frac{1}{A_-}\left[\lambda f_{i,j} + \frac{\mu}{h^2}(w^2_{i,j}(u_{i+1,j} + u_{i,j+1}) + w^2_{i-1,j} u_{i-1,j} + w^2_{i,j-1} u_{i,j-1})\right], \tag{3.10a}$$

$$w_{i,j} = \frac{1}{B}\left[1 + \frac{4\rho^2}{h^2}(w_{i+1,j} + w_{i-1,j} + w_{i,j+1} + w_{i,j-1})\right]. \tag{3.10b}$$

In order to minimize the grid effect, we alternate the above scheme with the following one, obtained by substituting D_+ for D_- (and vice-versa) in the discretization of the above Euler-Lagrange equation. Expressed in terms of

$$A_+ := \lambda + \frac{\mu}{h^2}(2w^2_{i,j} + w^2_{i+1,j} + w^2_{i,j+1}),$$

we have

$$u_{i,j} = \frac{1}{A_+}\left[\lambda f_{i,j} + \frac{\mu}{h^2}(w^2_{i,j}(u_{i-1,j} + u_{i,j-1}) + w^2_{i+1,j} u_{i+1,j} + w^2_{i,j+1} u_{i,j+1})\right], \tag{3.11a}$$

with the same equation for the $w_{i,j}$'s as before,

$$w_{i,j} = \frac{1}{B}\Big[1 + \frac{4\rho^2}{h^2}(w_{i+1,j} + w_{i-1,j} + w_{i,j+1} + w_{i,j-1})\Big]. \tag{3.11b}$$

REMARK 3.5. (On the edge detector $w_{i,j}$). Recall that the AT-weight $1 - w_\lambda$ is supported along the edges of f. Similarly, the discrete AT-weights, $1 - w_{i,j}$, provide a multiscale decomposition for discrete edge indicators, in the sense that they vary between being ≈ 0 in regions of smoothness, and being ≈ 1 near edges, depending on how large the local gradients of u are. Indeed, consider the fixed-point iterations for the w-part of the Euler-Lagrange equations (3.9b), (3.10b), (3.11b), which we rewrite as

$$w_{i,j}^{n+1} = \frac{1}{1 + 4\mu\rho\sqrt{(D_0^x u_{i,j})^2 + (D_0^y u_{i,j})^2} + \dfrac{16\rho^2}{h^2}}\Big[1 + \frac{16\rho^2}{h^2}\overline{w}_{i,j}^n\Big],$$

where $\overline{w}_{i,j}^n := \frac{1}{4}(w_{i+1,j}^n + w_{i-1,j}^n + w_{i,j+1}^n + w_{i,j-1}^n)$ is the local average. It follows that the $w_{i,j}$'s remain in $[0,1]$ with a limiting value

$$w_{i,j}^n \overset{n\to\infty}{\longrightarrow} \frac{4\rho}{4\rho + \mu h^2\sqrt{(D_0^x u_{i,j})^2 + (D_0^y u_{i,j})^2}}.$$

Thus, in smooth regions where $\big(|D_0^x u_{i,j}| + |D_0^y u_{i,j}|\big)h^2/\rho \ll 1$, the discrete weights, $1 - w_{i,j}$, are of order $h^2/\rho \approx 0$, and near edges where $\big(|D_0^x u_{i,j}| + |D_0^y u_{i,j}|\big)h^2/\rho \gg h$, the discrete weights $1 - w_{i,j} \gg h$.

In Figures 3.1 and 3.3 we demonstrate the hierarchical $\mathcal{AT}_\rho$-decompositions. We can clearly see the convergence to the respective images, while in Figures 3.2 and 3.4 we obtain the hierarchical representation, $\sum 2^{-j} w_j$, of the contours of the corresponding images.

