University of California
Los Angeles

A Multiscale Image Representation Using Hierarchical (BV, L^2) Decompositions

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Suzanne Nezzar

2003

The dissertation of Suzanne Nezzar is approved.

Luminita A. Vese

Stanley J. Osher

Tony F. Chan

Henry Huang

Eitan Tadmor, Committee Chair

University of California, Los Angeles

2003

To my family -

May Allah bless them always

Table of Contents

List of Figures

Acknowledgments

In the name of Allah, the most Gracious, the most Merciful

My first praise is to Allah, for His countless blessings, including my love for mathematics. Without His help and guidance none of this would be possible.
Next, I would like to thank my family: my parents, Amar and Cynthia, my siblings, Mona and Ali, and my husband, Hafid, for their faith and support.
I would like to extend my gratitude to my advisor, Eitan Tadmor, for his advice and guidance throughout my graduate studies. His enthusiasm and encouragement were a blessing. In my last year here at UCLA, Luminita Vese stepped in and became my second advisor while Eitan was away, and for that I would like to express my deepest appreciation; without your help I wouldn't have been able to finish. Next, I would like to thank Stanley Osher, who not only co-developed the Total Variation Minimization method I use in my work, but also suggested Luminita as an advisor to me. Also, I would like to extend my thanks to Tony Chan and Henry Huang for giving their time to be my committee members. Another professor I would like to acknowledge is Robert Brown, with whom I had the pleasure of training new teaching assistants. For all of the professors and teaching assistants I taught with, it was a pleasure to work with you.
Now for all of my roommates these past years: Mona Nezzar, Banafshe Sharifian, Fizzah Raza, Nessreen Sharaf, May Meroueh, Faeka Husain, Amina Inloes, Nargis, Roya, Edina Lekovic, Sawssan Ahmed, and my roommates down the hall: Uzma Naz, Safia Mullick, Fauzia Khan and Heena Hameed - you are all truly my sisters, jazakum Allah khair.
Also, I would like to thank all of my MSA UCLA sisters and brothers, it has been my pleasure to be amongst you - you are my family away from home. May Allah

bless you all.

I am also grateful to my colleagues: Jared Tanner, Sung Ha Kang, Rachel Caiden, Erik Downs, Charles Garcia, Heidi Graziano, Isabelle Saber, Pete Sung, Wayland Oong, and my officemates: Heidi Hu, Lyman Chayfee and Bill Cowieson. You have all helped me in some way to make it through, so thank you.

I would also like to thank all of the people who work in the Mathematics Department here at UCLA for all of your help, and I would especially like to extend my appreciation to Magda Albert, Stephanie Stern, Anna Ylvisaker, Babette Dalton and Mary Edwards for going out of their way for me.

I thank Allah for knowing all of you, it has been a blessing. May Allah bless all of you.

Vita

1975	Born, Hayward, California, USA
1995-1996	Leaders for Tomorrow Santa Clara University
1996	George W. Evans Memorial Prize Santa Clara University
1996-1997	Leaders for Tomorrow Santa Clara University
1997	George W. Evans Memorial Prize Santa Clara University
1997	Orella Prize Santa Clara University
1997	B.S. Mathematics, B.S. Computer Science Santa Clara University
1997-2001	Teaching and Research Assistant Department of Mathematics, UCLA
2001	Teaching Assistant Coordinator Department of Mathematics, UCLA
2002	Teaching Assistant Department of Mathematics, UCLA
2002	Robert J. Sorgenfrey Distinguished Teaching Award Department of Mathematics, UCLA

2002-2003 Dissertation Year Fellowship
Graduate Division, UCLA

Publications

Eitan Tadmor, Suzanne Nezzar and Luminita Vese, *A Multiscale Image Representation Using Hierarchical* (BV, L^2) *Decompositions.* UCLA CAM Report, June 2003.

Abstract of the Dissertation

A Multiscale Image Representation Using Hierarchical (BV, L^2) Decompositions

by

Suzanne Nezzar

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2003

Professor Eitan Tadmor, Chair

We propose a new model for multiscale image representation using hierarchical (BV, L^2) decompositions. We begin with the total variation minimization model of Rudin-Osher-Fatemi, in which we take a given image $f \in L^2(\Omega)$ and decompose it into the sum $u+v$, where $u \in BV(\Omega)$ is the minimizer of the Rudin-Osher-Fatemi functional and $v \in L^2(\Omega)$ is the residual (i.e. $v := f - u$). This decomposition depends on a parameter λ. We propose to iterate this process for different monotone values of λ, applying the decomposition to the residual of the previous step. In this manner, we obtain a multiscale representation of f. We will discuss some theoretical properties of this method as well as show numerical results of this new decomposition applied to both synthetic and real images.

We have also applied our method to color (vector-valued) images, blurred images and images with multiplicative noise. Finally, we will show that this decomposition can also be applied to other variational models by showing some results for a multiscale Mumford-Shah/Ambrosio-Tortorelli image segmentation model.

CHAPTER 1

Introduction

The field of image processing is full of many interesting tasks. We see its application in many different fields such as medical imaging, astronomy, computer graphics, and security identification. Among the tasks performed in image processing are image restoration, decomposition, segmentation and compression. Our research has been focused on image decomposition and image restoration of noisy and/or blurry images.

There are many different approaches to work in image processing. There are Fourier and Wavelet transformations which have inspired the JPEG image representation. These methods have also led us to new methods such as Ridgelets and Curvelets. Also, there are statistical methods used for texture analysis. Another approach to image processing is Variational PDE methods. In our research we use this last method. There are many advantages to using PDEs. First, since we consider the image as a continuous function, we are allowed to use concepts such as curvature, gradients, diffusion and level sets. Also, it has been shown that these methods give sharp edges between homogeneous regions.

Images could be realized as general L^2-objects, $f \in L^2$, representing the grayscale of the observed image. Likewise, color images are typically realized in terms of vector-valued L^2-RGB scale functions, $f = (f_1, f_2, f_3) \in L^2$. In practice, the more notable features of images could be identified within a proper subclass

of all L^2 objects. Most notable are the edges of an image, which are known to be well quantified within the smaller subclass of functions of Bounded Variation (BV). Quantifying the precise L^2 subclasses for these different features is still the subject of current research. We argue that a large class of notable images form an 'intermediate' space between the larger $L^2(\mathbb{R}^2)$ and the smaller $BV(\mathbb{R}^2)$. The standard tool for studying such 'intermediate' spaces is interpolation. To this end, one starts with a pair of given spaces, $Y \subset X$, and forms a scale of intermediate spaces, $(X,Y)_\theta, \theta \in [0,1]$, ranging from $(X,Y)_{\theta=0} = X$ to $(X,Y)_{\theta=1} = Y$, expressed in terms of the appropriate K-functional

$$K(f,\lambda) \equiv K(f,\lambda;X,Y) := \inf_{u+v=f} \Big\{ \|v\|_X + \lambda \|u\|_Y \Big\}.$$

In this dissertation, the K-functional is replaced by the closely related J-functional of the form

$$J_p(f,\lambda;X,Y) := \inf_{u+v=f} \Big\{ \lambda \|v\|_X^p + \|u\|_Y \Big\}.$$

The functional $J_2(f,\lambda)$ measures how well an L^2 object can be approximated by its BV features, $J_2(f,\lambda) \sim \lambda^\theta$ as $\lambda \uparrow \infty$. The classical argument addresses this question of convergence rate in terms of the smoothness properties of f. In modern theory the roles are reversed: one defines the scale of smoothness spaces, $(L^2, BV)_\theta$ in terms of the behavior of $J(f,\lambda)$. Here we will introduce a new multiscale decomposition whose properties quantify images in intermediate space which we denote as 'multiscale (BV, L^2) decompositions'.

This dissertation has been organized into four chapters. We begin by reviewing some existing image restoration methods in this chapter. In chapter 2, we focus our attention on one of these methods, namely, the Rudin-Osher-Fatemi total variation method, and discuss some of its properties. In the following chapter, we propose our multiscale image representation using hierarchical decompositions

and state some of its properties. We further discuss the numerical implementation of this method and present some results in the last chapter.

1.1 Some Variational PDE Based Image Restoration Models

We consider a given observed image $f : \Omega \to \mathbb{R}$ as a function in $L^2(\Omega)$, where Ω is an open and bounded subset of $\mathbb{R}^2$ (Ω is generally a rectangle in $\mathbb{R}^2$). The image f is comprised of some combination of homogeneous regions, contours and oscillatory patterns such as noise or texture. This scalar-valued (or gray-scale) image f is represented in the following manner: at each pixel (x, y), the value $f(x, y)$ represents the intensity of the light at that pixel. In general, gray-scale images take values between 0-black and 255-white.

1.1.1 Rudin-Osher-Fatemi Total Variation Minimization

Assume f is in the following form:

$$f = Ku + v,$$

where u is the true image, K is a linear operator representing the blur, and v is additive noise of zero mean. In order to recover u, we need to have some information about K and v. One successful model for the recovery of u was proposed by Rudin-Osher-Fatemi in [21]. The objective is to minimize the following functional with respect to u:

$$F(u) = \lambda \int_\Omega (f - Ku)^2 dxdy + \int_\Omega |\nabla u| dxdy,$$

where $\lambda > 0$ acts as a scaling parameter, $\int_\Omega (f - Ku)^2 dxdy$ is a fidelity term, and $\int_\Omega |\nabla u| dxdy$ is a regularizing term called the total variation of u. If $u \in BV(\Omega)$,

then this energy is finite and has a minimizer in this space (see [10], [1], or [26] for the general case).

1.1.2 Vese-Osher Modification of Rudin-Osher-Fatemi Minimization

This method in [27] is an improvement in texture capturing of the Rudin-Osher-Fatemi method. This model decomposes the image, f, into a 'cartoon' representation, u, and texture represented by $\text{div}\vec{g}$. The goal is to minimize the following functional in terms of u, g_1 and g_2

$$F_p(u, g_1, g_2) = \int_\Omega |\nabla u| dxdy + \lambda \int_\Omega |f - (u + \text{div}\vec{g})|^2 dxdy + \mu \Big(\int_\Omega \Big(\sqrt{g_1^2 + g_2^2} \Big)^p dxdy \Big)^{\frac{1}{p}},$$

where $\vec{g} = (g_1, g_2)$, for $g_1, g_2 \in L^\infty(\Omega)$, $\lambda, \mu > 0$ are tuning parameters and $p \to \infty$. The first term insures that $u \in BV(\Omega)$, the second term gives us $f \approx u + \text{div}\vec{g}$, while the third term is a penalty on the norm in $W^{-1,p}(\Omega)$ of $\text{div}\vec{g}$. (Note that here $K = I$.)

1.1.3 Osher-Solé-Vese Modification of Rudin-Osher-Fatemi Minimization

This is another model, in [18], which improves the texture capturing of the Rudin-Osher-Fatemi model. In this method, the original image, f, is also decomposed into a 'cartoon' representation, u, and texture v, where $v := f - u$. This method minimizes the following functional with respect to u

$$\begin{aligned} F(u) &= \int_\Omega |\nabla u| dxdy + \lambda \int_\Omega |\nabla \triangle^{-1}(f - u)|^2 dxdy \\ &= \|u\|_{BV(\Omega)} + \lambda \|f - u\|^2_{H^{-1}(\Omega)}. \end{aligned}$$

(Note that $K = I$ in this case as well.)

1.1.4 Mumford-Shah Model

The Mumford-Shah functional from [17] is defined by

$$F^{MS}(u,S) = \int_{\Omega\backslash S} (\alpha|\nabla u|^2 + \beta(u-f)^2)dxdy + \mathcal{H}^1(S),$$

where u is a piecewise smooth optimal approximation of f, f is the given image, S is the set of contours of u, $\mathcal{H}^1$ is the Hausdorff 1-dimensional measure in $\mathbb{R}^2$, and $\alpha, \beta > 0$ are fixed positive parameters. We also note that $\Omega \subset \mathbb{R}^2$ is open and bounded, S varies in the class of closed subsets of Ω and u varies in $C^1(\Omega\backslash S)$. The first term imposes the condition that u be smooth over $\Omega\backslash S$, the second term is the fidelity term, and the last term establishes that the length of the contours is minimal.

By setting $S = S_u$, the jump set of u, and minimizing over u, we have existence of minimizers in a weak formulation of this problem (see [12] and [3]). The functional associated with the weak formulation is:

$$G^{MS}(u) = \int_{\Omega\backslash S_u} (\alpha|\nabla u|^2 + \beta|u-f|^2)dxdy + \mathcal{H}^1(S_u),$$

where $f \in L^\infty(\Omega)$ and each minimizer of this functional belongs to the space of piecewise C^1 functions.

1.1.5 Ambrosio-Tortorelli Approximation

Ambrosio and Tortorelli proposed two approximations to the Mumford-Shah model in [4] and [5]. We will use the simpler approximation in [5] which is

$$G_\rho^{AT}(u,w) = \int_\Omega [\rho|\nabla w|^2 + \alpha(w^2|\nabla u|^2 + \frac{(w-1)^2}{4\alpha\rho}) + \beta|u-f|^2]dxdy.$$

Here the function w represents the contours of u. They show that if (u_ρ, w_ρ) minimizes G_ρ^{AT}, then (passing to subsequences) u_ρ is an approximation of u, a

minimizer of G^{MS}, and w_ρ goes to 1, as $\rho \to 0$, in the $L^2(\Omega)$-topology (i.e. $\int_\Omega |u_\rho - u|^2 dxdy \to 0$ and $\int_\Omega |w_\rho - 1|^2 dxdy \to 0$ as $\rho \to 0$). Here w_ρ is different from 1 (and less than one) only in a small neighborhood of S_u, which shrinks as $\rho \to 0^+$.

By looking at the Euler-Lagrange equations associated with the minimizers of these variational methods, we find something interesting. Since the functionals are convex, they lead us to nonlinear PDEs, $P(u, \nabla u, \nabla^2 u) = 0$. The solution of these PDEs can be sought as the steady-state solution to the time-dependent parabolic equation $u_t = P(u, \nabla u, \nabla^2 u)$. This brings us to the following reconstruction models which are formulated directly in terms of a time-dependent equation.

1.1.6 Perona-Malik Equation

In [19], Perona-Malik suggest using a nonlinear smoothing method defined by

$$\frac{\partial u}{\partial t} = \text{div}(c(|\nabla u|^2)\nabla u),$$

with $u(0, x, y) = f(x, y)$ and where c is defined as one of the following:

$$c(s) = \frac{1}{1 + s/k}, \text{ or } c(s) = e^{-s/k},$$

and k is a constant. The main idea of this method is to smooth the homogeneous regions while enhancing the boundaries.

1.1.7 Catté-Lions-Morel-Coll Equation

In [9], Catté et al. suggest using a slightly different equation than Perona-Malik, but the idea is the same: smooth homogeneous regions and enhance the bound-

ary. Their method introduces some regularization via convolution with the Gaussian kernel, G_σ, in order to obtain a well-posed problem, and is as follows, for $u(0,x,y) = f(x,y)$,

$$\frac{\partial u}{\partial t} = \text{div}(c(|\nabla G_\sigma * u|^2)\nabla u),$$

where G_σ is defined as

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

1.1.8 Alvarez-Guichard-Lions-Morel Equation

This method from [2] is an affine invariant, contrast invariant operator

$$\frac{\partial u}{\partial t} = |\nabla u| \text{div}(\frac{\nabla u}{|\nabla u|})^{\frac{1}{3}},$$

with $u(0,x,y) = f(x,y)$. It is equivalent to curve evolution of all level lines of u and is a multiscale analysis of the data f. In this model, t is a scale parameter such that for larger t, only larger objects of $u(t,x,y)$ are kept (i.e. t is related to the size of objects).

All of the models listed in this chapter extract from a given image, f, a single image u which depends on a scale parameter which is related with the size of the objects kept in u. Now if we write $f = u + v$, then by keeping only u, we lose all of the valuable details in v. In this dissertation, we propose a new multiscale image decomposition method which considers a sequence of increasing scales in order to recover the information lost in v. (We would like to note that the size of the scaling parameter is inversely proportional to the size of the objects.)

CHAPTER 2

The Rudin-Osher-Fatemi Total Variation Decomposition

2.1 Introduction

We will begin by restating the Rudin-Osher-Fatemi method from [21]. Let f, the observed image, be defined as

$$f = Ku + v,$$

where u is the true image, K is a linear operator representing the blur, and v is additive noise of zero mean. We assume some knowledge about the type of blur, K and type of noise, v. This method is to minimize the following functional with respect to u:

$$\inf_{u \in BV(\Omega)} F(u) = \lambda \int_{\Omega} (f - Ku)^2 dxdy + \int_{\Omega} |\nabla u| dxdy, \tag{2.1}$$

where $\lambda > 0$ acts as a scaling parameter. The fidelity term, $\int_{\Omega}(f - Ku)^2 dxdy$, insures that we minimize the noise and the regularizing term, $\int_{\Omega} |\nabla u| dxdy$, is needed to make the problem well-posed. In minimizing $F(u)$, we will show that we obtain the decomposition $f = Ku + v$, where $u \in BV(\Omega)$ and $Ku, v \in L^2(\Omega)$.

2.2 The Space $BV(\Omega)$

Definition 2.2.1 *We define the space of functions of bounded variation, $BV(\Omega)$, as*

$$BV(\Omega) = \Big\{ u \in L^1(\Omega); \int_\Omega |\nabla u| < \infty \Big\},$$

where ∇u is the distributional gradient of u, and

$$\int_\Omega |\nabla u| = \sup_{\vec{\varphi}} \Big\{ \int_\Omega u \ div\ \vec{\varphi} dxdy; \vec{\varphi} \in C_0^1(\Omega; \mathbb{R}^2), |\vec{\varphi}| \leq 1 \Big\}.$$

Property 2.2.1 *$BV(\Omega)$ is a Banach space endowed with the norm $\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + \int_\Omega |\nabla u|$. However, we will not use this norm, but instead follow the notation of Meyer ([15]), and use $\|u\|_{BV(\Omega)} = \int_\Omega |\nabla u|$.*

Given this notation for $\|u\|_{BV(\Omega)}$, we can rewrite the Rudin-Osher-Fatemi functional as

$$F(u) = \lambda \|v\|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)}.$$

2.3 Existence and Uniqueness of the Rudin-Osher-Fatemi Minimizer

We will follow the existence and uniqueness proofs of a generalized form of (2.1) in [26] and [10].

Let Ω be an open, bounded, and connected subset of $\mathbb{R}^2$, with Lipschitz boundary $\partial\Omega$. We use standard notation for the Sobolev and Lebesgue spaces

$W^{1,2}(\Omega)$ and $L^2(\Omega)$. For the theoretical study of the problem, we consider $\lambda = 1$ for simplicity.

To ensure the existence and the uniqueness of the minimizer for (2.1) in $BV(\Omega)$, we make the following assumptions on K:

H1. $K : L^2(\Omega) \to L^2(\Omega)$ is a linear and continuous operator.

H2. $K\chi_\Omega \neq 0$.

H3. K is injective.

