Proceedings of ALGORITMY 2009
pp. 226–235

MIXED REGULARIZATION METHOD FOR IMAGE RESTORATION

IVAN CIMRÁK* AND VALDEMAR MELICHER†

Abstract. We present a new mixed regularization method for image recovery. The method is based on the combination of the bounded variation regularization and the quadratic H^k regularization. We show motivation for two distinctive terms in the energy functional, one for each regularization. We obtain rigorous results on well-posedness and stability of the underlying minimization problem. The numerical results for several case-studies give significant improvement over standard single regularizations.

Key words. Bounded variation regularization, image reconstruction, denoising

AMS subject classifications. 68U10, 49M27, 65K10

1. Model problem. In the image restoration, two basic methods are widely used: The PDE based method and the energy method. We focus on the latter one. In this approach one constructs an energy functional which is afterwards minimized. The obtained minimum is then the reconstructed image.

The energy corresponding to a particular problem often consists of two terms: A fidelity term to the data and a regularization term. The latter one ensures the desired smoothness of the solution. Regularization of the problem is necessary since the data is often noisy and in that case the problem becomes an ill-posed inverse problem.

We denote by u the original picture and by f the noisy data. The energy to be minimized thus typically takes the following form

$$E = \int_\Omega \|f-u\|_2^2 + \alpha \mathcal{R}(u). \tag{1.1}$$

The first term measures the fidelity to the data and the second term is the regularization term. The parameter α is a real positive weighting factor which controls the tradeoff between the smoothness and the fidelity.

The choice of the regularization term is crucial for the quality of the solution. For example in the case of

$$\mathcal{R}(u) = \|\nabla u\|_2^2,$$

the noise can be satisfactory removed, but at the same time the solution is smoothed and belongs to the Sobolev space $H^1(\Omega)$. This effect called also an *oversmoothing effect* occurs whenever the regularization term is chosen to be

$$\mathcal{R}(u) = \|\mathcal{L}u\|_2^2,$$

where $\mathcal{L}$ is a differential operator. We refer to this kind of regularizations as H^k regularizations.

Real pictures however contain sharp edges. Therefore we need to choose other regularization which leads to the solution belonging to the class of discontinuous functions.

*I. Cimrák is supported by the Fund for Scientific Research - Flanders FWO, Belgium.
†V. Melicher is supported by the GOA project BOF 07/GOA/006 of Ghent University.

In the pioneering work of Rudin et al. [10], the authors introduced the bounded variation (BV) regularization in the context of image processing. The BV regularization successfully preserves sharp edges in the picture. The regularization term takes the form

$$\mathcal{R}(u) = J(u),$$

where $J(u)$ is the BV semi-norm. For detailed definitions of the space $BV(\Omega)$ and its norm and semi-norm see Section 2. For smooth functions, the BV semi-norm can be expressed as

$$J(u) = \int_\Omega |\nabla u|.$$

Minimizing of $J(u)$ penalizes functions with large gradients, however it allows for the jumps as presented in the following example.

EXAMPLE 1. *One of the solutions to the following one dimensional problem on the interval* $\Omega = (0, 1)$

$$\inf_{u \in BV(\Omega), u(0)=0, u(1)=1} J(u),$$

is the function

$$u(s) = \begin{cases} 0 & \text{for } 0 \le s < b, \\ 1 & \text{for } b \le s \le 1, \end{cases}$$

where b is a real number between 0 and 1.

The BV regularization has drawbacks too. The above example illustrates the so called *staircase effect*. Not only the function presented in the example is the solution but also any "staircase-like" non-decreasing piecewise constant function with value 0 for $s = 0$ and value 1 for $s = 1$. In two dimensions, the staircase effect results in flat regions with constant color. This effect occurs when the regularization term in (1.1) is chosen to be $\mathcal{R}(u) = J(u)$.