4. Hierarchical decomposition of images with multiplicative noise

Following [31], [28], we consider a *multiplicative* degradation model where we are given an image $f = u \cdot v$, with $u > 0$ being the original image and with v modeling the multiplicative noise, normalized such that $\int_\Omega v(x,y)dxdy = 1$, where for simplicity we assume that $|\Omega| = 1$. Let u_λ be the minimizer of the corresponding total variation functional in the multiplicative case [31], [9] (in a simplified form, without a mean constraint),

$$\mathcal{M}(f,\lambda;BV,L^2) := \inf_{u\in BV_+(\Omega)}\Big\{\lambda\Big\|\frac{f}{u} - 1\Big\|^2_{L^2(\Omega)} + |u|_{BV(\Omega)}\Big\}. \tag{4.1}$$

If $f > 0$ a.e. in Ω, then (4.1) has at least one minimizer $u \geq 0$ [9].

Setting $v_\lambda := \frac{f}{u_\lambda}$ we end up with the one-scale decomposition $f = u_\lambda v_\lambda$. We construct the hierarchical decomposition as in TNV, except that sums and differences are replaced by products and quotients. Thus, the iterative step at scale $\lambda_j = \lambda_0 2^j$ reads $v_j = u_{j+1} v_{j+1}$, leading to the multiplicative hierarchical decomposition