We will also use the following result:

Result 2.3.1 *For a sequence* $\{u_n\}$, *if* $\left[\|u_n\|_{L^1(\Omega)} + \int_\Omega |\nabla u_n| dxdy\right] \leq M,\ \forall n$, *then there exists* $u \in BV(\Omega)$ *and a subsequence of* $\{u_n\}$ *(still denoted* $\{u_n\}$*) such that* $u_n \to u$ *strongly in* $L^1(\Omega)$, $\nabla u_n \rightharpoonup \nabla u$ *weakly in the sense of measures and*

$$\int_\Omega |\nabla u| dxdy \leq \liminf_{n\to\infty} \int_\Omega |\nabla u_n| dxdy.$$

2.3.1 Existence

Note: In this proof we use M, a strictly positive constant which can be different from line to line.

Let $\{u_n\}_{n\geq 1}$ be a minimizing sequence for (2.1). Then $u_n \in BV(\Omega)$ and we have

$$\int_\Omega |\nabla u_n| dxdy \leq M,\ \forall n \geq 1.$$

We now want to prove that $|\int_\Omega u_n dxdy| \leq M,\ \forall n \geq 1$.

Let

$$w_n = \frac{\int_\Omega u_n}{|\Omega|}\chi_\Omega \text{ and } v_n = u_n - w_n.$$

Then $\int_\Omega v_n = 0$ and $\nabla v_n = \nabla u_n$. Hence, $\int_\Omega |\nabla v_n| dxdy \leq M$. Using the Poincaré-Wirtinger inequality, we obtain

$$\|v_n\|_{L^2(\Omega)} \leq M.$$

We also have

$$\begin{aligned} M &\geq \|Ku_n - f\|^2_{L^2(\Omega)} = \|Kv_n + Kw_n - f\|^2_{L^2(\Omega)} \\ &\geq (\|Kv_n - f\|_{L^2(\Omega)} - \|Kw_n\|_{L^2(\Omega)})^2 \\ &\geq \|Kw_n\|_{L^2(\Omega)}(\|Kw_n\|_{L^2(\Omega)} - 2\|Kv_n - f\|_{L^2(\Omega)}) \\ &\geq \|Kw_n\|_{L^2(\Omega)}[\|Kw_n\|_{L^2(\Omega)} - 2(\|K\| \cdot \|v_n\|_{L^2(\Omega)} + \|f\|_{L^2(\Omega)})]. \end{aligned}$$

Let $x_n = \|Kw_n\|_{L^2(\Omega)}$ and $a_n = \|K\| \cdot \|v_n\|_{L^2(\Omega)} + \|f\|_{L^2(\Omega)}$. Then

$$x_n(x_n - 2a_n) \leq M, \text{ with } 0 \leq a_n \leq \|K\| \cdot M + \|f\|_{L^2(\Omega)} = M', \ \forall n \geq 1.$$

Hence, we obtain

$$0 \leq x_n \leq a_n + \sqrt{a_n^2 + M} \leq M'',$$

which implies

$$\|Kw_n\|_{L^2(\Omega)} = \Big| \int_\Omega u_n dxdy \Big| \cdot \frac{\|K\chi_\Omega\|_{L^2(\Omega)}}{|\Omega|} \leq M'', \ \forall n \geq 1,$$

and thanks to assumption H2, we obtain that $|\int_\Omega u_n dxdy|$ is uniformly bounded. Again, by the Poincaré-Wirtinger inequality, we have

$$\Big\| u_n - \frac{\int_\Omega u_n}{|\Omega|} \Big\|_{L^2(\Omega)} \leq C \int_\Omega |\nabla u_n| dxdy \leq C \cdot M.$$

Finally, we obtain

$$\|u_n\|_{L^2(\Omega)} = \Big\| u_n - \frac{\int_\Omega u_n}{|\Omega|} + \frac{\int_\Omega u_n}{|\Omega|} \Big\|_{L^2(\Omega)} \leq \Big\| u_n - \frac{\int_\Omega u_n}{|\Omega|} \Big\|_{L^2(\Omega)} + \Big| \int_\Omega u_n \Big| \leq M'''.$$

Therefore, u_n is bounded in $L^2(\Omega)$ and, in particular, in $L^1(\Omega)$. Then u_n is also bounded in $BV(\Omega)$. From Result (2.3.1) and the fact that Ku_n converges weakly

to Ku in $L^2(\Omega)$ from assumption H1 we have

$$\int_\Omega (Ku - f)^2 dxdy \leq \liminf_{n\to\infty} \int_\Omega (Ku_n - f)^2 dxdy$$

and

$$\int_\Omega |\nabla u| \leq \liminf_{n\to\infty} \int_\Omega |\nabla u_n|,$$

that is to say

$$F(u) \leq \liminf_{n\to\infty} F(u_n)$$

and u is a minimum of F.

2.3.2 Uniqueness

Let u, $w \in BV(\Omega)$ be two solutions of the minimization problem (2.1).
We will first show that $Ku = Kw$. If not (i.e. $Ku \neq Kw$), then

$$F(\frac{1}{2}u + \frac{1}{2}w) < \frac{1}{2}F(u) + \frac{1}{2}F(w) = \inf F,$$

because F is the sum of two convex functions with independent variables, Ku and ∇u, the first one being strictly convex. However, this inequality cannot be true if u and w are minimizers of F. Then, $Ku = Kw$.
Now, since K is injective, we have $u = w$.

2.4 Some Mathematical Properties of the Rudin-Osher-Fatemi Model

From now on, we will assume that there is no blurring of u, or $K = I$, unless otherwise stated. First we will look at two properties from [6].

Property 2.4.1 *The L^2 norm of u_λ is bounded by a constant independent of λ.*

Proof Since $u_\lambda \in BV(\Omega)$ is the unique minimizer of (2.1), we have for all $w \in BV(\Omega)$,

$$\lambda \int_\Omega (f-u_\lambda)^2 dxdy + \int_\Omega |\nabla u_\lambda| dxdy \leq \lambda \int_\Omega (f-w)^2 dxdy + \int_\Omega |\nabla w| dxdy.$$

By letting $w = 0$, we obtain

$$\int_\Omega (f-u_\lambda)^2 dxdy + \int_\Omega |\nabla u_\lambda| dxdy \leq \int_\Omega f^2 dxdy,$$

which implies that

$$\int_\Omega (f-u_\lambda)^2 dxdy \leq \int_\Omega f^2 dxdy,$$

from which we can deduce, using the Schwarz inequality, that

$$\int_\Omega u_\lambda^2 dxdy \leq 4 \int_\Omega f^2 dxdy,$$

and hence the L^2 norm of u_λ is bounded by a constant independent of λ.

Property 2.4.2 *For every λ, we have $\int_\Omega u_\lambda dxdy = \int_\Omega f dxdy$.*

Proof The Euler-Lagrange equation for the Rudin-Osher-Fatemi method is

$$\begin{cases} u = f + \frac{1}{2\lambda} \text{ div } \left(\frac{\nabla u}{|\nabla u|} \right) \text{ in } \Omega \\ \frac{\partial u}{\partial N} = 0 \text{ on } \partial\Omega. \end{cases}$$

By integrating the first equation over Ω, we obtain

$$\int_\Omega u dxdy = \int_\Omega f dxdy + \frac{1}{2\lambda} \int_\Omega \text{ div } \Big(\frac{\nabla u}{|\nabla u|} \Big) dxdy.$$

Using Green's formula and the boundary condition, we have the following

$$\int_\Omega \text{ div } \Big(\frac{\nabla u}{|\nabla u|} \Big) dxdy = \int_{\partial\Omega} \frac{1}{|\nabla u|} \frac{\partial u}{\partial N} d(\partial\Omega) = 0,$$

which concludes the proof.

Now we discuss some more properties of this model, shown by Meyer in [15]. In this section, we are working on the domain $\Omega = \mathbb{R}^2$. We would like to note that if the image is initially defined only on a rectangle, that it can be extended to all of $\mathbb{R}^2$ by reflection.

Definition 2.4.1 *Let G denote the Banach space consisting of all generalized functions $f(x,y)$ which can be written as*

$$f(x,y) = \frac{\partial}{\partial x} g_1(x,y) + \frac{\partial}{\partial y} g_2(x,y), \quad g_1, g_2 \in L^\infty(\mathbb{R}^2). \tag{2.2}$$

The norm $\|f\|_*$ of f in G is defined as the lower bound of all L^∞ norms of the functions $|\vec{g}|$ where $\vec{g} = (g_1, g_2)$, $|\vec{g}(x,y)| = \sqrt{g_1(x,y)^2 + g_2(x,y)^2}$ and where the infimum is computed over all decompositions (2.2) of f.

Lemma 2.4.1 *If $g \in L^2(\mathbb{R}^2)$, then*

$$|\int f(x,y) g(x,y) dxdy| \leq \|f\|_{BV} \|g\|_*.$$

Proof For proving this observation, we first observe that it is true if $f \in W^{1,1}$. Then we replace f by convolutions with an approximation to the identity and obtain a sequence f_j in $W^{1,1}$ such that $f_j \rightharpoonup f$ as $j \to \infty$. Since $g \in L^2(\mathbb{R}^2)$, we can pass to the weak limit.

Theorem 2.4.1 *If $\|f\|_* > \frac{1}{2\lambda}$, then the Rudin-Osher-Fatemi decomposition $f = u + v$ is characterized by the following two conditions:*

$$\|v\|_* = \frac{1}{2\lambda} \text{ and } \int u(x,y) v(x,y) dxdy = \frac{1}{2\lambda} \|u\|_{BV}. \tag{2.3}$$

If $\|f\|_* \leq \frac{1}{2\lambda}$, then $u = 0$ and $v = f$.

Proof The proof of this theorem uses three steps. First, since u minimizes the functional $F(u)$, we have

$$\|u+\epsilon h\|_{BV}+\lambda\|v-\epsilon h\|_{L^2}^2 \geq \|u\|_{BV}+\lambda\|v\|_{L^2}^2. \tag{2.4}$$

We can see that $\|u+\epsilon h\|_{BV} \leq \|u\|_{BV}+|\epsilon|\|h\|_{BV}$, so we now have

$$|\epsilon|\|h\|_{BV}+\lambda\|v-\epsilon h\|_{L^2}^2 \geq \lambda\|v\|_{L^2}^2.$$

By expanding the squared L^2 norm, we find that

$$\epsilon\int vhdxdy \leq |\epsilon|\frac{1}{2\lambda}\|h\|_{BV}+\frac{\epsilon^2}{2}\|h\|_{L^2}^2.$$

If we let $\epsilon \to 0$, then we obtain

$$|\int vhdxdy| \leq \frac{1}{2\lambda}\|h\|_{BV},$$

which implies that the norm of v in the dual space of $W^{1,1}$ does not exceed $1/2\lambda$ (i.e. $\|v\|_* \leq 1/2\lambda$).

The second step of the proof consists of taking $h=u$ in (2.4). If $\epsilon>0$, then we obtain $\int vudxdy \leq \frac{1}{2\lambda}\|u\|_{BV}$, but if $\epsilon<0$, we get $\int vudxdy \geq \frac{1}{2\lambda}\|u\|_{BV}$. When combined, we conclude that

$$\int vudxdy = \frac{1}{2\lambda}\|u\|_{BV}. \tag{2.5}$$

Now, $\|v\|_* \leq 1/2\lambda$ (from step one) implies that $\|v\|_* = 1/2\lambda$ otherwise equality couldn't be obtained in (2.5).

Our last step is to assume that (2.3) is true and show that $f=u+v$ is the Rudin-Osher-Fatemi decomposition. We write

$$\begin{aligned}\|u+\epsilon h\|_{BV}+\lambda\|v-\epsilon h\|_{L^2}^2 \quad &\geq \quad 2\lambda\int(u+\epsilon h)vdxdy+\lambda\|v\|_{L^2}^2 \\ &\quad -2\lambda\epsilon\int vhdxdy+\lambda\epsilon^2\|h\|_{L^2}^2\end{aligned}$$

$$
\begin{aligned}
&= 2\lambda \int uv dxdy + \lambda \|v\|_{L^2}^2 + \lambda \epsilon^2 \|h\|_{L^2}^2 \\
&= \|u\|_{BV} + \lambda \|v\|_{L^2}^2 + \lambda \epsilon^2 \|h\|_{L^2}^2 \\
&\geq \|u\|_{BV} + \lambda \|v\|_{L^2}^2.
\end{aligned}
$$

The first inequality follows from $\|v\|_* = 1/2\lambda$ and Lemma (2.4.1).

Example 2.4.1 *Let D be a disc centered at the origin and with radius R. We now apply the Rudin-Osher-Fatemi method to $f = \alpha\chi_D$ where α is a positive constant. Let us first assume $\lambda R \geq 1/\alpha$. If we are working on the entire domain $\mathbb{R}^2$, then the Rudin-Osher-Fatemi decomposition of $f = \alpha\chi_D$ is given by*

$$f = u + v, \ u = (\alpha - (\lambda R)^{-1})\chi_D,$$

$$v = (\lambda R)^{-1}\chi_D.$$

If $\lambda R \leq 1/\alpha$, Theorem (2.4.1) implies $u = 0$, $v = f$ which means that f is too small to be treated as an object.

We would also like to note that if we are on a bounded set Ω instead, we have

$$f = u + v, \ u = (\alpha - (\lambda R)^{-1})\chi_D + \frac{\pi R}{\lambda Area(\Omega \backslash D)}\chi_{\Omega \backslash D},$$

$$v = (\lambda R)^{-1}\chi_D - \frac{\pi R}{\lambda Area(\Omega \backslash D)}\chi_{\Omega \backslash D}.$$

We refer the reader to [22], [23] and [24] for more examples of the Rudin-Osher-Fatemi minimization.

Theorem 2.4.2 *A regularized solution $\hat{u}$ to the ill-posed inverse problem in (2.1)*

$$f = Ku + v$$

is characterized by the following two properties

$$\begin{cases} \|K^*f\|_* \leq (2\lambda)^{-1} \Rightarrow \hat{u} = 0,\ v = f \\ \|K^*f\|_* \geq (2\lambda)^{-1} \Rightarrow \|K^*\hat{v}\|_* = (2\lambda)^{-1}\ \textit{and} \\ \langle \hat{u}, K^*\hat{v}\rangle = (2\lambda)^{-1}\|\hat{u}\| \end{cases}$$

Proof If u is the unique minimizer of

$$F(u) = \|u\|_{BV} + \lambda\|f - Ku\|_{L^2}^2,$$

and $v = f - Ku$, then

$$F(u) \leq F(u + \epsilon h),$$

for $\epsilon \in \mathbb{R}$. So, we have

$$\begin{aligned} \|u\|_{BV} + \lambda\|v\|_{L^2}^2 &\leq \|u + \epsilon h\|_{BV} + \lambda\|v - \epsilon Kh\|_{L^2}^2 \\ &\leq \|u + \epsilon h\|_{BV} + \lambda\|v\|_{L^2}^2 + \lambda\epsilon^2\|Kh\|_{L^2}^2 - 2\lambda\epsilon\langle v, Kh\rangle \\ &\leq \|u\|_{BV} + |\epsilon|\|h\|_{BV} + \lambda\|v\|_{L^2}^2 + \lambda\epsilon^2\|Kh\|_{L^2}^2 - 2\lambda\epsilon\langle v, Kh\rangle. \end{aligned}$$

Hence, we have

$$2\lambda\epsilon\langle v, Kh\rangle \leq |\epsilon|\|h\|_{BV} + \lambda\epsilon^2\|Kh\|_{L^2}^2. \tag{2.6}$$

Now, we just follow the proof of Theorem (2.4.1) to obtain

$$2\lambda\langle v, Kh\rangle \leq \|h\|_{BV}$$

. This says that $2\lambda\|K^*\hat{v}\|_* \leq 1$. Now, we go back to (2.6) and let $h = u$ and again follow the proof of Theorem (2.4.1) and obtain $2\epsilon\lambda\langle v, Ku\rangle \leq \epsilon\|u\|$ and then our desired result, $(2\lambda)\|K^*\hat{v}\|_* = 1$.

CHAPTER 3

Multiscale Image Representation Using Hierarchical (BV, L^2) Decompositions

3.1 Introduction

The total variation minimization model of Rudin-Osher-Fatemi is very good at denoising images, while at the same time preserving edges. However, if f contains small textured patterns, they end up in the residual, v, along with the noise. This means that u will be a 'cartoon' representation of f. If λ is a large parameter, then u is close to f and not much change has been applied to f. However, if λ is a small parameter, then the regularizing term is stronger and the image u will contain only the large details of f. Sometimes the parameter λ can be estimated if some statistical information about the noise is known as in [21] or [10].
We propose a hierarchical decomposition of f. By extracting more than one u component from f, we obtain improved results and useful decompositions.

3.2 Description of the Hierarchical (BV, L^2) Decomposition

Given $f \in L^2(\Omega)$ and a small starting parameter $\lambda > 0$, consider the decomposition of f provided by the total variation minimization (Rudin-Osher-Fatemi

model) in the following sense. Let u_λ, be the unique minimizer of

$$\inf_{u \in BV(\Omega)} F_\lambda(u) = \lambda \int_\Omega (f-u)^2 dxdy + \int_\Omega |\nabla u| dxdy, \tag{3.1}$$

or in other words, the unique minimizer of

$$\inf_{u \in BV(\Omega)} F_\lambda(u) = \lambda \|v\|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)},$$

and denote $v_\lambda := f - u_\lambda$. Then we have $f = u_\lambda + v_\lambda$, with $u_\lambda \in BV(\Omega)$ and $v_\lambda \in L^2(\Omega)$.

Formally minimizing F_λ with respect to u yields the following associated Euler-Lagrange equation:

$$\begin{cases} u_\lambda = f + \frac{1}{2\lambda} \text{ div } \left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|} \right) \text{ in } \Omega \\ \frac{\partial u_\lambda}{\partial N} = 0 \text{ on } \partial\Omega. \end{cases}$$

The main idea is as follows: substitute f by v_λ and λ by 2λ in (3.1) and repeat the process.

This means that in the second iteration, we want to solve

$$\inf_{u \in BV(\Omega)} F_{2\lambda}(u) = 2\lambda \int_\Omega (v_\lambda - u)^2 dxdy + \int_\Omega |\nabla u| dxdy.$$

This will, in turn, give us a new unique minimizer, $u_{2\lambda}$, and hence a new residual, $v_{2\lambda} := v_\lambda - u_{2\lambda}$. Since $v_\lambda = u_{2\lambda} + v_{2\lambda}$, we have

$$f = u_\lambda + v_\lambda = u_\lambda + u_{2\lambda} + v_{2\lambda}.$$

Now let $\lambda_k = 2^k \lambda$ (so $\lambda_0 = \lambda$). After k steps, we have $v_{\lambda_{k-1}}$ and we want to minimize

$$\inf_{u \in BV(\Omega)} F_{\lambda_k}(u) = \lambda_k \int_\Omega (v_{\lambda_{k-1}} - u)^2 dxdy + \int_\Omega |\nabla u| dxdy. \tag{3.2}$$

Let u_{λ_k} be the unique minimizer of (3.2), and $v_{\lambda_k} := v_{\lambda_{k-1}} - u_{\lambda_k}$, the residual.