To illustrate the oversmoothing and the staircase effect we subsequently applied both regularizations on the benchmark picture Lenna. Figure 1.1a depicts the original image and 1.1b shows the version degraded by 5% noise. In Figure 1.1c one can see the smoothed reconstruction when the H^2 regularization is used. In this case $\mathcal{L}$ was taken to be the Laplacian so $\mathcal{L}u = \Delta u$. Here, the noise has almost been eliminated however the resulting image is blurred. The BV reconstruction in Figure 1.1d nicely preserves the edges, however the staircase effect is clearly present leading to a non-satisfactory reconstruction.

Natural solution to eliminate both undesired effects is to introduce a function Φ and to set the regularization term to be

$$\mathcal{R}(u) = \int_\Omega \Phi(|\nabla u|).$$

The function $\Phi(s)$ should locally behave such that it is close to $\Phi(s) = s$ on the places where an edge occurs resulting in the local BV regularization, and it is close to $\Phi(s) = s^2$ on the places where no edge is expected resulting in the local H^1 regularization.

There are many choices for Φ, however, they must conform to some hypotheses to give reasonable results in image recovery. These hypotheses have been formulated in [3] to satisfy the following principle in image analysis:

The reconstructed image must be formed by homogeneous regions, separated by sharp edges.

In [9] the authors defined more similar hypotheses in order to conform to the more specific principle:

The restored object can be decomposed into three scales–flat, grey and sharp edges.

Charbonnier, in [7], presents many choices used in image reconstruction as well as a comparative study. Similar studies have been done in [5, 9].

The above presented Φ-approach has a disadvantage that at one location in the image, both regularizations are applied simultaneously, of course with different positive weights. Moreover, the function Φ in [3] is supposed to be of class $\mathcal{C}^2$ which means that Φ does not vary quickly from place to place, and that the weights for both regularization also vary smoothly. This can be a problem, when you want to distinguish clearly between the BV regularization and the H^2 regularization. We try to eliminate this drawback and we allow for discontinuous Φ. Further we do not formulate the problem in terms of Φ but we formulate it in terms of a domain decomposition.

We propose the following two-step procedure.

ALGORITHM 1.

Step 1 *Divide the domain into two non-overlapping regions in such a way that in the first region denoted by Ω_1, the solution is supposed to have discontinuities, and in the second region denoted by Ω_2, the solution is supposed to be smooth.*

Step 2 *Regularize the problem with two different regularizations. On Ω_1 apply the BV regularization that is capable to preserve the discontinuities, and on Ω_2 use the H^k regularization, which does not have the staircase effect.*

We will refer to this method as the mixed BV-H^k regularization. The implementation of the first step of the procedure will be discussed in Section 3. Some theoretical results concerning the second step will be presented in Section 2.

To motivate the study of this method, we present the results for the same benchmark picture Lenna presented before. In Figure 1.2a one can see the division of the whole domain into two parts, the black one, where the H^2 regularization has been used and the white one, where the BV regularization has been applied. Such a division will be referred to as a *mask*. The resulting image is depicted in Figure 1.2b. One can see the qualitatively better result when the smoothing and the edge-enhancing regularizations have been used at right locations.

2. Basic definitions and theoretical considerations. In this section we first briefly introduce the space $BV(\Omega)$. We do not present any details about the classical Sobolev spaces $H^k(\Omega)$, the overview can be found e.g. in [2]. Next we rigorously define the minimization problem appearing in the second step of the Algorithm 1 with mixed BV-H^k regularization and finally we sketch theoretical results concerning stability and well-posedness of the minimization problem.

Space of BV functions. Denote by $C_c^\infty(\Omega, \mathbb{R}^n)$ the set of functions in $C^\infty(\mathbb{R}^n)$ with compact support in Ω.

DEFINITION 2.1. *By $BV(\Omega)$ we denote the subspace of functions $u \in L^1(\Omega)$ such that the following quantity is finite*

$$J(u) = \sup \left\{ \int_\Omega u(x) \nabla \cdot (\xi(x)) \mathrm{d}x \mid \xi \in C_c^\infty(\Omega, \mathbb{R}^n), \|\xi\|_{L^\infty(\Omega, \mathbb{R}^n)} \leq 1 \right\}.$$

We define the norm $\|u\|_{BV} = \|u\|_{L^1} + J(u)$.