$$f = u_0 u_1 \cdots u_k \times v_k, \qquad \lambda_j = \lambda_0 2^j.$$

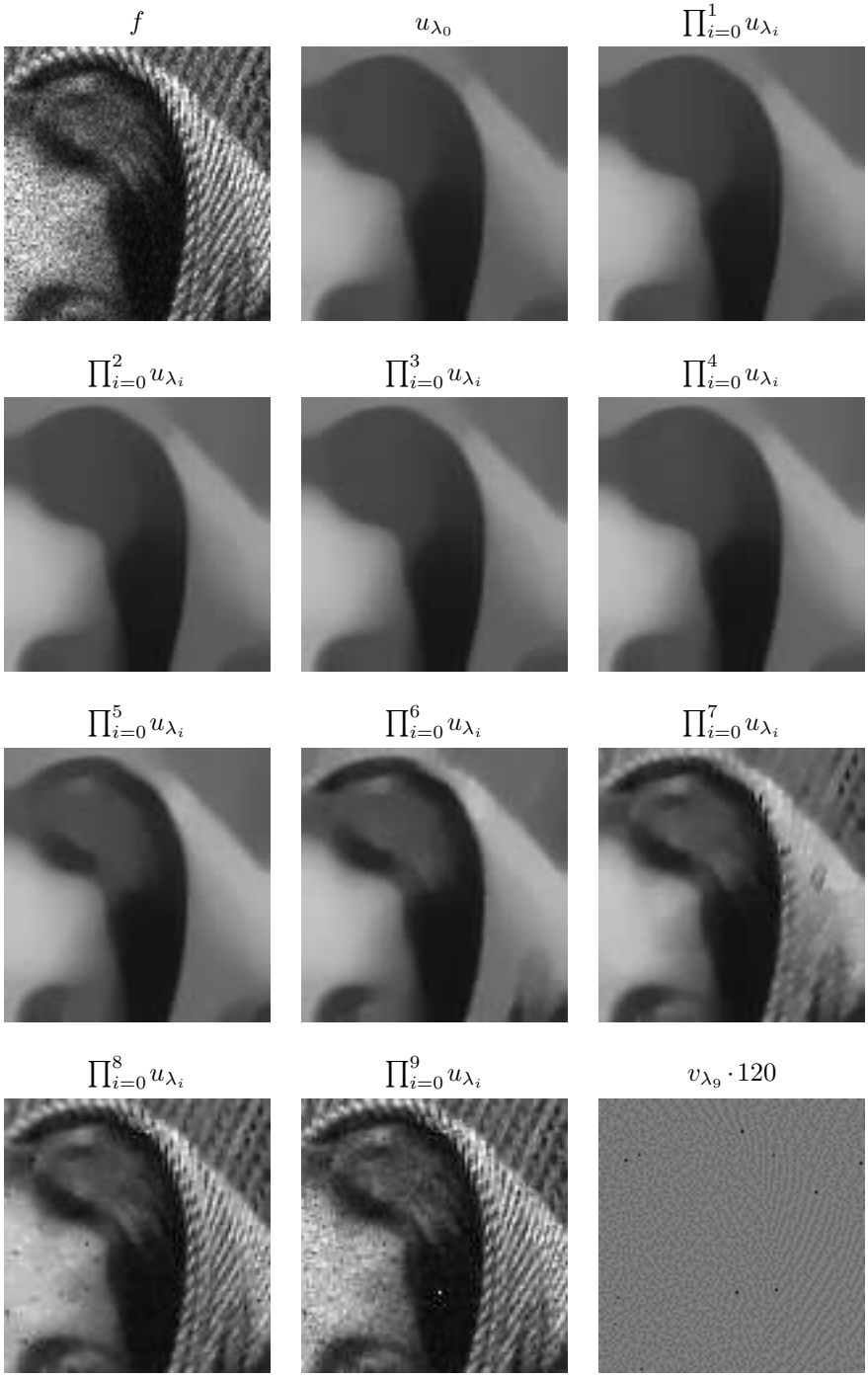

FIG. 4.1. *The recovery of u given an initial image of a woman with multiplicative noise, for 10 steps. Parameters:* $\lambda_0 = .02$, *and* $\lambda_k = 2^k \lambda_0$.

4.1. Characterization of $\mathcal{M}$-minimizers. We begin with characterization of the $\mathcal{M}(f,\lambda;BV,L^2)$-minimizers. We show that $[u_\lambda,v_\lambda]$ is a minimizer if it is an extremal pair, properly interpreted in terms of the dual BV norm $\|\cdot\|_*$ (consult [20, Definition 10]).

THEOREM 4.1. *If u is a minimizer of (4.1), then*

$$|u|_{BV(\Omega)}\cdot\Big\|\frac{f}{u^2}\Big(\frac{f}{u}-1\Big)\Big\|_*=\frac{|u|_{BV(\Omega)}}{2\lambda}\quad\text{and}\quad\Big\langle\frac{f}{u^2}\Big(\frac{f}{u}-1\Big),u\Big\rangle=\frac{1}{2\lambda}|u|_{BV(\Omega)}. \tag{4.2}$$

Proof. Let

$$g(\epsilon):=\Big(\frac{f}{u+\epsilon\varphi}-1\Big)^2. \tag{4.3}$$

Taylor's expansion gives $g(\epsilon)=\Big(\frac{f}{u}-1\Big)^2-2\epsilon\Big(\frac{f}{u}-1\Big)\frac{f\varphi}{u^2}+\frac{\epsilon^2}{2}g''(\epsilon_\xi)$ and hence

$$\lambda\Big\|\frac{f}{u+\epsilon\varphi}-1\Big\|^2_{L^2(\Omega)}\le\lambda\Big\|\frac{f}{u}-1\Big\|^2_{L^2(\Omega)}-2\lambda\epsilon\Big\langle\frac{f}{u}-1,\frac{f\varphi}{u^2}\Big\rangle+\lambda\frac{\epsilon^2}{2}\max_x|g''(x)|.$$

This inequality, together with the fact that u as an $\mathcal{M}$-minimizer satisfies, for all $\varphi\in BV(\Omega)$,

$$\lambda\Big\|\frac{f}{u}-1\Big\|^2_{L^2(\Omega)}+|u|_{BV(\Omega)}\le\lambda\Big\|\frac{f}{u+\epsilon\varphi}-1\Big\|^2_{L^2(\Omega)}+|u+\epsilon\varphi|_{BV(\Omega)},$$

implies that

$$2\lambda\epsilon\Big\langle\frac{f}{u}-1,\frac{f\varphi}{u^2}\Big\rangle\le\epsilon|\varphi|_{BV(\Omega)}+\lambda\frac{\epsilon^2}{2}\max_x|g''(x)|. \tag{4.4}$$