By putting all the pieces together, we obtain the following hierarchical decomposition of f:

$$
\begin{aligned}
f &= u_{\lambda_0} + v_{\lambda_0} \\
&= u_{\lambda_0} + u_{\lambda_1} + v_{\lambda_1} \\
&= u_{\lambda_0} + u_{\lambda_1} + u_{\lambda_2} + v_{\lambda_2} \\
&\vdots \\
&= u_{\lambda_0} + u_{\lambda_1} + u_{\lambda_2} + \cdots + u_{\lambda_{k-1}} + u_{\lambda_k} + v_{\lambda_k},
\end{aligned}
$$

or $v_{\lambda_k} = f - (u_{\lambda_0} + u_{\lambda_1} + u_{\lambda_2} + \cdots + u_{\lambda_{k-1}} + u_{\lambda_k})$ (the residual at the $(k+1)$-st step).

This procedure defines a multiscale image decomposition of f.

Images could be realized as general L^2-objects, $f \in L^2$, representing the grayscale of the observed image. Likewise, color images are typically realized in terms of vector-valued L^2-RGB scale functions, $f = (f_1, f_2, f_3) \in L^2$. In practice, the more notable features of images could be identified within a proper subclass of all L^2 objects. Most notable are the edges of an image, which are known to be well quantified within the smaller subclass of functions of Bounded Variation (BV). Quantifying the precise L^2 subclasses for these different features is still the subject of current research. We argue that a large class of notable images form an 'intermediate' space between the larger $L^2(\mathbb{R}^2)$ and the smaller $BV(\mathbb{R}^2)$. The standard tool for studying such 'intermediate' spaces is interpolation. To this end, one starts with a pair of given spaces, $Y \subset X$, and forms a scale of intermediate spaces, $(X, Y)_\theta, \theta \in [0, 1]$, ranging from $(X, Y)_{\theta=0} = X$ to $(X, Y)_{\theta=1} = Y$, expressed in terms of the appropriate K-functional

$$
K(f, \lambda) \equiv K(f, \lambda; X, Y) := \inf_{u+v=f} \Big\{ \|v\|_X + \lambda \|u\|_Y \Big\}.
$$

In this dissertation, the K-functional is replaced by the closely related J-functional of the form

$$J_p(f,\lambda;X,Y) := \inf_{u+v=f} \Big\{ \lambda \|v\|_X^p + \|u\|_Y \Big\}.$$

The functional $J_2(f,\lambda)$ measures how well an L^2 object can be approximated by its BV features, $J_2(f,\lambda) \sim \lambda^\theta$ as $\lambda \uparrow \infty$. The classical argument addresses this question of convergence rate in terms of the smoothness properties of f. In modern theory the roles are reversed: one defines the scale of smoothness spaces, $(L^2, BV)_\theta$ in terms of the behavior of $J(f,\lambda)$. Here we will introduce a new multiscale decomposition whose properties quantify images in intermediate space which we denote as 'multiscale (BV, L^2) decompositions'.

Remark 3.2.1 *By setting $\lambda_k = 2^k\lambda_0$, we obtain a decomposition using dyadic scales.*

Remark 3.2.2 *We can also obtain a finer or coarser decomposition by changing the factor of 2 in the expression of the scaling parameter $\lambda_k = 2^k\lambda_0$. Let $\lambda_k = \alpha^k\lambda_0$, for $\alpha > 1$ instead. For a finer decomposition, let $1 < \alpha < 2$ and for a coarser decomposition, let $\alpha > 2$. This will also hold for all of the multiscale decomposition methods to follow.*

3.2.1 Some Mathematical Properties of Our Scheme

Theorem 3.2.1 *If $\|f\|_* > \frac{1}{2\lambda_0}$, then $\forall k \geq 0$, we have $\|v_{\lambda_k}\|_* > \frac{1}{2\lambda_{k+1}}$.*

Proof From Theorem (2.4.1), we have

$$\|v_{\lambda_k}\|_* = \frac{1}{2\lambda_k} > \frac{1}{2\lambda_{k+1}},$$

since $\lambda_{k+1} = 2\lambda_k$.

This means that if we can make the first decomposition, we can make all the proceeding ones.

Theorem 3.2.2 *Let* $\|f\|_* > \frac{1}{2\lambda_0}$. *If we obtain a multiscale decomposition of* $f = u_{\lambda_0} + u_{\lambda_1} + \cdots + u_{\lambda_{k-1}} + u_{\lambda_k} + v_{\lambda_k}$, *then*

$$\|v_{\lambda_k}\|_* \to 0 \text{ as } k \to \infty.$$

Proof This follows directly from Theorem (2.4.1), since

$$\|v_{\lambda_k}\|_* = \|v_{2^k\lambda_0}\|_* = \frac{1}{2^{k+1}\lambda_0}.$$

Now by letting $k \to \infty$, we obtain the desired result.

Corollary 3.2.1 *Under the same conditions as Theorem (3.2.2), we have*

$$\|f - \sum_{i=0}^{k} u_{\lambda_i}\|_* \to 0 \text{ as } k \to \infty.$$

Proof Since $v_{\lambda_k} = f - \sum_{i=0}^{k} u_{\lambda_i}$, it follows from Theorem (3.2.2).

Theorem 3.2.3 *Let* $\langle h, g\rangle = \int hg$ *denote the inner product of* h *and* g. *For any* $h \in BV$,

$$|\langle h, f - \sum_{i=0}^{k} u_{\lambda_i}\rangle| \to 0, \text{ as } k \to \infty,$$

and hence, $\lim_{k\to\infty} v_{\lambda_k} \perp BV$.

Proof From Lemma (2.4.1),

$$|\langle h, f - \sum_{i=0}^{k} u_{\lambda_i}\rangle| \le \|h\|_{BV}\|f - \sum_{i=0}^{k} u_{\lambda_i}\|_*,$$

but $\|f - \sum_{i=0}^{k} u_{\lambda_i}\|_* \to 0$ as $k \to \infty$ from Corollary (3.2.1) and $h \in BV$, so the results follows.

Corollary 3.2.2 *For any $h \in BV$,*

$$|\langle h, f\rangle - \sum_{i=0}^{k} \langle h, u_{\lambda_i}\rangle| \to 0, \text{ as } k \to \infty,$$

or in other words

$$\sum_{i=0}^{k} \langle h, u_{\lambda_i}\rangle \to \langle h, f\rangle \text{ as } k \to \infty.$$

Proof Just substitute

$$\langle h, f - \sum_{i=0}^{k} u_{\lambda_i}\rangle = \langle h, f\rangle - \sum_{i=0}^{k} \langle h, u_{\lambda_i}\rangle$$

in Theorem (3.2.3) and the proof is done.

Theorem 3.2.4 *The L^2-norm of both sequences, $\{u_{\lambda_k}\}$ and $\{v_{\lambda_k}\}$, is uniformly bounded.*

Proof First, we will look at the sequence $\{v_{\lambda_k}\}$. Let $u_{\lambda_{k+1}}$ be the unique minimizer of $\int |\nabla u| + \lambda_{k+1} \int (v_{\lambda_k} - u)^2$. Then if $u = 0$,

$$\begin{aligned} \|u_{\lambda_{k+1}}\|_{BV} + \lambda_{k+1}\|v_{\lambda_k} - u_{\lambda_{k+1}}\|_{L^2}^2 &= \|u_{\lambda_{k+1}}\|_{BV} + \lambda_{k+1}\|v_{\lambda_{k+1}}\|_{L^2}^2 \\ &\leq \lambda_{k+1}\|v_{\lambda_k}\|_{L^2}^2. \end{aligned}$$

From this we obtain

$$\|v_{\lambda_{k+1}}\|_{L^2}^2 \leq \|v_{\lambda_k}\|_{L^2}^2.$$

Thus $\{\|v_{\lambda_k}\|_{L^2}^2\}$ is a decreasing sequence which is bounded below by 0, and hence it is uniformly bounded.

As for the sequence $\{u_{\lambda_k}\}$, using Property (2.4.1), we have

$$\|u_{\lambda_{k+1}}\|_{L^2}^2 \leq 4\|v_{\lambda_k}\|_{L^2}^2,$$

and so it is also uniformly bounded.

Example 3.2.1 *Let D be a disc centered at the origin and with radius R. We now apply the multiscale decomposition algorithm to $f = \alpha\chi_D$ where α is a positive constant. Let us assume $\lambda_0 R \geq 1/\alpha$. If we are working on the entire domain $\mathbb{R}^2$, then the multiscale decomposition of $f = \alpha\chi_D$ is given by*

$$f = u_{\lambda_0} + u_{\lambda_1} + \cdots + u_{\lambda_{k-1}} + u_{\lambda_k} + v_{\lambda_k},$$

$$\sum_{i=0}^{k} u_{\lambda_i} = (\alpha - (\lambda_k R)^{-1})\chi_D,$$

$$v_{\lambda_k} = (\lambda_k R)^{-1}\chi_D.$$

Proof From Property (2.4.1), we have

$$\begin{aligned} u_{\lambda_0} &= (\alpha - (\lambda_0 R)^{-1})\chi_D \text{ and} \\ v_{\lambda_0} &= (\lambda_0 R)^{-1}\chi_D. \end{aligned}$$

In the next step, we decompose v_{λ_0} instead of f, so by Property (2.4.1) again,

$$\begin{aligned} u_{\lambda_1} &= ((\lambda_0 R)^{-1} - (\lambda_1 R)^{-1})\chi_D \text{ and} \\ v_{\lambda_1} &= (\lambda_1 R)^{-1}\chi_D. \end{aligned}$$

If we continue this process for $(k-1)$ more steps, we obtain

$$\begin{aligned} u_{\lambda_k} &= ((\lambda_{k-1} R)^{-1} - (\lambda_k R)^{-1})\chi_D \text{ and} \\ v_{\lambda_k} &= (\lambda_k R)^{-1}\chi_D. \end{aligned}$$

Now we have $f = \sum_{i=0}^{k} u_{\lambda_i} + v_{\lambda_k}$, where

$$\begin{aligned} \sum_{i=0}^{k} u_{\lambda_i} &= (\alpha - (\lambda_k R)^{-1})\chi_D, \text{ and} \\ v_{\lambda_k} &= (\lambda_k R)^{-1}\chi_D. \end{aligned}$$

We can clearly see that the multiscale decomposition obtains a better approximation of f than the Rudin-Osher-Fatemi method in this example. If we look

ahead a little bit to the next chapter, we can see an example of this property in figure (4.11), in which we plot the images of $\sum_{j=0}^{i} u_{\lambda_j}$ and v_{λ_i} for $i = 0, \ldots, 4$.

3.3 Description of the Hierarchical (BV, L^2) Decomposition for Vector-Valued Images

We would now like to extend this technique from scalar-valued, gray-scale images, to vector-valued, color images. In particular, we would like to use the RGB representation for color images, so $\vec{f} = (f_1, f_2, f_3) = (f_R, f_G, f_B) \in L^2(\Omega)^3$. Note that even though we will only discuss the color image case, this discussion could be made in any dimension.

We will now extend the total variation minimization to vector-valued functions. First, we say that a vector-valued function $\vec{u} = (u_1, u_2, u_3) \in L^1(\Omega)^3$ is of bounded variation if the quantity $\int_\Omega |\nabla \vec{u}| dxdy$ defined by

$$\int_\Omega |\nabla \vec{u}| dxdy = \sup_{\vec{\varphi}} \Big\{ \sum_{i=1}^{3} \int_\Omega u_i \text{ div } \vec{\varphi}_i dxdy; \vec{\varphi}_i \in C_0^\infty(\Omega; \mathbb{R}^2), |(\vec{\varphi}_1, \vec{\varphi}_2, \vec{\varphi}_3)| \leq 1 \Big\}.$$

is finite, and we will again denote it by $\|\vec{u}\|_{BV(\Omega)}$. The new minimization model can be written as

$$\begin{aligned} \inf_{\vec{u}=(u_1,u_2,u_3)\in BV(\Omega)} F_\lambda(\vec{u}) &= \int_\Omega \sqrt{|\nabla u_1|^2 + |\nabla u_2|^2 + |\nabla u_3|^2} dxdy \\ &+ \lambda \sum_{i=1}^{3} \int_\Omega |u_i - f_i|^2 dxdy. \end{aligned}$$

This can also be expressed as

$$\inf_{\vec{u}=(u_1,u_2,u_3)\in BV(\Omega)} F_\lambda(\vec{u}) = \|\vec{u}\|_{BV(\Omega)} + \lambda \sum_{i=1}^{3} \|u_i - f_i\|^2_{L^2(\Omega)}.$$

Formally minimizing the energy above with respect to u_1, u_2 and u_3 gives the following system of coupled Euler-Lagrange equations

$$\begin{aligned} u_1 &= f_1 + \frac{1}{2\lambda}\ \text{div}\ \Big(\frac{\nabla u_1}{|\nabla \vec{u}|}\Big), \\ u_2 &= f_2 + \frac{1}{2\lambda}\ \text{div}\ \Big(\frac{\nabla u_2}{|\nabla \vec{u}|}\Big), \\ u_3 &= f_3 + \frac{1}{2\lambda}\ \text{div}\ \Big(\frac{\nabla u_3}{|\nabla \vec{u}|}\Big). \end{aligned}$$

The hierarchical decomposition is constructed as in the previous case, so again $\vec{f} = \vec{u}_{\lambda_0} + \vec{u}_{\lambda_1} + \vec{u}_{\lambda_2} + \cdots + \vec{u}_{\lambda_{k-1}} + \vec{u}_{\lambda_k} + \vec{v}_{\lambda_k}$.

3.4 Description of the Hierarchical (BV, L^2) Decomposition for Blurred Scalar-Valued Images

Given $f \in L^2(\Omega)$, a small starting parameter $\lambda > 0$, and the blurring kernel K, consider the decomposition of f provided by the total variation minimization model. Let u_λ, be the unique minimizer of

$$\inf_{u \in BV(\Omega)} F_\lambda(u) = \lambda \int_\Omega (f - Ku)^2 dxdy + \int_\Omega |\nabla u| dxdy, \tag{3.3}$$

or equivalently, the unique minimizer of

$$\inf_{u \in BV(\Omega)} F_\lambda(u) = \lambda \|f - Ku\|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)},$$

Let $v_\lambda := f - Ku_\lambda$, so $f = Ku_\lambda + v_\lambda$, with $u_\lambda \in BV(\Omega)$ and $Ku_\lambda, v_\lambda \in L^2(\Omega)$. Formally minimizing F_λ with respect to u yields the following associated Euler-Lagrange equation:

$$\begin{cases} K^*Ku_\lambda = K^*f + \frac{1}{2\lambda}\ \text{div}\ \Big(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\Big)\ \text{in}\ \Omega \\ \frac{\partial u_\lambda}{\partial N} = 0\ \text{on}\ \partial\Omega. \end{cases}$$

Again, we construct the hierarchical decomposition as before, and so

$$\begin{aligned} f &= Ku_{\lambda_0} + Ku_{\lambda_1} + Ku_{\lambda_2} + \cdots + Ku_{\lambda_{k-1}} + Ku_{\lambda_k} + v_{\lambda_k} \\ &= K(u_{\lambda_0} + u_{\lambda_1} + u_{\lambda_2} + \cdots + u_{\lambda_{k-1}} + u_{\lambda_k}) + v_{\lambda_k}. \end{aligned}$$

3.5 Description of the Hierarchical (BV, L^2) Decomposition for Images with Multiplicative Noise

Let $f \in L^2(\Omega)$, where $f = u \cdot v$, u is the true image and v is multiplicative noise with mean 1.

Consider the decomposition of f provided by the Rudin-Osher model in [20]. Let u_λ, be the unique minimizer of

$$\inf_{u \in BV(\Omega)} F_\lambda(u) = \lambda \int_\Omega (\frac{f}{u} - 1)^2 dxdy + \int_\Omega |\nabla u| dxdy, \tag{3.4}$$

or equivalently, the unique minimizer of

$$\inf_{u \in BV(\Omega)} F_\lambda(u) = \lambda \| \frac{f}{u} - 1 \|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)}.$$

Set $v_\lambda := \frac{f}{u_\lambda}$. Now, $f = u_\lambda \cdot v_\lambda$, with $u_\lambda \in BV(\Omega)$ and $v_\lambda \in L^2(\Omega)$.

Again, here $\lambda > 0$ is a scaling parameter. Formally minimizing F_λ with respect to u yields the following associated Euler-Lagrange equation:

$$\begin{cases} \left(\frac{f}{u_\lambda} - 1\right) \cdot \left(-\frac{f}{u_\lambda^2}\right) = \frac{1}{2\lambda} \text{ div } \left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\right) \text{ in } \Omega \\ \frac{\partial u_\lambda}{\partial N} = 0 \text{ on } \partial\Omega. \end{cases}$$

Again, we construct the hierarchical decomposition as before, except we replace sums with products and subtraction with division.

Thus, $f = u_{\lambda_0} \cdot u_{\lambda_1} \cdot u_{\lambda_2} \cdot \ldots \cdot u_{\lambda_{k-1}} \cdot u_{\lambda_k} \cdot v_{\lambda_k}$.

3.6 Multiscale Ambrosio-Tortorelli Approximation of the Mumford-Shah Functional

In this section, we would like to show that we are able to create a multiscale version of other existing variational PDE models. We have applied the same technique to the Ambrosio-Tortorelli approximation to the Mumford-Shah model as the Rudin-Osher-Fatemi model. Now recall that the Mumford-Shah functional is defined by

$$F^{MS}(u,S)=\int_{\Omega\setminus S}(\alpha|\nabla u|^2+\beta(u-f)^2)dxdy+\mathcal{H}^1(S),$$

where f is the given image, u is a piecewise smooth optimal approximation of f, S is the set of contours of u, $\mathcal{H}^1$ is the Hausdorff 1-dimensional measure in $\mathbb{R}^2$, and $\alpha,\beta>0$ are fixed positive parameters. Also, remember the Ambrosio-Tortorelli functional

$$G_\rho^{AT}(u,w)=\int_\Omega[\rho|\nabla w|^2+\alpha(w^2|\nabla u|^2+\frac{(w-1)^2}{4\alpha\rho})+\beta|u-f|^2]dxdy, \qquad (3.5)$$

where w represents the contours of u. The Euler-Lagrange equations associated with (3.5) are

$$\begin{cases}\beta u-\alpha\nabla(w^2\nabla u)=\beta f\\ -\triangle w+\frac{1+4\alpha\rho|\nabla u|^2}{4\rho^2}(w-\frac{1}{1+4\alpha\rho|\nabla u|^2})=0.\end{cases}$$

Definition 3.6.1 *The space $SBV(\Omega)$ is defined as follows*

$$SBV(\Omega)=\Big\{u\in BV(\Omega):Du=\nabla u dxdy+(u^+-u^-)\mathcal{H}^1(J_u)\Big\},$$

where $\mathcal{H}^1$ is the Hausdorff 1-dimensional measure in $\mathbb{R}^2$, J_u is the set of jumps of u and (u^+-u^-) is the magnitude of the jump.