We point out that $BV(\Omega)$ endowed with the norm $\|u\|_{BV}$ is a Banach space and for the cases when $\Omega \subset \mathbb{R}^2$ we have that $BV(\Omega) \subset L^2(\Omega)$ [1]

From [1] we present two important results concerning weak compactness and lower semi-continuity of elements from the BV theory.

LEMMA 2.2. *The norm* $\|\cdot\|_{BV}$ *is weakly lower semi-continuous with respect to the* L^p *topology for* $1 \leq p < \infty$.

LEMMA 2.3. *Let* S *be a* BV*-bounded set of functions. Then* S *is relatively compact in* $L^p(\Omega)$ *for* $1 \leq p < n/(n-1)$. S *is bounded and thus relatively weakly compact in* $L^p(\Omega)$ *for* $p = n/(n-1)$ *and* $n \geq 2$.

These two lemmas are key tools in the study of well-posedness and stability of underlying minimization problems.

Minimization problem, well-posedness, stability. Consider a domain Ω that is split into two non-overlapping subdomains Ω_1, Ω_2 so that $\Omega = \Omega_1 \cup \Omega_2$. The first subdomain should cover the discontinuities of the solution, the second subdomain represents the part, where our solution is expected to be smooth. At this point we do not exploit how this domain decomposition can be achieved. This topic is answered in Section 3.

Further take $U_1 = BV(\Omega_1), U_2 = H^k(\Omega_2)$ where k is the desired order of expected regularity of the solution. We construct U by

$$U = \{u \in L^2(\Omega) : u|_{\Omega_1} \in U_1, u|_{\Omega_2} \in U_2\}. \tag{2.1}$$

We introduce two projectors being the restrictions on Ω_1 and on Ω_2 defined by $P_1 : U \to U_1,\ P_1 u = u|_{\Omega_1}$ and $P_2 : U \to U_2,\ P_2 u = u|_{\Omega_2}$. Using the above notations we can write that

$$U = U_1 \oplus U_2.$$

Next to the strong topologies on U_1 and U_2 generated by the norms, we introduce weaker topologies. By τ_1 we thus denote the topology generated by the weak norm on $L^2(\Omega_1)$ and by τ_2 we denote the topology generated by the weak norm on $H^k(\Omega_2)$. Further we denote by τ_U the topology that is naturally generated by τ_1 and τ_2 on the direct sum U, so that a sequence $u_k \to u$ with respect to τ_U, when $P_i u_k \to P_i u$ with respect to τ_{U_i} topology for $i = 1, 2$.

We rigorously define our minimization problem for mixed BV-H^k regularization. Consider regularized least-square solution defined by

$$\min_{u \in U} \{\|u - f\|_2^2 + \alpha_1 \|P_1 u\|_{BV(\Omega_1)} + \alpha_2 \|P_2 u\|_{H^k(\Omega_2)}\}. \tag{2.2}$$

In other words we seek a function u that fits the data f and moreover its restriction on Ω_1 belongs to $BV(\Omega_1)$ and its restriction on Ω_2 belongs to $H^k(\Omega_2)$.

Notice that we do not impose any continuity properties of the solution on the boundary between Ω_1 and Ω_2. This is also not necessary, since the BV regularization allows for the discontinuities. So if there was such a continuity assumption on the boundary, the solution would be anyway allowed to be discontinuous in the very close neighbourhood of this boundary.

The case when $\Omega_2 = \emptyset$ was studied by Acar in [1]. He studied more general setting when the fidelity term was extended by a linear operator to $\|Au - f\|_2^2$. It is possible to extend his results to our case however we use the framework introduced by

Hofmann et al. in [8]. They study a minimization problem in the setting of general Banach spaces

$$\min_{u\in\mathcal{D}}\{\|F(u)-f\|_V^r+\alpha\mathcal{R}(u)\} \tag{2.3}$$

where $F:\mathcal{D}(F)\subseteq U\to V$ is the (in general nonlinear) forward operator mapping between Banach spaces U and V and where they have $1\leq r<\infty$ for the exponent in (2.3). $\mathcal{R}$ is a convex and proper stabilizing functional with the domain

$$\mathcal{D}(R):=\{u\in U:R(u)\neq+\infty\},$$

and $\mathcal{D}=\mathcal{D}(F)\cap\mathcal{D}(\mathcal{R})$.