Dividing by ϵ and letting $\epsilon\downarrow 0_+$ (while noticing that $\lim_{\epsilon\to0}\frac{\epsilon^2}{2}\max_x|g''(x)|=0$), yield that for any $\varphi\in BV(\Omega)$,

$$2\lambda\Big\langle\frac{f}{u^2}\Big(\frac{f}{u}-1\Big),h\Big\rangle\le|\varphi|_{BV(\Omega)};$$

thus

$$\Big\|\frac{f}{u^2}\Big(\frac{f}{u}-1\Big)\Big\|_*\le\frac{1}{2\lambda}. \tag{4.5}$$

Now let $\varphi=u$ in (4.4). Then, dividing by $\epsilon<0$, and letting $\epsilon\uparrow0_-$, we obtain

$$\Big\langle\frac{f}{u^2}\Big(\frac{f}{u}-1\Big),u\Big\rangle\ge\frac{|u|_{BV(\Omega)}}{2\lambda}. \tag{4.6}$$

Combining (4.5) and (4.6) concludes the proof. ∎

As before, Theorem 4.1 could be refined, depending on the amount of texture present in f. For example, compared with Corollary 2.5, we have that f consists mostly of texture, if and only if its $\mathcal{M}$-minimizer is the trivial one which, in the multiplicative case, is given by $u\equiv1$. The following theorem confirms one direction of this implication.

THEOREM 4.2. *If $u\equiv 1$ is a minimizer of (4.1), then $\|f(f-1)\|_*\le\frac{1}{2\lambda}$.*

Proof. Since $u\equiv 1$ is assumed to be an $\mathcal{M}$-minimizer, we have for any $\varphi\in BV(\Omega)$ that

$$\lambda\Big\|\frac{f}{1+\epsilon\varphi}-1\Big\|^2_{L^2(\Omega)}+|1+\epsilon\varphi|_{BV(\Omega)}\ge\lambda\|f-1\|^2_{L^2(\Omega)}.$$

Revisiting $g(\epsilon)$ in (4.3) with $u\equiv 1$, we have

$$g(\epsilon):=\Big(\frac{f}{1+\epsilon\varphi}-1\Big)^2=(f-1)^2-2\epsilon(f-1)f\varphi+\frac{\epsilon^2}{2}g''(\epsilon_\xi).$$

The last two relations yield

$$\lambda\|f-1\|^2_{L^2(\Omega)}-2\lambda\epsilon\langle\varphi,f(f-1)\rangle+\lambda\epsilon^2\int_\Omega\frac{\varphi^2}{2}g''(1)dx \tag{4.7}$$

$$+\lambda\int_\Omega\mathcal{O}((\epsilon\varphi)^3)dx+\cdots+\epsilon|\varphi|_{BV(\Omega)}\ge\lambda\|f-1\|^2_{L^2(\Omega)}. \tag{4.8}$$

Divide by ϵ and let $\epsilon\downarrow 0_+$ to obtain that for all $\varphi\in BV(\Omega)$

$$|\varphi|_{BV(\Omega)}\ge 2\lambda\langle\varphi,f(f-1)\rangle,$$

which means that $\|f(f-1)\|_*\le\frac{1}{2\lambda}$. ∎

4.2. Discretization of Euler-Lagrange equation and numerical results. Formally minimizing $\mathcal{M}(f,\lambda)$ with respect to u yields the following associated Euler-Lagrange equation:

$$\Big(\frac{f}{u_\lambda}-1\Big)\cdot\Big(-\frac{f}{u_\lambda^2}\Big)=\frac{1}{2\lambda}\ \mathrm{div}\ \Big(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\Big).$$

When working on a bounded domain Ω, we augment the Euler-Lagrange equation by the following Neumann boundary condition:

$$\frac{\partial u_\lambda}{\partial n}\Big|_{\partial\Omega}=0.$$

The hierarchical decomposition, $f\sim\prod_{j=0}^k u_j$, is obtained, in which the u_j's are constructed as (approximate) solutions of the recursive relation governed by the Euler-Lagrange equation.