3.6.1 Description of the Hierarchical (SBV, L^2) Decomposition

Given $f \in L^2(\Omega)$, a small starting parameter $\beta > 0$, a fixed α and small fixed ρ, consider the decomposition of f provided by the Ambrosio-Tortorelli minimization in the following sense. Let pair, (u_β, w_β), be a minimizer of

$$\inf_{u \in H^1} G_\beta(u, w) = \int_\Omega [\rho |\nabla w|^2 + \alpha(w^2 |\nabla u|^2 + \frac{(w-1)^2}{4\alpha\rho}) + \beta |u - f|^2] dxdy, \quad (3.6)$$

and denote $v_\beta := f - u_\beta$. Then we have $f = u_\beta + v_\beta$, with $u_\beta, w_\beta \in H^1(\Omega)$ and $v_\beta \in L^2(\Omega)$.

Formally minimizing G_β with respect to u and w yields the following associated Euler-Lagrange equations:

$$\begin{cases} \beta u_\beta - \alpha \nabla(w_\beta^2 \nabla u_\beta) = \beta f \\ -\triangle w_\beta + \frac{1+4\alpha\rho|\nabla u_\beta|^2}{4\rho^2}(w_\beta - \frac{1}{1+4\alpha\rho|\nabla u_\beta|^2}) = 0. \end{cases}$$

We construct the hierarchical decomposition in the same manner as before, so $f = u_{\beta_0} + u_{\beta_1} + u_{\beta_2} + \cdots + u_{\beta_{k-1}} + u_{\beta_k} + v_{\beta_k}$. Now, what about the w_{β_i}'s? The combination of all of the w_{β_i}'s gives the set of edges of the image u.

Remark 3.6.1 *For this multiscale decompostion we can also obtain a finer or coarser decomposition by changing the scaling parameter β by something smaller than 2 for a finer decomposition and something greater than 2 for a coarser decomposition.*

CHAPTER 4

Numerical Experiments

In this chapter, we give the details of the numerical scheme and algorithm used for the decompositions. We also present some results of the decompositions applied to both synthetic and real images. In each case, we discretize (using finite differences) the Euler-Lagrange equation associated with that minimization to solve every step of the hierarchical decomposition.

4.1 Numerical Discretization of the Hierarchical Decomposition of Gray-Scale Images Without Blurring

From [21], at each step of our scheme, we would like to find the unique minimizer, u_λ, of

$$\inf_{u \in BV(\Omega)} F_\lambda(u) = \lambda \|f - u\|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)},$$

where f is the original image and $\lambda > 0$ is a scaling parameter. The associated Euler-Lagrange equation of the Rudin-Osher-Fatemi model is

$$\begin{cases} u_\lambda = f + \frac{1}{2\lambda} \text{ div } \left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|} \right) \text{ in } \Omega \\ \frac{\partial u_\lambda}{\partial N} = 0 \text{ on } \partial\Omega. \end{cases}$$

For the discretization of the Rudin-Osher-Fatemi model, we follow [26] and [7].

First, we remove the singularity when $|\nabla u_\lambda| = 0$, by approximating $F(u)$ by $F_\epsilon(u)$, where

$$F_\epsilon(u) = \lambda \int_\Omega (f-u)^2 dxdy + \int_\Omega \sqrt{\epsilon^2 + |\nabla u|^2} dxdy.$$

Then, the Euler-Lagrange equation minimizing $F_\epsilon(u)$ is:

$$\begin{aligned} u_\lambda &= f + \frac{1}{2\lambda} \mathrm{div}\Big(\frac{\nabla u_\lambda}{\sqrt{\epsilon^2 + |\nabla u_\lambda|^2}} \Big) \text{ in } \Omega, & (4.1) \\ \frac{\partial u_\lambda}{\partial n} &= 0 \text{ on } \partial\Omega. & (4.2) \end{aligned}$$

Assume for simplicity $\Omega = (0,1)^2$, $h > 0$ and let $x_i = ih$, $y_j = jh$, $h = 1/M$, for $0 \le i, j \le M$, be the discrete points. We recall the following notations:

$$\begin{aligned} u_{i,j} &\approx u(x_i, y_j), \\ f_{i,j} &\approx f(x_i, y_j), \\ \triangle^x_\pm f_{i,j} &= \pm(f_{i\pm1,j} - f_{i,j}), \\ \triangle^y_\pm f_{i,j} &= \pm(f_{i,j\pm1} - f_{i,j}), \\ \triangle^x_0 f_{i,j} &= (f_{i+1,j} - f_{i-1,j})/2, \text{ and} \\ \triangle^y_0 f_{i,j} &= (f_{i,j+1} - f_{i,j-1})/2. \end{aligned}$$

The discrete form of the Euler-Lagrange equation is:

$$\begin{aligned} u_{i,j} &= f_{i,j} + \frac{1}{2\lambda h} \triangle^x_- \Big[\frac{1}{\sqrt{\epsilon^2 + (\frac{\triangle^x_+ u_{i,j}}{h})^2 + (\frac{\triangle^y_0 u_{i,j}}{h})^2}} \frac{\triangle^x_+ u_{i,j}}{h} \Big] \\ &\quad + \frac{1}{2\lambda h} \triangle^y_- \Big[\frac{1}{\sqrt{\epsilon^2 + (\frac{\triangle^x_0 u_{i,j}}{h})^2 + (\frac{\triangle^y_+ u_{i,j}}{h})^2}} \frac{\triangle^y_+ u_{i,j}}{h} \Big] \\ &= f_{i,j} + \frac{1}{2\lambda h^2} \frac{u_{i+1,j} - u_{i,j}}{\sqrt{\epsilon^2 + (\frac{u_{i+1,j} - u_{i,j}}{h})^2 + (\frac{u_{i,j+1} - u_{i,j-1}}{2h})^2}} \\ &\quad - \frac{1}{2\lambda h^2} \frac{u_{i,j} - u_{i-1,j}}{\sqrt{\epsilon^2 + (\frac{u_{i,j} - u_{i-1,j}}{h})^2 + (\frac{u_{i-1,j+1} - u_{i-1,j-1}}{2h})^2}} \end{aligned}$$

$$+\frac{1}{2\lambda h^2}\frac{u_{i,j+1}-u_{i,j}}{\sqrt{\epsilon^2+(\frac{u_{i+1,j}-u_{i-1,j}}{2h})^2+(\frac{u_{i,j+1}-u_{i,j}}{h})^2}}$$
$$-\frac{1}{2\lambda h^2}\frac{u_{i,j}-u_{i,j-1}}{\sqrt{\epsilon^2+(\frac{u_{i+1,j-1}-u_{i-1,j-1}}{2h})^2+(\frac{u_{i,j}-u_{i,j-1}}{h})^2}},$$

We use a fixed point Gauss-Seidel iteration method for the above equation and so we now introduce the following linearized equation:

$$\begin{aligned}
u_{i,j}^{n+1} &= f_{i,j}+\frac{1}{2\lambda h^2}\frac{u_{i+1,j}^{n}-u_{i,j}^{n+1}}{\sqrt{\epsilon^2+(\frac{u_{i+1,j}^{n}-u_{i,j}^{n}}{h})^2+(\frac{u_{i,j+1}^{n}-u_{i,j-1}^{n}}{2h})^2}} \\
&\quad -\frac{1}{2\lambda h^2}\frac{u_{i,j}^{n+1}-u_{i-1,j}^{n}}{\sqrt{\epsilon^2+(\frac{u_{i,j}^{n}-u_{i-1,j}^{n}}{h})^2+(\frac{u_{i-1,j+1}^{n}-u_{i-1,j-1}^{n}}{2h})^2}} \\
&\quad +\frac{1}{2\lambda h^2}\frac{u_{i,j+1}^{n}-u_{i,j}^{n+1}}{\sqrt{\epsilon^2+(\frac{u_{i+1,j}^{n}-u_{i-1,j}^{n}}{2h})^2+(\frac{u_{i,j+1}^{n}-u_{i,j}^{n}}{h})^2}} \\
&\quad -\frac{1}{2\lambda h^2}\frac{u_{i,j}^{n+1}-u_{i,j-1}^{n}}{\sqrt{\epsilon^2+(\frac{u_{i+1,j-1}^{n}-u_{i-1,j-1}^{n}}{2h})^2+(\frac{u_{i,j}^{n}-u_{i,j-1}^{n}}{h})^2}},
\end{aligned}$$

Introducing the notations:

$$\begin{aligned}
c_1 &= \frac{1}{\sqrt{\epsilon^2+(\frac{u_{i+1,j}^{n}-u_{i,j}^{n}}{h})^2+(\frac{u_{i,j+1}^{n}-u_{i,j-1}^{n}}{2h})^2}}, \\
c_2 &= \frac{1}{\sqrt{\epsilon^2+(\frac{u_{i,j}^{n}-u_{i-1,j}^{n}}{h})^2+(\frac{u_{i-1,j+1}^{n}-u_{i-1,j-1}^{n}}{2h})^2}}, \\
c_3 &= \frac{1}{\sqrt{\epsilon^2+(\frac{u_{i+1,j}^{n}-u_{i-1,j}^{n}}{2h})^2+(\frac{u_{i,j+1}^{n}-u_{i,j}^{n}}{h})^2}}, \\
c_4 &= \frac{1}{\sqrt{\epsilon^2+(\frac{u_{i+1,j-1}^{n}-u_{i-1,j-1}^{n}}{2h})^2+(\frac{u_{i,j}^{n}-u_{i,j-1}^{n}}{h})^2}},
\end{aligned}$$

and solving for $u_{i,j}^{n+1}$, we obtain:

$$\begin{aligned}
u_{i,j}^{n+1} &= \Big(\frac{1}{1+\frac{1}{2\lambda h^2}(c_1+c_2+c_3+c_4)}\Big) \\
&\quad \cdot \quad \Big[f_{i,j}+\frac{1}{2\lambda h^2}(c_1u_{i+1,j}^{n}+c_2u_{i-1,j}^{n}+c_3u_{i,j+1}^{n}+c_4u_{i,j-1}^{n})\Big].
\end{aligned}$$

We let $u^0_{i,j} = f_{i,j}$. Then, we note that if $m_1 \leq f_{i,j} \leq m_2$, for any $0 \leq i, j \leq M$, we have $m_1 \leq u^n_{i,j} \leq m_2$, for any $n \geq 0$. We use the above equation for $u^{n+1}_{i,j}$ for all points (x_i, y_j) such that $1 \leq i, j \leq M-1$ (i.e. inside the boundary). The question is how do we deal with the boundary, $\partial\Omega$? Since we have Neumann boundary conditions, our first approach was to set

$$\begin{cases} u^{n+1}_{i,0} = u^{n+1}_{i,1} \text{ for } i = 0, \ldots, M \\ u^{n+1}_{0,j} = u^{n+1}_{1,j} \text{ for } j = 0, \ldots, M \\ u^{n+1}_{i,M} = u^{n+1}_{i,M-1} \text{ for } i = 0, \ldots, M \\ u^{n+1}_{M,j} = u^{n+1}_{M-1,j} \text{ for } j = 0, \ldots, M \end{cases}$$

If we look at the residual in figure (4.4), we see that this approach resulted in good results everywhere except the boundary. In order to fix this problem we decided to extend the boundaries of f by adding rows of points to all sides of Ω (we chose to add 10 rows to each side in our experiments, however it is possible to use less). Extension of Ω was achieved using reflection along the boundary as shown in figure (4.1). This new approach solved our boundary problem as we can see from the residual in figure (4.5).

Another issue that comes with numerical discretization is how can we implement the algorithm such that we see the least amount of grid effect? We attempted to address this issue by rotating the starting point of the algorithm between the four corners of the grid and by alternating whether we run the algorithm row by row or column by column. We illustrate this in figure (4.2).

Since this is a fixed point iteration, it is important to decide how many iterations we need to make until we converge to u_λ. One way to do this is to iterate until

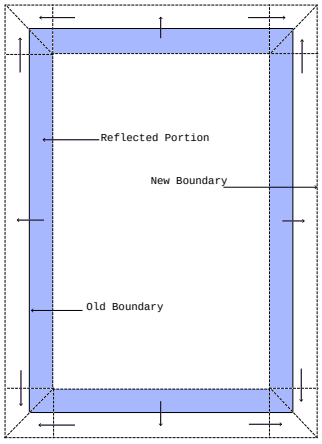

Figure 4.1: We use reflection to extend the boundary of Ω.

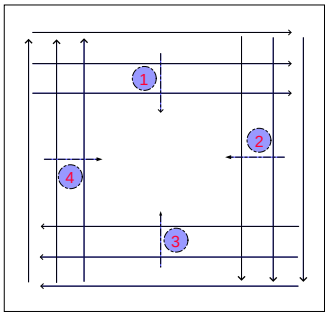

Figure 4.2: This shows the rotation of the starting point of the algorithm in order to minimize grid effect.

$\|u^{n+1} - u^n\| \leq TOL$, where TOL is some given tolerance. However, in practice, we usually use a large fixed number of iterations.

So far, we have specified how we implement the algorithm for a fixed λ. In order to make this a multiscale decomposition, we re-iterate this process, each time updating the value of f and λ in the following way:

$$\begin{cases} f_{new} = v_\lambda = f_{old} - u_\lambda \\ \lambda_{new} = 2\lambda_{old} \end{cases}$$

In other words, we take the residual of the previous step and apply the Rudin-Osher-Fatemi minimization using a larger (here doubled) scaling parameter, λ. Let $\lambda_i = 2^i\lambda_0$, for all i. The ending result after k steps is a multiscale representation of f, defined by

$$f = u_{\lambda_0} + u_{\lambda_1} + u_{\lambda_2} + \cdots + u_{\lambda_{k-1}} + v_{\lambda_{k-1}}.$$

Now the question arises: how many steps, k, should we do? We have considered a couple of different stopping criteria. The first one is to stop when $\|v_{\lambda_{k-1}}\|_* \leq TOL$. We know that this will eventually be achieved because $\|v_{\lambda_{k-1}}\|_* = \frac{1}{2^k\lambda}$. Another is $\|u_{\lambda_k} - u_{\lambda_{k-1}}\| \leq TOL$. However, it is not clear that this will ever be realized for a general image and any given tolerance. It is also possible to look at $\|v_{\lambda_k}\|_{L^2} \leq TOL$. But, again we don't have a guarantee this will ever occur.

We would now like to look further at some of our results. If we look at figure (4.6), it is clear that as we recover more texture, we also start to recover the noise which is of the same scale as that texture. This leads to a balancing act between these two elements: how much texture are you willing to give up to eliminate more noise. Also, in order to obtain better results in the case of noisy initial data, it might be better to use a finer decomposition of f (see Remark (3.2.2)).

In figure (4.7), we see how our multiscale decomposition adds more and more detail at each stage of the algorithm. Now, if we look at figures (4.8) and (4.9),

we are able to see the different pieces originally discarded with the v term that come back into our final result.

One last thing we would like to point out here is how the smaller values of the scaling factor, λ, correspond to the larger objects in the image. In the last row of figure (4.10), we see how all of the objects of smaller scales are only brought into u when we consider larger values for λ, whereas, the larger objects appear in the first few terms, or when λ is smaller, so in fact we are creating a separation of scales.

4.1.1 Localization of the Algorithm

We address one last question: can we localize this algorithm such that we do more work in a region with more texture and less in a homogeneous region? We looked to the ideas in [11]. Let f be a function defined on Ω, a rectangle in $\mathbb{R}^2$. They propose a way to split Ω into regions depending on how much texture they have. They calculate either $\|f\|_{BV}$ or $\|f - ave(f)\|_{L^2}$ (where $ave(f)$ is the average of f on the region) on the region they are considering. Let $\Omega_0 = \Omega$. If $\|f - ave(f)\|_{L^2(\Omega_0)} < TOL$ (or $\|f\|_{BV(\Omega_0)} < TOL$), they stop since the region is homogeneous enough. Otherwise, they split the current region, Ω_0 into four equal rectangles, Ω_{0i}, for $i = 1, 2, 3, 4$, and the process continues on each of these smaller regions. In figures (4.12) and (4.13), we can see both of their criteria in action. The first image is our starting image, f. The second and fourth rows correspond to a representation (for three different values for the tolerance, where $\epsilon = TOL$) of how many times the region had to be looked at by $\|f - ave(f)\|_{L^2}$ and $\|f\|_{BV}$ respectively, to obtain the desired tolerance. In this case, the regions with darker colors were worked on less than the regions with lighter colors. We can see that they give us what we would expect: the areas with more texture

or variation have the lighter colors, whereas the more homogeneous regions have the darker colors. (The values represent the number of times that region was visited.) In the third row we show a way to approximate the image f for the case when we have decomposed Ω using $\|f - ave(f)\|_{L^2}$. For each of the resulting small rectangles, we approximate f by its average on that small rectangle.

We have chosen to implement this idea into our algorithm. We have tried two different criteria to determine whether or not to work further on a given region. The first was to look at the BV-norm of the residual, $\|v_\lambda\|_{BV}$. The other value we considered was $F(u_\lambda)$. It is important to note that we used images that were of the size $2^i \times 2^i$ in our experiments to make it easier to implement (if the image's size is not of this form, reflection can be used to extend the image).

Let's say we are looking at $F(u_\lambda)$, then our algorithm works as follows: first find u_λ. If $F(u_\lambda) \leq TOL$, we stop, otherwise we split the region it into four equal regions as shown in figure (4.3). In other words, if our current image is of size 32×32 plus an extended boundary of 5 rows on each side, making it 42×42, then the four equal pieces would be of size 16×16 plus an extended of boundary of 5 rows on each side, making it 26×26.

The algorithm will now create four different images. Note that there is overlapping between these new images. The reasons for the overlapping are first, as we saw earlier, we need to expand the boundaries in order to obtain good results there, and also to ensure that when we put the pieces back together again, there won't be any obvious lines where the split was made. (Note that when the pieces are put back together, we ignore the expanded boundaries.) Now, for each of these regions, we first check to see if $F(u_\lambda) \leq TOL$. If it is, we stop. If not, we continue our multiscale decomposition, so we find $u_{2\lambda}$ which minimizes $F(u) = 2\lambda\|v_\lambda - u\|_{L^2} + \|u\|_{BV}$ (in the smaller region). Now we check

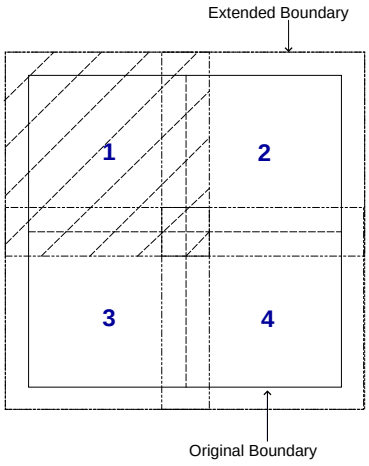

Figure 4.3: Method of splitting the domain into four equal regions. The shaded region represents the first image.