Their results can be extended for the case when $\mathcal{R}$ consists of a sum of several stabilizing functionals $\mathcal{R}_i$ and the Banach space U can be decomposed into a direct sum of their domains. The detailed description of this extension will be published elsewhere and frequently uses Lemmas 2.2 and 2.3. The consequence of such an extension is that the theoretical results concerning the well-posedness and stability are valid also for the case of several stabilizing functionals.

In our case, $F=I,\ r=2,\ U=V=L^2(\Omega),\mathcal{D}=U_1\oplus U_2$ and

$$\mathcal{R}(u)=\alpha_1\|P_1u\|_{BV(\Omega_1)}+\alpha_2\|P_2u\|_{H^k(\Omega_2)}.$$

Theoretical results from [8] adapted for our setting are formulated in the following theorem.

THEOREM 2.4.

(well-posedness) Assume that α_1 and α_2 are positive and that $f\in L^2(\Omega)$. Then there exists a minimizer of (2.2).

(stability) The minimizers of (2.2) are stable with respect to the data f. That means that if $\{f_k\}$ is a sequence converging to f in $L^2(\Omega)$ with respect to the norm topology, then every sequence $\{u_k\}$ satisfying

$$u_k=\arg\min_{u\in U}\{\|u-f_k\|_2^2+\alpha_1\|P_1u\|_{BV(\Omega_1)}+\alpha_2\|P_2u\|_{H^k(\Omega_2)}\} \tag{2.4}$$

has a subsequence, which converges with respect to the τ_U topology, and the limit of each τ_U convergent subsequence is a minimizer of (2.3).

3. Numerical implementation. The implementation of the second step from Algorithm 1 brings no novelty or difficulty. Once we have obtained well-posedness and stability, we can work with gradient based methods to obtain the minimum of the energy functional. More precisely, we apply the conjugate gradients (CG) method.

In the H^2 part of the minimization, the computation of the gradient of E is straightforward since the energy functional is quadratic. The BV regularization term however is not differentiable. We therefore replace the $|\nabla u|$ term with its differentiable approximation $(|\nabla u|^2+\epsilon)^{1/2}$ for some small positive ϵ. There are also methods where no such regularization is necessary for example the method of Chambolle [6].

More interesting task is the implementation of the first step from Algorithm 1. We need to find out how to choose the mask. The mask is represented by an array of the same size as the noisy image has. This array contains only zeros and ones. For the purpose we described in the first section, we need to detect the edges in the image. Image detection approaches include among others the first order methods when the edge is located at the places where the norm of the first derivatives has a peak which

exceeds a given threshold. Second order methods detect the edges as the zero points of the second derivatives.

We use the first approach, however we first need to pre-process the image. We suppose that the image is noisy and therefore there is no way to use first order methods, since the noisy image can be highly oscillating.

We therefore apply the BV regularization to the whole image since we know that this kind of regularization preserves the edges in the picture. For the purposes of the edge detection it is not a disadvantage when the reconstructed image contains flat regions where smooth solution is expected.

From this first iteration we can extract the edges and define the mask and afterwards we can continue and use the mixed BV-H^2 regularization.

We can rewrite the Algorithm 1 in more details.

ALGORITHM 2. *Start from the noisy data f.*

Step 1 *(a)* **On the noisy data f apply the BV regularization.** *To do this, set $\Omega_2 = \emptyset$. Then replace the non-differentiable term $|\nabla u|$ with its differentiable approximation $(|\nabla u|^2+\epsilon)^{1/2}$. Afterwards, using the CG algorithm, obtain the solution of the minimization problem (2.2) denoted by u_{BV}.*

(b) **Get the mask u_m.** *On the (already regularized) u_{BV} apply the first order method to determine the edges and the mask.*