To discretize the Euler-Lagrange equation, we begin by regularization of $\mathcal{M}(f,\lambda)$ to avoid the singularity when $|\nabla u_\lambda|=0$. So, we have

$$\mathcal{M}^\epsilon(f,\lambda):=\inf_{u\in BV}\Big\{\lambda\Big\|\frac{f}{u}-1\Big\|^2_{L^2(\Omega)}+\int_\Omega\sqrt{\epsilon^2+|\nabla u|^2}dxdy\Big\}.$$

This yields the associated Euler-Lagrange equation:

$$-\frac{f^2}{u_\lambda^3}+\frac{f}{u_\lambda^2}=\frac{1}{2\lambda}\mathrm{div}\Big(\frac{\nabla u_\lambda}{\sqrt{\epsilon^2+|\nabla u_\lambda|^2}}\Big), \tag{4.9}$$

which we solve by a dynamic scheme $(x,y,t)\mapsto u(x,y,t)$:

$$\frac{\partial u}{\partial t}=\frac{f^2}{u^3}-\frac{f}{u^2}+\frac{1}{2\lambda}\mathrm{div}\Big(\frac{\nabla u}{\sqrt{\epsilon^2+|\nabla u|^2}}\Big),\quad u(x,y,0)=f(x,y). \tag{4.10}$$

Let $u_{i,j}^n \approx u(x_i, y_j, n\triangle t)$. The discretization that we have used is a linearized semi-implicit scheme:

$$\frac{u_{i,j}^{n+1}-u_{i,j}^n}{\triangle t} = \frac{f_{i,j}^2}{(u_{i,j}^n)^3} - \frac{f_{i,j}}{(u_{i,j}^n)^2} + \frac{\Delta t}{2\lambda}\left[\frac{u_{i+1,j}^n - u_{i,j}^{n+1}}{h|\nabla u^n|_{i,j}} - \frac{u_{i,j}^{n+1}-u_{i-1,j}^n}{h|\nabla u^n|_{i-1,j}}\right] + \frac{1}{2\lambda}\left[\frac{u_{i,j+1}^n - u_{i,j}^{n+1}}{h|\nabla u^n|_{i,j}} - \frac{u_{i,j}^{n+1}-u_{i,j-1}^n}{h|\nabla u^n|_{i,j-1}}\right],$$

where $|\nabla u^n|_{i,j}$ is the "numerical gradient" used before in (2.19).

We note in passing the issue of *stability*: in order to enable the necessary division by $u \neq 0$, we shift f away from zero, adding a positive constant which is subtracted from the final result.

We demonstrate our hierarchical decomposition to the image f in Figure 4.1. We can see that just as in the case with additive noise, we must pay a price for the recovered texture, namely the return of some noise. As in the case with additive noise, using a finer decomposition might give improved results.

Acknowledgments. The research of E. Tadmor was supported in part by ONR Grant # N00014-91-J-1076 and NSF grant #DMS07-07949. The research of L. Vese was supported in part by NSF grants 0312222 and 0714945, NIH grant #P20MH65166, and by a Sloan Fellowship. Part of the research was carried out while S. Nezzar and L. Vese were visiting the Center for Scientific Computation and Mathematical Modeling (CSCAMM) at the University of Maryland, College Park.

REFERENCES

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

[2] L. Ambrosio and V. Tortorelli, *On the approximation of free discontinuity problems*, Bollettino U.M.I., (7)6-B, 105–123, 1992.

[3] C. Bennet and R. Sharpley, *Interpolation of Operators*, Academic Press, 1988.

[4] J. Bergh and J. Lofstrom, *Interpolation Spaces*, Springer Verlag, 1976.

[5] F. Andreu-Vaillo, V. Caselles and J.M. Mazón, *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals*, Birkhäuser Verlag, 2004.

[6] G. Aubert and L. Vese, *A variational method in image recovery*, SIAM J. Numer. Anal., 34, 1948–1979, 1997.

[7] A.Z. Averbuch, R.R. Coifman, F.G. Meyer and J.O. Strömberg, *Multi-layered image representation: application to image compression*, in Proceedings of the International Conference on Image Processing (ICIP 1998), Chicago, IL, 1998, IEEE Computer Society Press, Los Alamitos, CA, 2, 292–296, 1998.

[8] A. Buades, B. Coll and J.M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4(2), 490–530, 2005.