$F(u_{2\lambda}) \leq TOL$ and continue the splitting process if necessary. It is clear that this process can not go on forever because the size of f is finite. In our algorithm, the smallest region we work on is 2×2, and so if we have not reached the desired tolerance by that point, we just continue with our original implementation (without decomposing Ω). We see that in this manner we obtain a method which is essentially localized.

In figure (4.15), we use $\|v_\lambda\|_{BV}$, and in figure(4.14), we consider $F(u_\lambda)$. The first column represents the old algorithm (before we introduced spacial decomposition) for different values for the tolerance (where $\epsilon = TOL$). The middle column represents the way in which Ω was decomposed (again lighter means more texture), whereas the last column contains the new u. The resulting image u is therefore comprised of some regions which may need as much as 10 terms, while others only need 3. What is remarkable is how close this approximation is to the result of our original method, even with a fairly large tolerance. What is also interesting to note is that both approaches for the decomposition of Ω yield

similar results.

4.2 Numerical Discretization of the Hierarchical Decomposition of Color Images

Let us recall the system of Euler-Lagrange equations associated with

$$\inf_{\vec{u}=(u_1,u_2,u_3)\in BV(\Omega)} F_\lambda(\vec{u}) = \|\vec{u}\|_{BV(\Omega)} + \lambda \sum_{i=1}^{3} \|u_i - f_i\|^2_{L^2(\Omega)}$$

are

$$\begin{aligned} u_1 &= f_1 + \frac{1}{2\lambda}\ \mathrm{div}\ \left(\frac{\nabla u_1}{|\nabla \vec{u}|}\right), \\ u_2 &= f_2 + \frac{1}{2\lambda}\ \mathrm{div}\ \left(\frac{\nabla u_2}{|\nabla \vec{u}|}\right), \\ u_3 &= f_3 + \frac{1}{2\lambda}\ \mathrm{div}\ \left(\frac{\nabla u_3}{|\nabla \vec{u}|}\right). \end{aligned}$$

The numerical implementation of this case is exactly the same as in the previous case except that we solve three coupled equations simultaneously. Ultimately, the only things that change are the value of the c_i's. If $\vec{u} = (u_1, u_2, u_3)$, then for u_k, we have the following

$$\begin{aligned} c_{k;1} &= \frac{1}{\sqrt{\epsilon^2 + \sum_{m=1}^{3}\left((\frac{\triangle_+^x u^n_{m;i,j}}{h})^2 + (\frac{\triangle_0^y u^n_{m;i,j}}{h})^2\right)}}, \\ c_{k;2} &= \frac{1}{\sqrt{\epsilon^2 + \sum_{m=1}^{3}\left((\frac{\triangle_+^x u^n_{m;i-1,j}}{h})^2 + (\frac{\triangle_0^y u^n_{m;i-1,j}}{h})^2\right)}}, \\ c_{k;3} &= \frac{1}{\sqrt{\epsilon^2 + \sum_{m=1}^{3}\left((\frac{\triangle_0^x u^n_{m;i,j}}{h})^2 + (\frac{\triangle_+^y u^n_{m;i,j}}{h})^2\right)}}, \\ c_{k;4} &= \frac{1}{\sqrt{\epsilon^2 + \sum_{m=1}^{3}\left((\frac{\triangle_0^x u^n_{m;i,j-1}}{h})^2 + (\frac{\triangle_+^y u^n_{m;i,j-1}}{h})^2\right)}}. \end{aligned}$$

Solving for $u^{n+1}_{k:i,j}$, we obtain:

$$\begin{aligned} u^{n+1}_{k:i,j} &= \Big(\frac{1}{1+\frac{1}{2\lambda h^2}(c_{k:1}+c_{k:2}+c_{k:3}+c_{k:4})}\Big) \\ &\cdot \quad \Big[f_{k:i,j}+\frac{1}{2\lambda h^2}(c_{k:1}u^n_{k:i+1,j}+c_{k:2}u^n_{k:i-1,j}+c_{k:3}u^n_{k:i,j+1}+c_{k:4}u^n_{k:i,j-1})\Big]. \end{aligned}$$

If we look to figures (4.17) and (4.16), we see yet again how well the decomposition works, even for natural images.

4.3 Numerical Discretization of the Hierarchical Decomposition of Gray-Scale Images With Blurring

We follow the method of discretization in [26] and [7]. We implement our algorithm for this type of image in essentially the same way as for the case without noise. The only difference is that we have to deal with the blurring operator K, a Gaussian kernel. Recall that the Euler-Lagrange equation associated with

$$\inf_{u\in BV(\Omega)} F_\lambda(u) = \lambda\|f-Ku\|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)},$$

is

$$\begin{cases} K^*Ku_\lambda = K^*f + \frac{1}{2\lambda}\ \text{div}\left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\right) \text{ in } \Omega \\ \frac{\partial u_\lambda}{\partial N} = 0 \text{ on } \partial\Omega. \end{cases}$$

We will work with the case that K is a convolution-type integral operator, so in the numerical approximations, $(K_{mn})_{m,n=1,d}$ is a symmetric matrix with

$$\sum_{m,n=1}^{d} K_{mn} = 1,$$

and the approximation of Ku can be defined as

$$Ku_{i,j} = \sum_{m,n=1}^{d} K_{mn}u_{i+d/2-m,j+d/2-n}.$$

Since K is symmetric, we have $K^* = K$ and thus, $K^*Ku = KKu$, which is approximated by

$$KKu_{i,j} = \sum_{m,n=1}^{d} \sum_{r,t=1}^{d} K_{mn} K_{rt} u_{i+d-r-m,j+d-t-n}.$$

Using the same notation as before for c_i, we have

$$\begin{aligned} & 2\lambda h^2 K^* K u_{i,j}^{n+1} + (c_1(u_{i,j}^n) + c_2(u_{i,j}^n) + c_3(u_{i,j}^n) + c_4(u_{i,j}^n))u_{i,j}^{n+1} \\ = \; & c_1(u_{i,j}^n)u_{i+1,j}^{n+1} + c_2(u_{i,j}^n)u_{i-1,j}^{n+1} + c_3(u_{i,j}^n)u_{i,j+1}^{n+1} + c_4(u_{i,j}^n)u_{i,j-1}^{n+1} + 2\lambda h^2 K f_{i,j}. \end{aligned}$$

Figure (4.18) illustrates how the decomposition works for blurred images. We can see that it has become sharper, but it will still never recover the lost texture due to the blurring.

4.4 Numerical Discretization of the Multiplicative Hierarchical Decomposition

The associated Euler-Lagrange equation for

$$\inf_{u \in BV(\Omega)} F_\lambda(u) = \lambda \| \frac{f}{u} - 1 \|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)},$$

is

$$-\frac{f^2}{u^3} + \frac{f}{u^2} = \frac{1}{2\lambda} \mathrm{div}\Big(\frac{\nabla u}{\sqrt{\epsilon^2 + |\nabla u|^2}} \Big), \tag{4.3}$$

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

$$\frac{\partial u}{\partial t} = \frac{f^2}{u^3} - \frac{f}{u^2} + \frac{1}{2\lambda} \mathrm{div}\Big(\frac{\nabla u}{\sqrt{\epsilon^2 + |\nabla u|^2}} \Big), \;\; u(x, y, 0) = f(x, y). \tag{4.4}$$

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:

$$\begin{aligned}\frac{u_{i,j}^{n+1}-u_{i,j}^{n}}{\triangle t} &= \frac{f_{i,j}^{2}}{(u_{i,j}^{n})^{3}}-\frac{f_{i,j}}{(u_{i,j}^{n})^{2}} \\ &\quad +\frac{1}{2\lambda h^{2}}(c_1u_{i+1,j}^{n}+c_2u_{i-1,j}^{n}+c_3u_{i,j+1}^{n}+c_4u_{i,j-1}^{n}) \\ &\quad -\frac{1}{2\lambda h^{2}}(c_1+c_2+c_3+c_4)u_{i,j}^{n+1},\end{aligned}$$

or, solving again for $u_{i,j}^{n+1}$,

$$\begin{aligned}u_{i,j}^{n+1} &= \Big(\frac{1}{1+\frac{\triangle t}{2\lambda h^{2}}(c_1+c_2+c_3+c_4)}\Big)\cdot\Big[u_{i,j}^{n}+\triangle t\frac{f_{i,j}^{2}}{(u_{i,j}^{n})^{3}}-\triangle t\frac{f_{i,j}}{(u_{i,j}^{n})^{2}} \\ &\quad +\frac{\triangle t}{2\lambda h^{2}}(c_1u_{i+1,j}^{n}+c_2u_{i-1,j}^{n}+c_3u_{i,j+1}^{n}+c_4u_{i,j-1}^{n})\Big].\end{aligned}$$

Now, because we are dividing through by u, we need to make sure we don't divide by zero. One way to do that is to add some positive constant to f, and then at the end, subtract that constant from the final result.

In this section, we have applied our method to the image f in figure (4.19). 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 (see Remark 3.2.2).

4.5 Numerical Discretization of the Hierarchical Decomposition of the Ambrosio-Tortorelli Approximation to the Mumford-Shah Model

Our algorithm implements the Ambrosio-Tortorelli model

$$G_{\rho}^{AT}(u,w)=\int_{\Omega}[\rho|\nabla w|^{2}+\alpha(w^{2}|\nabla u|^{2}+\frac{(w-1)^{2}}{4\alpha\rho})+\beta|u-f|^{2}]dxdy,$$

which has the following Euler-Lagrange equations

$$\begin{cases} \beta u - \alpha\nabla(w^2\nabla u) = \beta f \\ -\triangle w + \frac{1+4\alpha\rho|\nabla u|^2}{4\rho^2}(w - \frac{1}{1+4\alpha\rho|\nabla u|^2}) = 0. \end{cases}$$

Discretizing the Euler-Lagrange equations, we obtain

$$\begin{cases} \beta f_{i,j} = \beta u_{i,j} - \alpha\frac{\triangle_-^x(w_{i,j}^2\frac{\triangle_+^x u_{i,j}}{h})}{h} - \alpha\frac{\triangle_-^y(w_{i,j}^2\frac{\triangle_+^y u_{i,j}}{h})}{h} \\ \frac{1}{h^2}\triangle_-^x\triangle_+^x w_{i,j} + \triangle_-^y\triangle_+^y w_{i,j} = \frac{1+4\alpha\rho\frac{(\triangle_0^x u_{i,j}+\triangle_0^y u_{i,j})}{h^2}}{4\rho^2}(w_{i,j} - \frac{1}{1+4\alpha\rho\frac{(\triangle_0^x u_{i,j}+\triangle_0^y u_{i,j})}{h^2}}). \end{cases}$$

Using the notation

$$\begin{aligned} C_1 &= \beta + \frac{\alpha}{h^2}(2w_{i,j}^2 + w_{i-1,j}^2 + w_{i,j-1}^2) \\ C_2 &= 1 + 4\alpha\rho\sqrt{\Big(\frac{u_{i+1,j} - u_{i-1,j}}{2h}\Big)^2 + \Big(\frac{u_{i,j+1} - u_{i,j-1}}{2h}\Big)^2} + \frac{16\rho^2}{h^2}, \end{aligned}$$

we have

$$\begin{cases} u_{i,j} = \frac{1}{C_1}\Big[\beta f_{i,j} + \frac{\alpha}{h^2}(w_{i,j}^2(u_{i+1,j} + u_{i,j+1}) + w_{i-1,j}^2 u_{i-1,j} + w_{i,j-1}^2 u_{i,j-1})\Big] \\ w_{i,j} = \frac{1}{C_2}\Big[1 + \frac{4\rho^2}{h^2}(w_{i+1,j} + w_{i-1,j} + w_{i,j+1} + w_{i,j-1})\Big]. \end{cases}$$

In order to minimize grid effect we alternate the above formulation with the following

$$\begin{aligned} C_1 &= \beta + \frac{\alpha}{h^2}(2w_{i,j}^2 + w_{i+1,j}^2 + w_{i,j+1}^2) \\ C_2 &= 1 + 4\alpha\rho\sqrt{\Big(\frac{u_{i+1,j} - u_{i-1,j}}{2h}\Big)^2 + \Big(\frac{u_{i,j+1} - u_{i,j-1}}{2h}\Big)^2} + \frac{16\rho^2}{h^2}, \end{aligned}$$

and

$$\begin{cases} u_{i,j} = \frac{1}{C_1}\Big[\beta f_{i,j} + \frac{\alpha}{h^2}(w_{i,j}^2(u_{i-1,j} + u_{i,j-1}) + w_{i+1,j}^2 u_{i+1,j} + w_{i,j+1}^2 u_{i,j+1})\Big] \\ w_{i,j} = \frac{1}{C_2}\Big[1 + \frac{4\rho^2}{h^2}(w_{i+1,j} + w_{i-1,j} + w_{i,j+1} + w_{i,j-1})\Big]. \end{cases}$$

This was obtained by substituting $\triangle_+$ for $\triangle_-$ (and vice-versa) in the discretization of the first Euler-Lagrange equation for this minimization problem. In figures (4.20) and (4.21), we show how well this method works. We can clearly see that we converge to the desired image as well as obtaining a function which represents the contours of u.

4.6 Conclusion

In conclusion, we have proposed and implemented a new multiscale image representation method based on hierarchical decompositions. We have shown that by extracting information from the previously discarded residuals of existing methods, we are able to obtain better decompositions. Finally, we have included some examples to illustrate our multiscale decomposition.

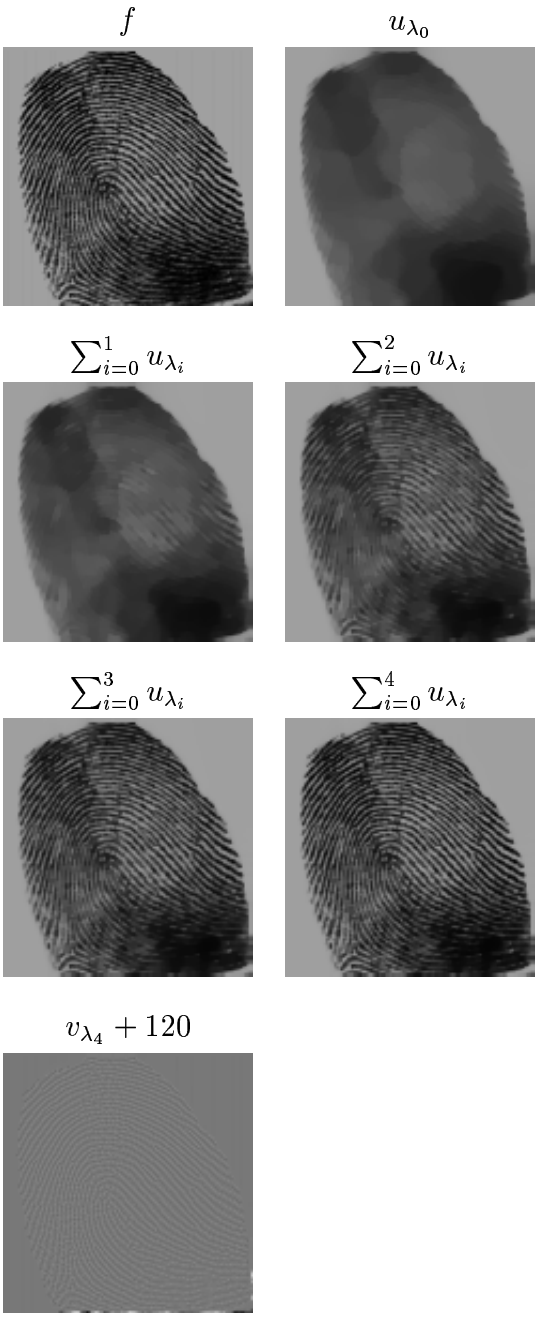

Figure 4.4: Decomposition of an initial image of a fingerprint with the original boundary conditions for 5 steps. Parameters: $\lambda_0 = .01$, and $\lambda_k = 2^k \lambda_0$.

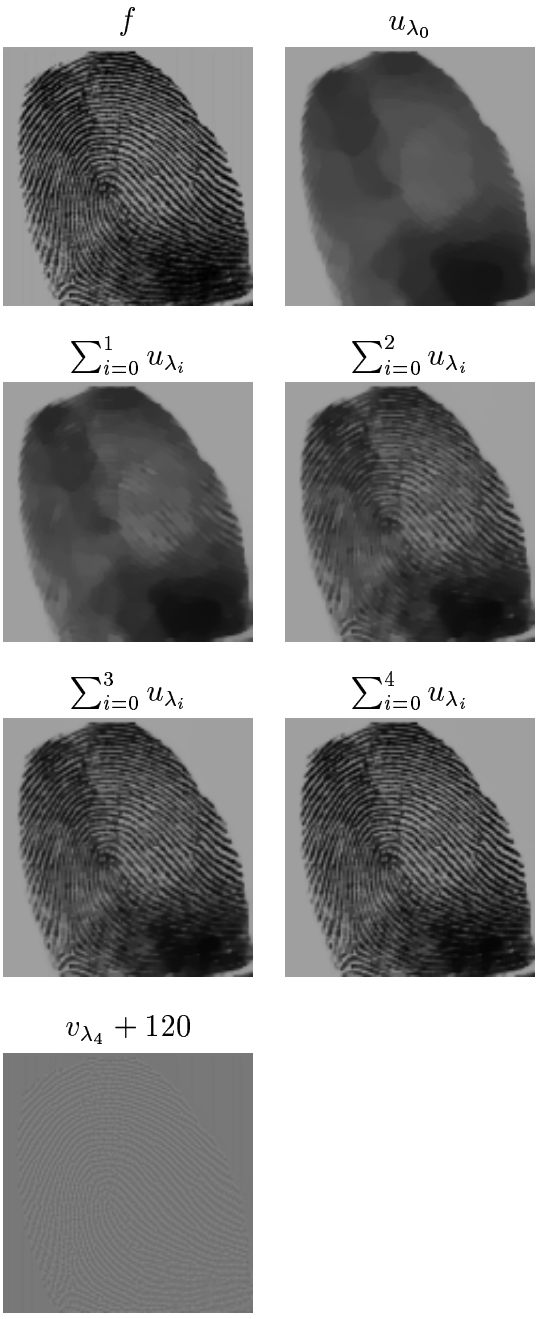

Figure 4.5: Decomposition of an initial image of a fingerprint with the improved boundary conditions for 5 steps. Parameters: $\lambda_0 = .01$, and $\lambda_k = 2^k \lambda_0$

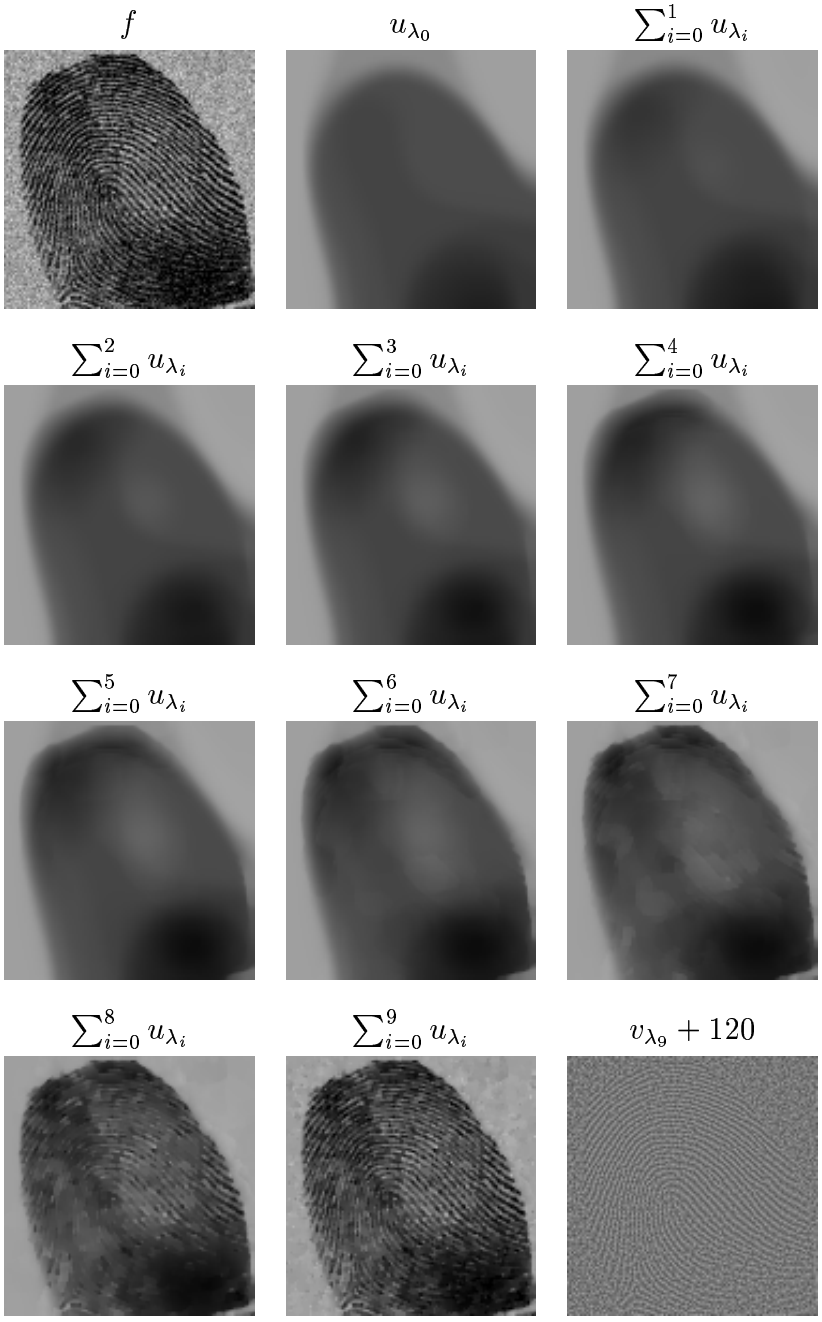

Figure 4.6: Decomposition of a noisy image of a fingerprint for 10 steps. Parameters: $\lambda_0 = .0001$, and $\lambda_k = 2^k \lambda_0$

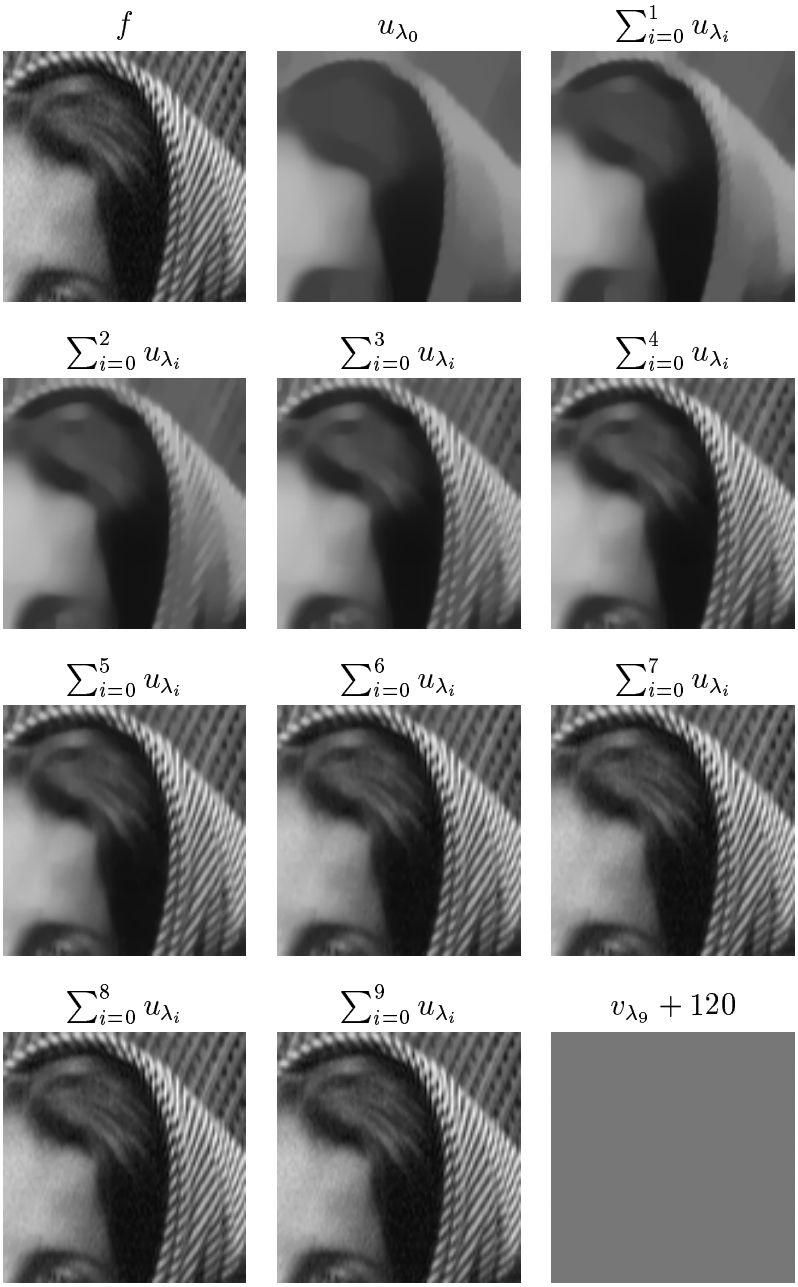

Figure 4.7: Decomposition of an initial image of a woman with improved boundary conditions for 10 steps. Parameters: $\lambda_0 = .005$, and $\lambda_k = 2^k \lambda_0$

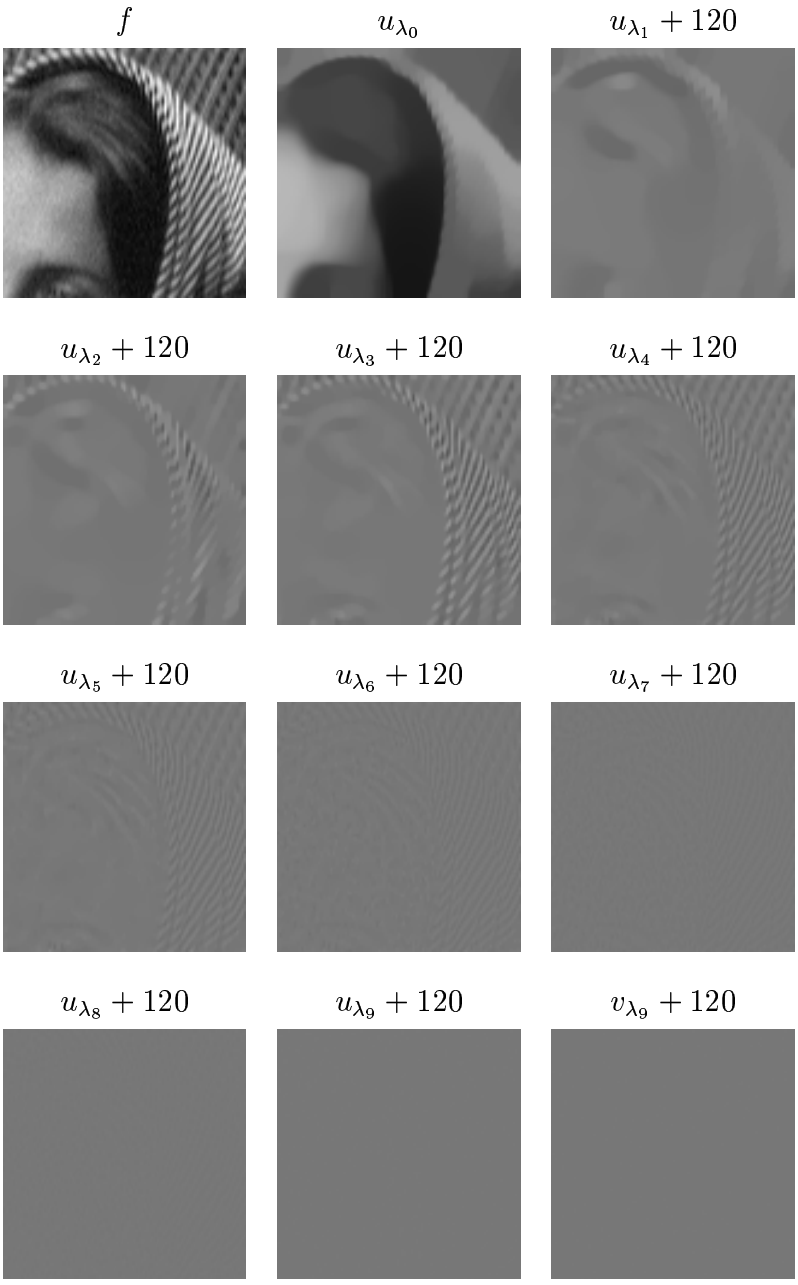

Figure 4.8: Representation of each u_{λ_i}, for $0 \leq k < 10$. Parameters: $\lambda_0 = .005$, and $\lambda_k = 2^k \lambda_0$

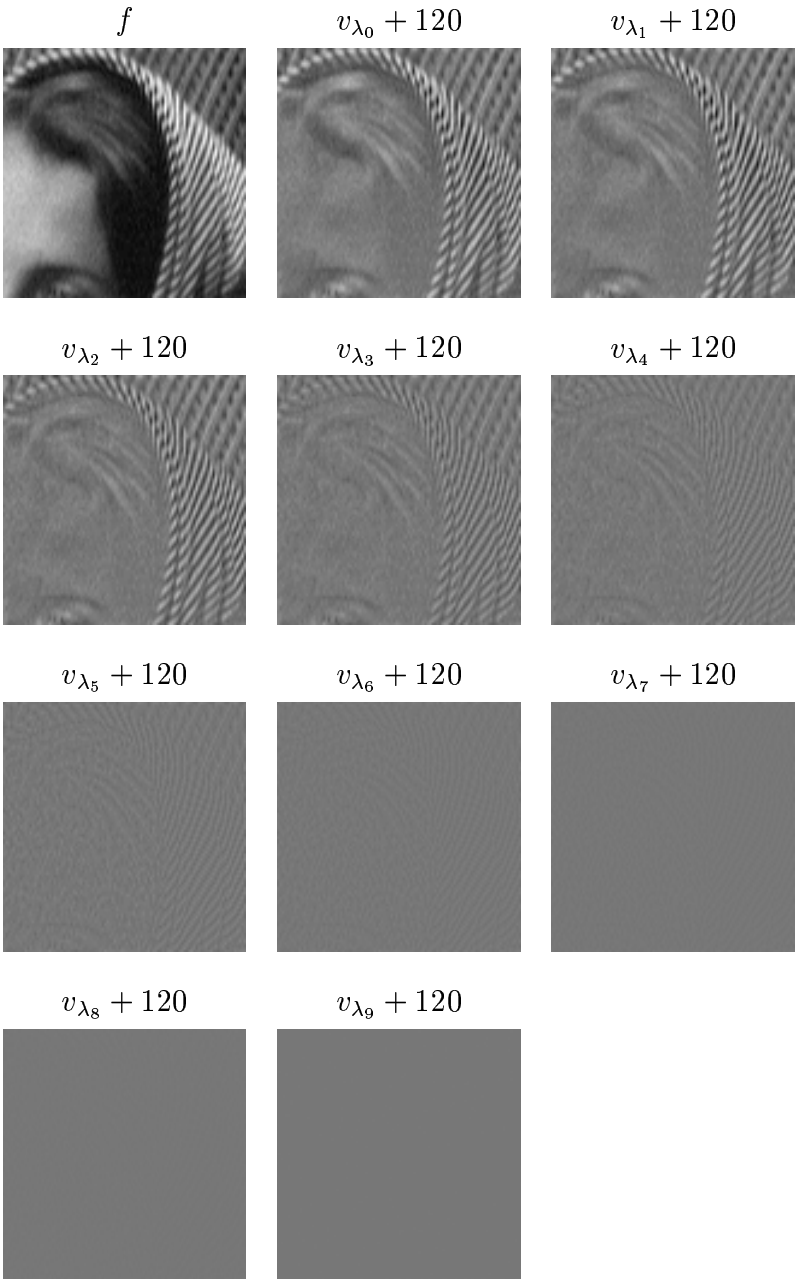

Figure 4.9: Representation of each v_{λ_i}, for $0 \leq k < 10$. Parameters: $\lambda_0 = .005$, and $\lambda_k = 2^k \lambda_0$

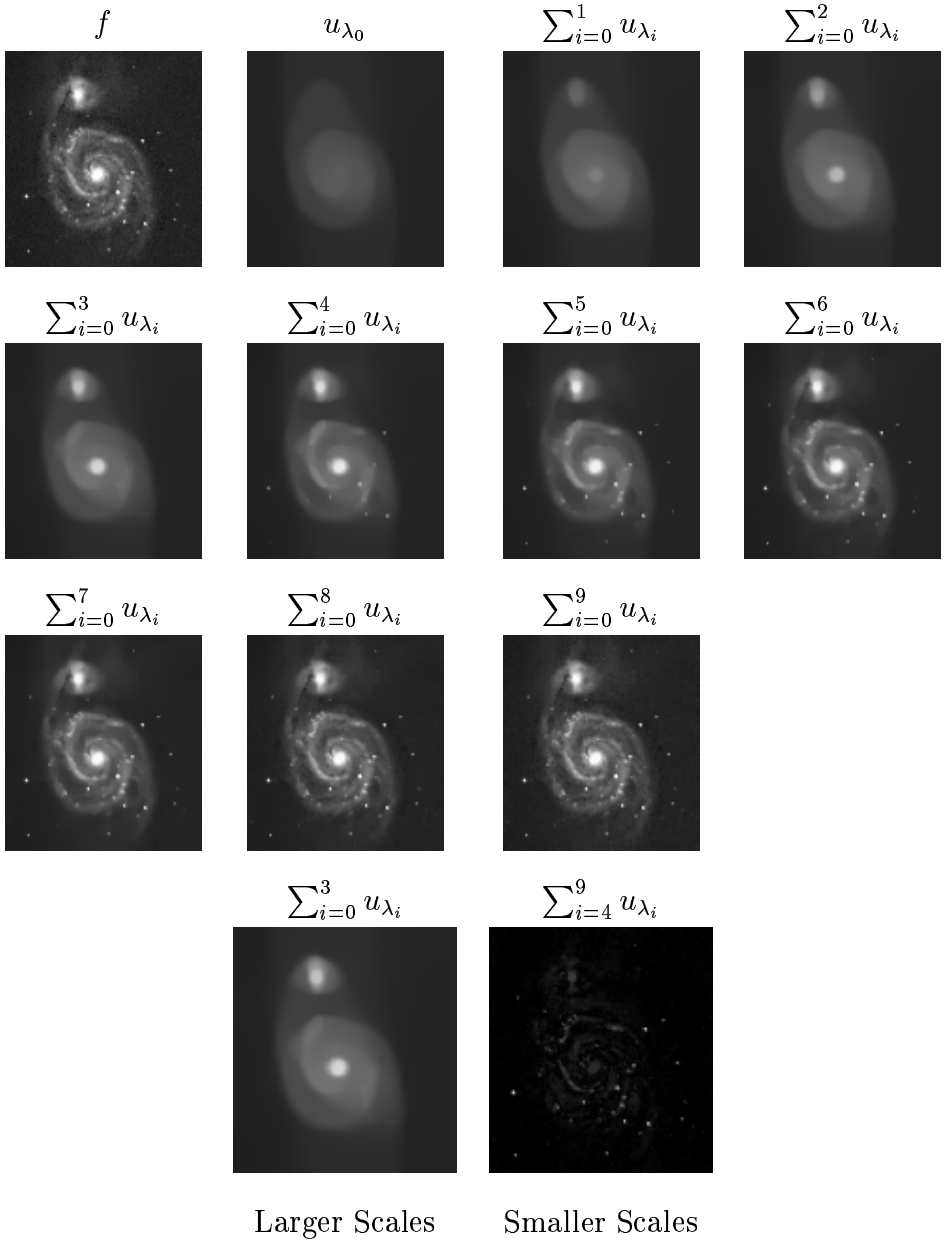

Figure 4.10: Decomposition of an image of a galaxy for 10 steps. Parameters: $\lambda_0 = .001$, and $\lambda_k = 2^k \lambda_0$. The bottom row represents the splitting of scales.

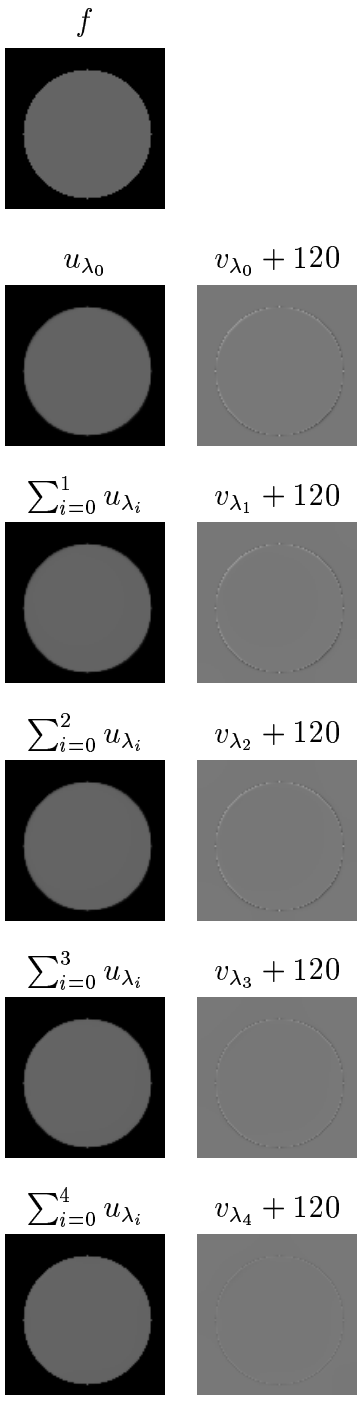

Figure 4.11: As described in Example (3.2.1), given an initial image of a circle, these represent the u_{λ_i} components and the residuals, v_{λ_i} for 5 steps. Parameters: $\lambda_0 = .01$, and $\lambda_k = 2^k \lambda_0$.