[9] A. Chambolle and P.L. Lions, *Image restoration by constrained total variation minimization and variants*, SPIE Electronic Imaging Proceedings, Vol. 2567, 1995.

[10] A. Chambolle and P.L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76, 167–188, 1997.

[11] F. Demengel and R. Temam, *Convex functions of a measure and applications*, Indiana Univ. Mathematics Journal, 33(5), 673–709, 1984.

[12] R.A. DeVore and G.G. Lorentz, *Constructive Approximation, Springer-Verlag*, Grundlehren der mathematischen Wissenschaften, 303, 1993.

[13] S. Esedoglu and S. Osher, *Decomposition of images by the anisotropic Rudin-Osher-Fatemi model*, Comm. Pure Appl. Math., LVII, 1609–1626, 2004.

[14] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721–741, 1984.

[15] C.W. Groetsch and O. Scherzer, *Inverse scale space theory for inverse problems*, Lecture Notes in Computer Science; Proc. Third International Conference on Scale-Space and Morphology in Computer Vision, 2106, 317–325, 2001.
[16] J. Jost, *Partial Differential Equations*, Springer, 2007.
[17] N. Kruglyak, *The K-functional and Calderón-Zygmund type decompositions*, in Interpolation Theory and Applications, L. De Carli and M. Milman, eds, Contemporary Mathematics, Amer. Math. Soc., 2007.
[18] D. Levy, D. Paquin, E. Schreibmann and L. Xing, *Multiscale image registration*, Mathematical Biosciences and Engineering, 3, 389–418, 2006.
[19] F.G. Meyer, A.Z. Averbuch and R.R. Coifman, *Multilayered image representation: application to image compression*, IEEE Trans. Image Process., 11, 1072–1080, 2002.
[20] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, University Lecture Series Volume 22, AMS, 2002.
[21] V.A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer Verlag, Heidelberg, 1984.
[22] V.A. Morozov, *Regularization Methods for Ill-Posed Problems*, Florida, CRC Press, 1993.
[23] D. Mumford and J. Shah, *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems*, Communications on Pure and Applied Mathematics, Vol. XLVII, 577–685, 1989.
[24] S. Nezzar, *A Multiscale Image Representation Using Hierarchical* (BV, L^2) *Decompositions*, Ph.D. Thesis, UCLA, May, 2003.
[25] S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4(2), 460–489, 2005.
[26] L. Rudin and V. Caselles, *Image Recovery via Multiscale Total Variation*, in Proceedings of the Second European Conference on Image Processing, Palma, Spain, 1995.
[27] D. Paquin, D. Levy and L. Xing, *Hybrid multiscale landmark and deformable registration*, Mathematical Biosciences and Engineering, 4, 711–737, 2007.
[28] L. Rudin, P.L. Lions and S. Osher, *Multiplicative denoising and deblurring: theory and algorithms*, in S. Osher and N. Paragios, Editors, Geometric Level Set Methods in Imaging, Vision and Graphics, Springer, 103–119, 2003.
[29] L. Rudin, *Images, Numerical Analysis of Singularities and Shock Filters*, Ph.D. Thesis, CS Dept., CalTech, Pasadena, 1987.
[30] L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60, 259–268, 1992.
[31] L. Rudin and S. Osher, *Total variation based image restoration with free local constraints*, Proc. IEEE ICIP, Austin (Texas) USA, Vol. I, 31–35, 1994.
[32] A.J. Smola and B. Scholkop, *A tutorial on support vector regression*, Stat. and Computing, 14, 199–222, 2004.
[33] E. Tadmor, S. Nezzar and L. Vese, *A multiscale image representation using hierarchical (BV,L2) decomposition*, Multiscale Modeling and Simulation: A SIAM Inter. J., 2(4), 554–579, 2004.
[34] A.N. Tikhonov and V.A. Arsenin, *Solution of Ill-posed Problems*, Winston & Sons, Washington, 1977.
[35] V. Vapnik, *Statistical Learning Theory*, John Wiley and Sons, 1998.
[36] L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Appl. Math. Optim., 44, 131–161, 2001.