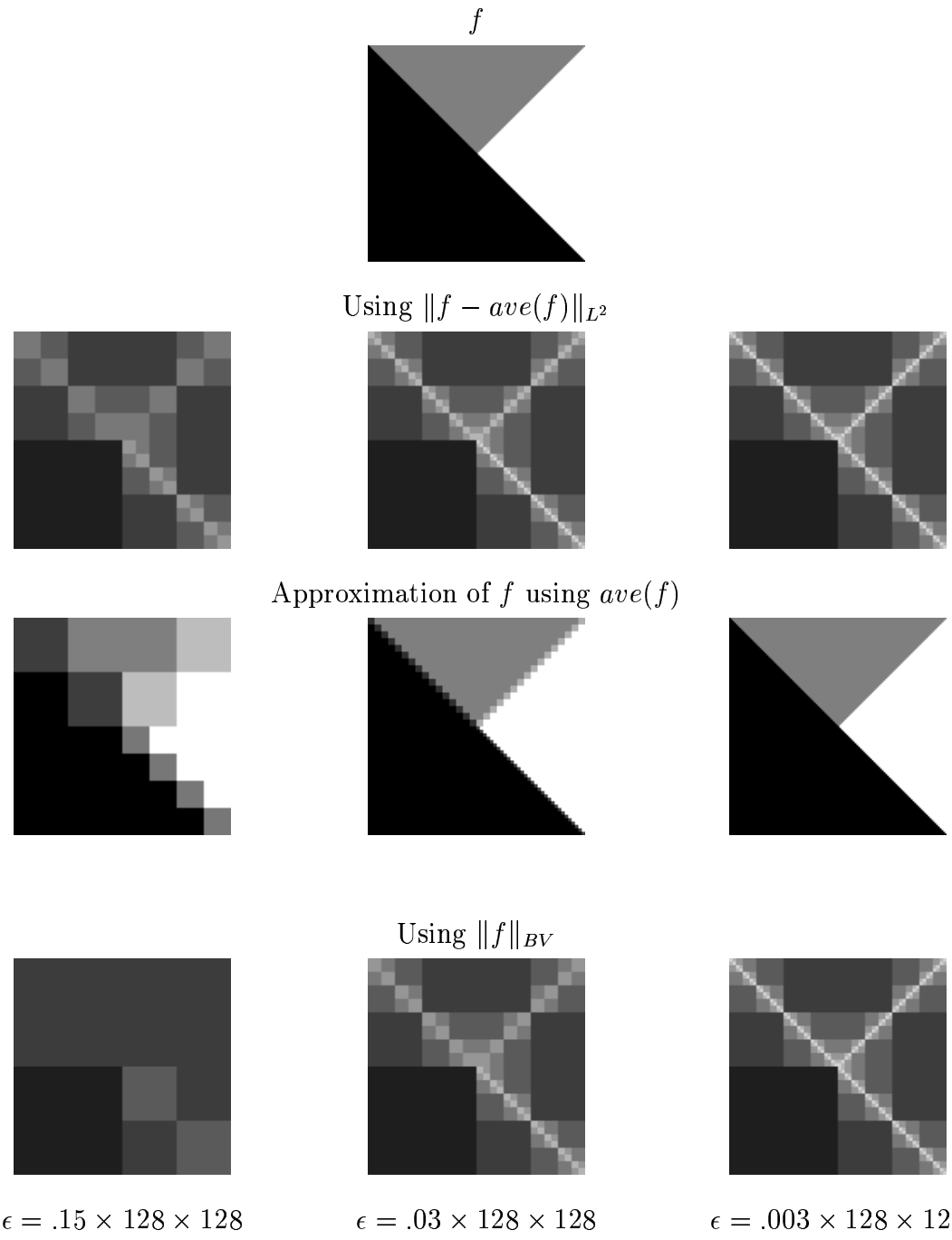

Figure 4.12: Decomposition of Ω by the methods described in [11] for a synthetic image.

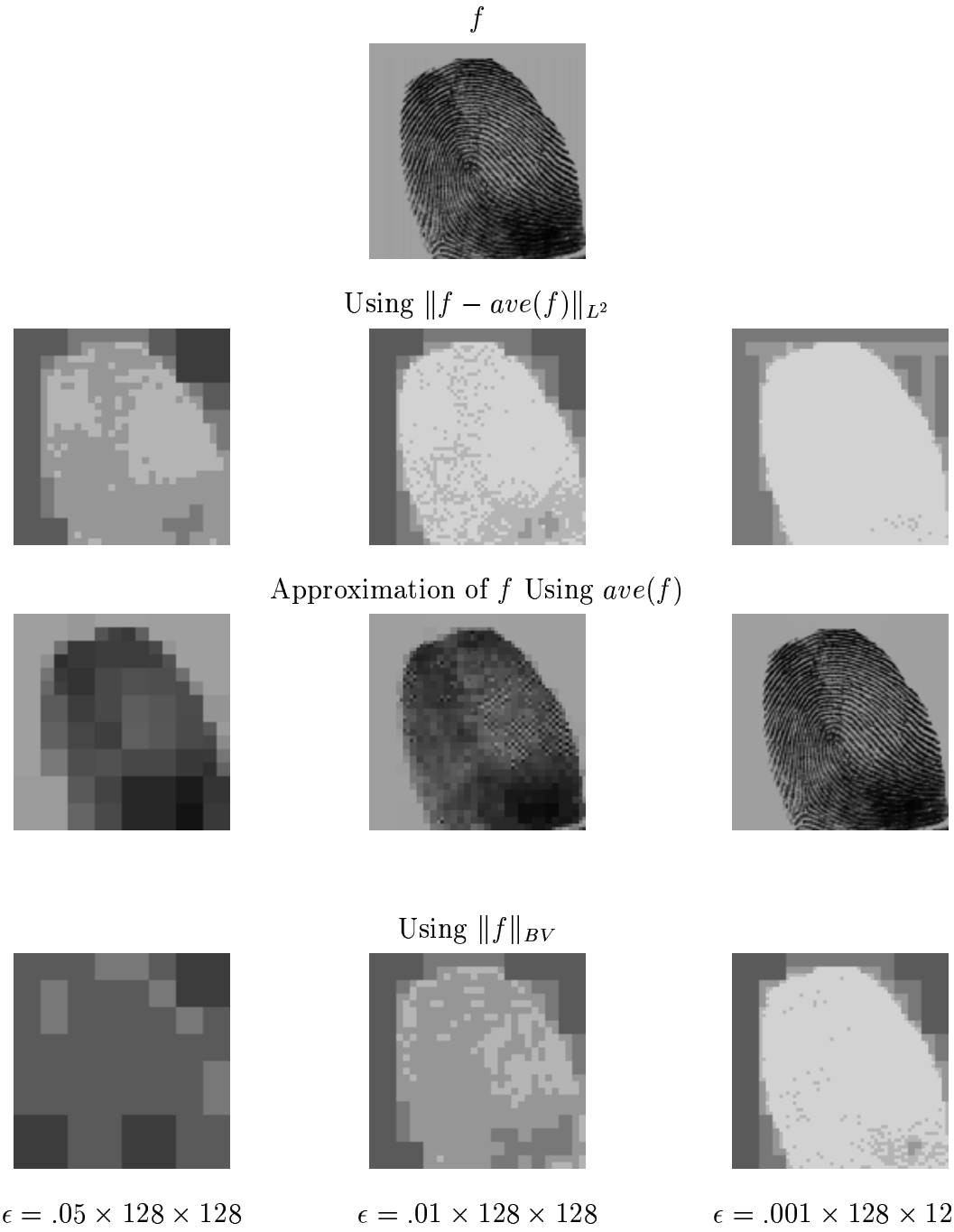

Figure 4.13: Decomposition of Ω by the methods described in [11] for a fingerprint image.

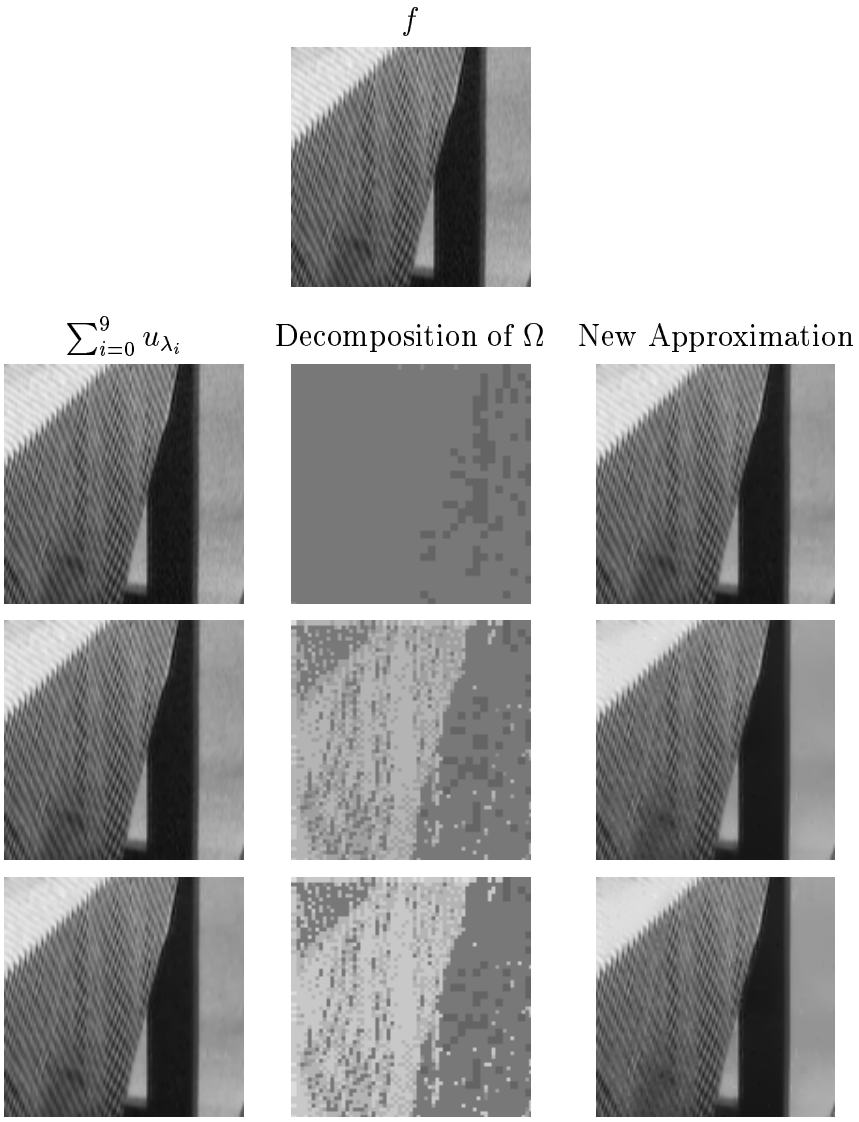

Figure 4.14: Decomposition of f and Ω using $F(u_\lambda)$ and the new u resulting from it, for a maximum of 10 steps. We use $\epsilon = 50 \times 128 \times 128$ for all calculations. Parameters: Row 1: $\lambda_0 = .01$, Row 2: $\lambda_0 = .001$ and Row 3: $\lambda_0 = .0005$, where $\lambda_k = 2^k \lambda_0$.

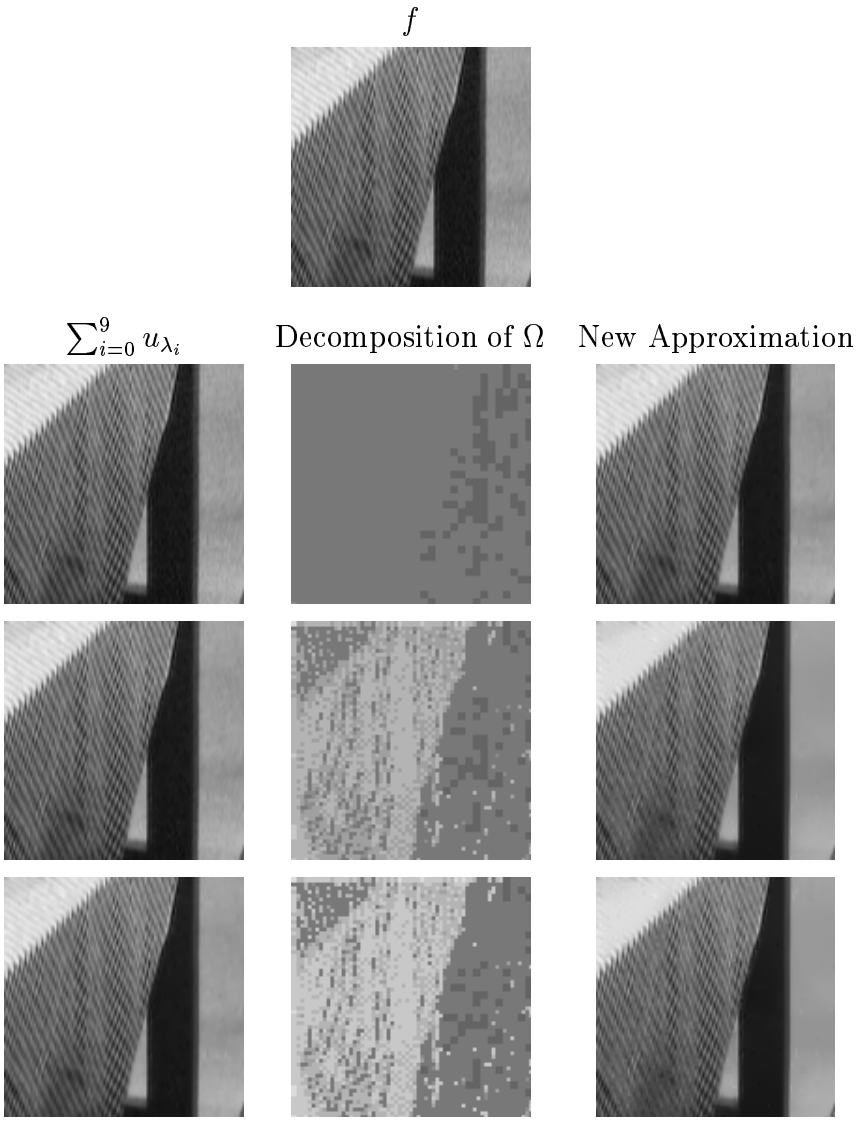

Figure 4.15: Decomposition of f and Ω using the total variation of v_λ and the new u resulting from it, for a maximum of 10 steps. We use $\epsilon = 50 \times 128 \times 128$ for all calculations. Parameters: Row 1: $\lambda_0 = .01$, Row 2: $\lambda_0 = .001$, and Row 3: $\lambda_0 = .0005$, where $\lambda_k = 2^k \lambda_0$

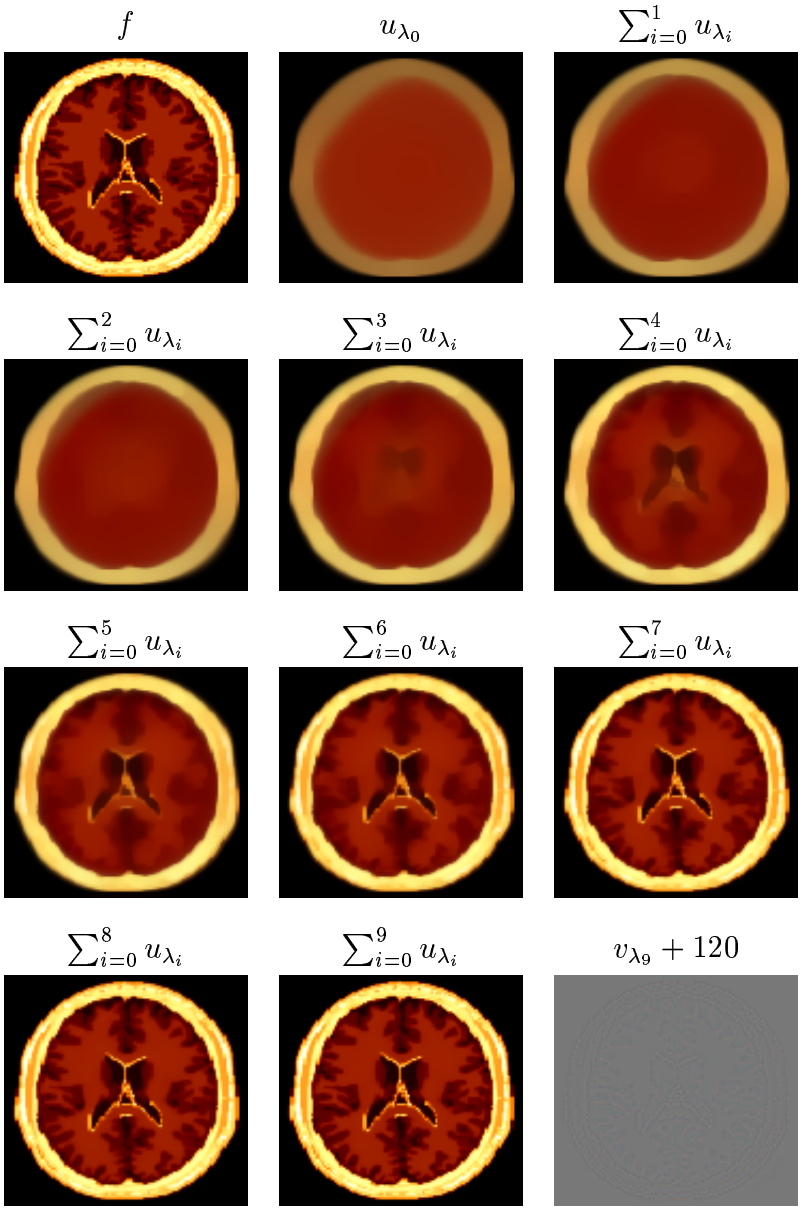

Figure 4.16: Decomposition of a vector-valued MRI image, for 10 steps. Parameters: $\lambda_0 = .00025$, and $\lambda_k = 2^k \lambda_0$

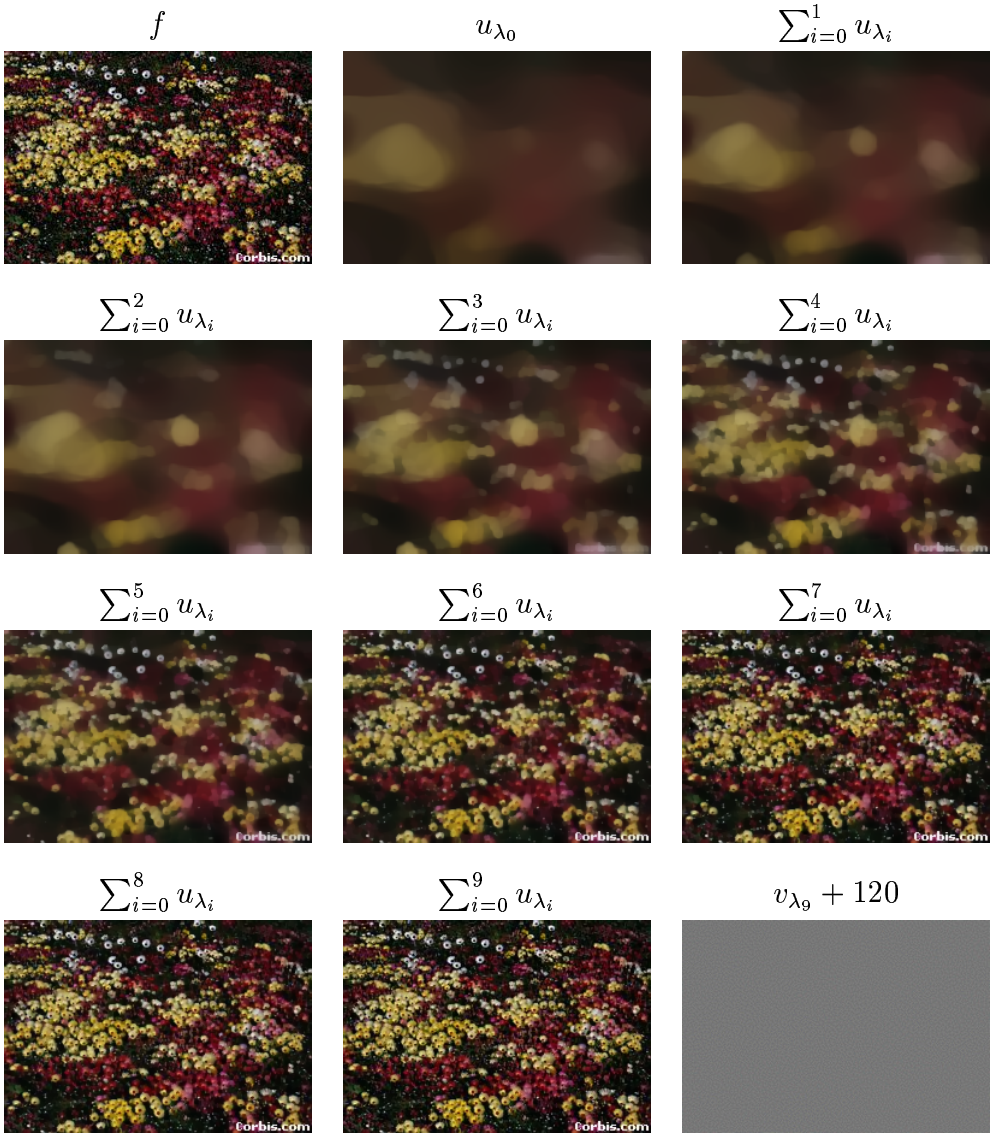

Figure 4.17: Decomposition of a vector-valued image of flowers for 10 steps. Parameters: $\lambda_0 = .00025$, and $\lambda_k = 2^k \lambda_0$

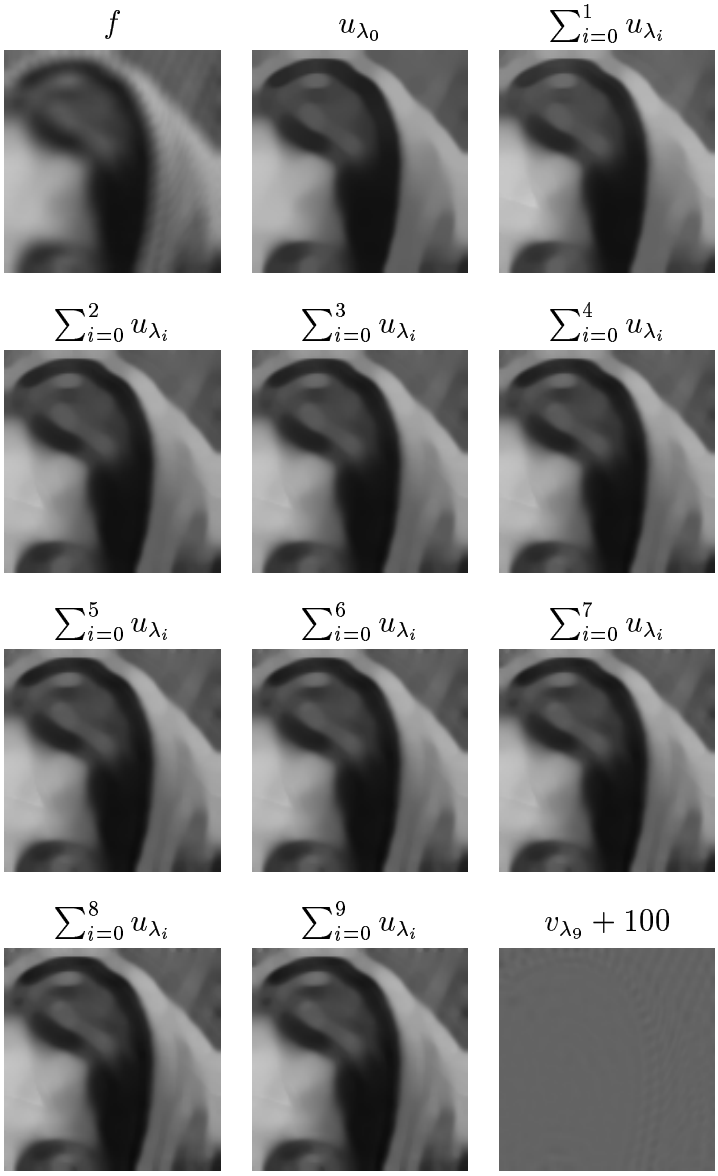

Figure 4.18: The recovery of u from a blurred initial image using 10 steps. Parameters: $\lambda_0 = .1$, $d = 5$ and $\lambda_k = 10^k \lambda_0$.

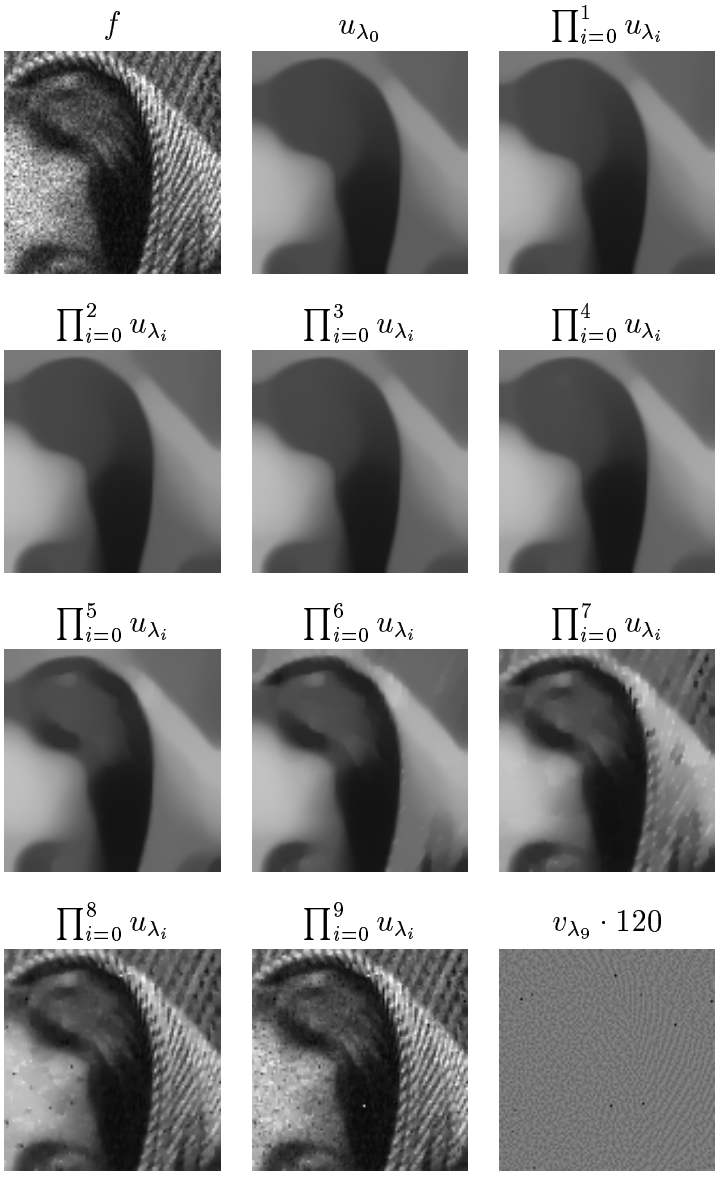

Figure 4.19: The recovery of u given an initial image of a woman with multiplicative noise, for 10 steps. Parameters: $\lambda_0 = .02$, and $\lambda_k = 2^k \lambda_0$. Note that by using a finer grid (see Remark (3.2.2)), it is possible to obtain a better denoised image.

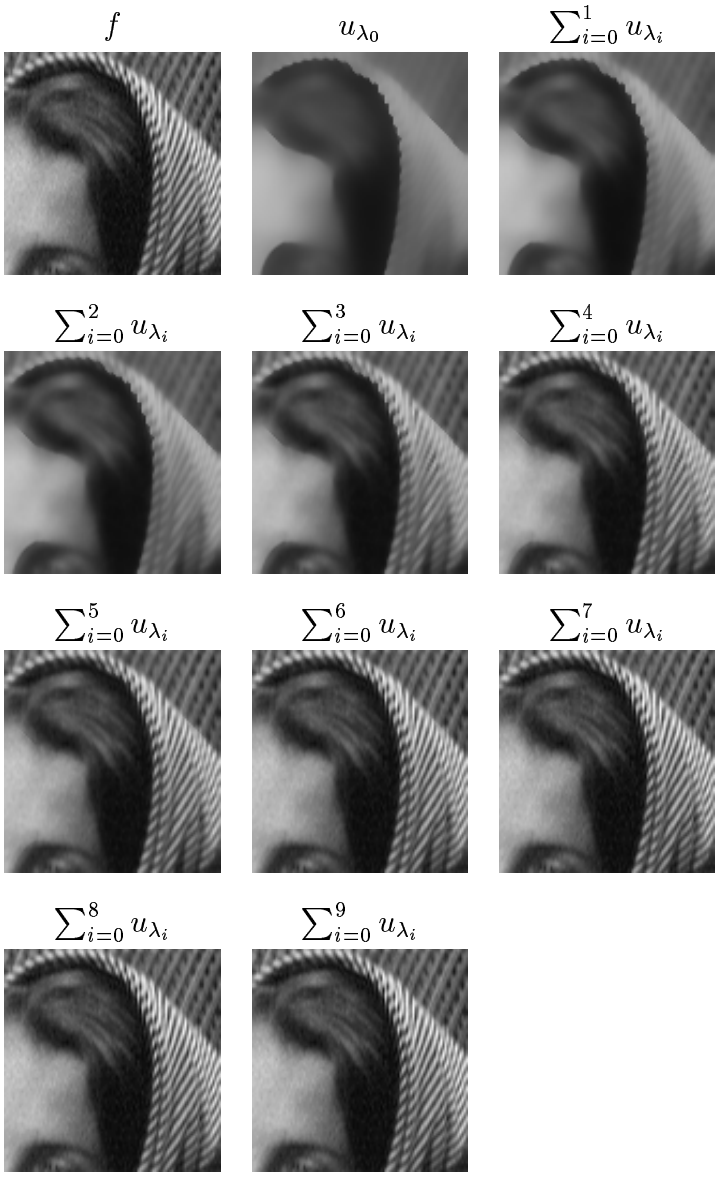

Figure 4.20: The sum of the u_i's using the Ambrosio-Tortorelli approximation of the image of a woman, using 10 steps. Parameters: $\beta_0 = .25$, $\alpha = 5$, $\rho = .0002$, and $\beta_k = 2^k \beta_0$

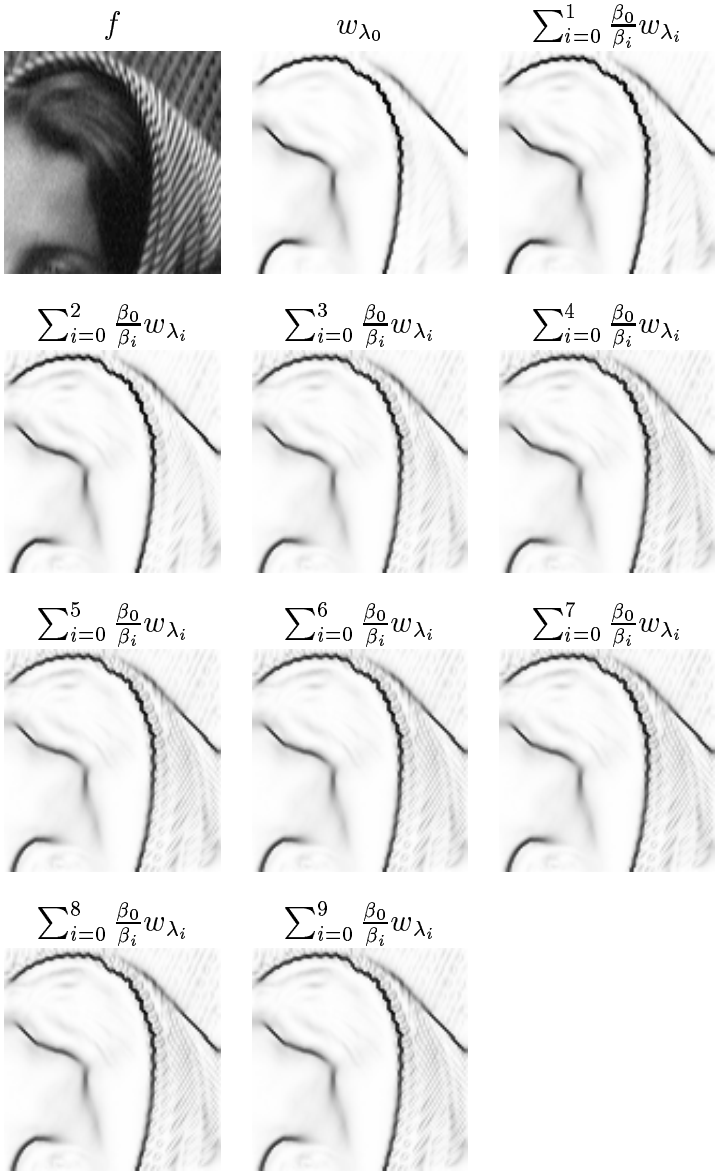

Figure 4.21: The weighted sum of the w_i's using the Ambrosio-Tortorelli approximation of the image of a woman, using 10 steps. Parameters: $\beta_0 = .25$, $\alpha = 5$, $\rho = .0002$, and $\beta_k = 2^k \beta_0$

References

[1] R. Acar and C.R. Vogel, *Analysis of Bounded Variation Penalty Methods of Ill-Posed Problems*, Inverse Problem 10, pp. 1217-1229, 1994.

[2] L Alvarez, F. Guichard, P.L. Lions and J.-M. Morel, *Axioms and Fundamental Equations of Image Processing*, Archive of Rational Mechanics and Analysis, 123(3), pp. 199-257, 1993.

[3] L. Ambrosio, *A Compactness Theorem for a Special Class of Functions of Bounded Variation*, Bollettino U.M.I. (7), 3-B, pp. 857-881, 1989.

[4] L. Ambrosio and V. Tortorelli, *Approximation of Functionals Depending on Jumps by Elliptic Functionals via* Γ*-Convergence*, Communications on Pure and Applied Mathematics, Vol. XLIII, pp. 999-1036, 1990.

[5] L. Ambrosio and V. Tortorelli, *On the Approximation of Free Discontinuity Problems*, Bollettino U.M.I. (7), 6-B, pp. 105-123, 1992.

[6] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Applied Mathematical Sciences v. 147 (Springer-Verlag New York, Inc.), 2002.

[7] G. Aubert and L. Vese, *A Variational Method in Image Recovery*, SIAM Journal on Numerical Analysis, 34(5), pp. 1948-1979, October 1997.

[8] A.Z. Averbuch, R.R. Coifman and F.G. Meyer, *Multilayered Image Representation: Application to Image Compression*, IEEE Transactions on Image Processing, Vol. 11, No. 9, September 2002.

[9] F. Catté, P.L. Lions, J-M. Morel and T. Coll, *Image Selective Smoothing and Edge Detection by Nonlinear Diffusion*, SIAM Journal on Numerical Analysis, 29(1), pp. 182-193, February 1992.

[10] A. Chambolle and P.L. Lions, *Image Recovery via Total Variation Minimization and Related Problems*, Numer. Math. 76, pp.167-188, 1997.

[11] A. Cohen, R. DeVore, P. Petrushev and H. Xu, *Nonlinear Approximation and the Space* $BV(\mathbb{R}^2)$, American Journal of Mathematics 121, pp. 587-628, 1999.

[12] G. Dal Maso, J.-M. Morel and S. Solimini, *A Variational Method in Image Segmentation. Existence and Approximation Results*, Acta Math.

[13] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, London 1992.

[14] F.G. Meyer, A.Z. Averbuch, J.O. Strömburg and R.R. Coifman, *Multi-Layered Image Representation: Application to Image Compression*, Proceedings, International Conference on Image Procession ICIP 1998, Volume: 2, pp. 292-296, 4-7 Oct. 1998.

[15] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*,SIAM Journal of University Lecture Series Volume 22, AMS 2002.

[16] J.-M. Morel and S. Solimini, *Variational Methods in Image Segmentation*, Progress in Nonlinear Differential Equations and Their Applications, Vol. 14, Birdhäuser Boston, 1995.

[17] D. Mumford and J. Shah, *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems*, Communications on Pure and Applied Mathematics, Vol. XLVII, pp. 577-685, 1989.

[18] S. Osher, A. Solé and L. Vese, *Image Decomposition and Restoration Using Total Variation Minimization and the H^{-1} Norm*, UCLA CAM Report 02-57, October 2002, to appear in SIAM MMS.

[19] P. Perona and J. Malik, *Scale-Space and Edge Detection Using Anisotropic Diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7), pp. 629-639, 1990.

[20] L. Rudin and S. Osher, *Total Variation Based Image Restoration with Free Local Constraints*, Proc. IEEE ICIP, Vol I, pp. 31-35, Austin (Texas) USA, 1994.

[21] L. Rudin, S. Osher and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, 60, pp. 259-268, 1992.

[22] D. Strong and T. Chan, *Relation of Regularization Parameter and Scale in Total Variation Based Image Denoising*, UCLA Math Department CAM Report 96-7, February 1996.

[23] D. Strong and T. Chan, *Exact Solutions to Total Variation Regularization Problems*, UCLA Math Department CAM Report 96-41, October 1996.

[24] D. Strong and T. Chan, *Spatially and Scale Adaptive Total Variation Based Regularization and Anisotropic Diffusion in Image Processing*, UCLA Math Department CAM Report 96-46, November 1996.

[25] E. Tadmor, S. Nezzar and L. Vese, *A Multiscale Image Representation Using Hierarchical (BV,L2) Decompositions*, UCLA CAM Report, June 2003.

[26] L. Vese, *A Study in the BV Space of a Denoising-Deblurring Variational Problem*, Applied Mathematics and Optimization, 44(2), pp. 131-161, 2001.

[27] L. Vese and S. Osher, *Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing*, UCLA CAM Report 02-19, May 2002, to appear in Journal of Scientific Computing.