* *Start with $\Omega_1 = \emptyset$*
* *For every point x decide if gradient exceeds the threshold $\mathrm{t_{tr}}$. If it is so than add this point to Ω_1.*
* *To obtain the mask, put also all neighbouring points of x to Ω_1, reason for this is described below.*

Step 2 *(a)* **Construct the energy functional.** *According to the mask u_m, split the domain Ω so we have $\Omega_2 \neq \emptyset$.*

(b) **Get the BV-H^2 regularized solution u_{mix}.** *Using the CG algorithm minimize the energy functional (2.2) and obtain u_{mix}.*

One can repeat the steps 1b, 2a, 2b several times so that the mask in step 1b will be obtained from u_{mix} instead of from u_m. In such a way, better results can be obtained.

From 1D experiments we have the following observation:

> *Once the location of an edge falls out of the BV region in the mask, the edge is lost.*

This observation is somehow clear. Imagine that the location of an edge is not detected correctly and falls into the region where the H^2 regularization is used. That means that this edge is immediately blurred and thus lost. Therefore it is necessary to determine the mask carefully such that all edges are included. Following this recommendation we extended Step 1b and Ω_1 includes not only the points on the edges but also their neighbouring points.

To determine the edges correctly we need to set the right threshold value $\mathrm{t_{tr}}$ in Step 1b. This can be done iteratively. We begin with some low value and thus the resulting mask will show same patterns as the noise does. Quantitatively, the mask and noise will have the same variance and mean. Increasing this threshold we eventually come to the value when the variances of the mask and the noise will not coincide any more. In other words, we need to find the value of the threshold, when the mask will not be a random picture any more.

In the energy functional, two weighting factors occur. To determine the correct values of α_1 and α_2, one can apply the generalized L-curve method [4]. Having these determined, one can apply the results from Theorem 2.4 to obtain the existence of the minimizers of (2.2) for specific values of α_1, α_2.

4. Experiments. We present two case-studies. One with an artificial image and one with a real-world image. The value of the constant ϵ was chosen to be 10^{-6}.

It is crucial to detect the edges correctly. In Figures 3.1e and 3.2e we see the resulting masks for two different noise levels of 5% and 20%. In the first case, the noise was low enough to determine all edges correctly. We started from low value of the threshold $\mathrm{t_{tr}}$ which gave noisy mask. Increasing the value of $\mathrm{t_{tr}}$ the mask became clean and with clear contours of edges. In the case with 20% noise, increasing $\mathrm{t_{tr}}$ does not significantly improved the mask. Accordingly, the mixed BV-H^2 reconstruction for 20% noise still keeps the undesired artifacts. In Figures 3.1 and 3.2 we see the influence of the noise on the resulting image.

We were not satisfied with the result depicted in Figure 3.2. The reason for such a bad reconstruction was in the bad detection of the edges. Edges were detected from the BV regularized image and clearly, the BV regularization can not handle such high noise. So we tried first to annihilate the noise by the H^2 regularization before it was processed by BV. As an input for the BV regularization in Step 1a of Algorithm 2 we took the H^2 regularized image instead of the noisy data f. Result can be seen in Figure 3.3a where the mask is clearly better than in Figure 3.2e. Similarly, the resulting image in Figure 3.3b is of higher quality.

Next case-study in Figure 4.1 shows the reconstruction of a noisy real image. We again see that the mask represents the edges in the image well and also the mixed BV-H^2 reconstruction is better than single BV or H^2 reconstruction. To see the differences clearly, please check the electronic version of this paper because of the possibly poor quality of the printed reproduction.

REFERENCES

[1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10(6):1217–1229, 1994.

[2] Robert A. Adams and John J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[3] G. Aubert and L. Vese. A variational method in image recovery. *SIAM J. Numer. Anal.*, 34(5):1948–1979, 1997.

[4] M. Belge, M.E. Kilmer, and E.L. Miller. Efficient determination of multiple regularization parameters in a generalized L-curve framework. *Inverse Problems*, 18(4):1161–1183, 2002.

[5] P. Blomgren, T. F. Chan, and P. Mulet. Extensions to total variation denoising. In F. T. Luk, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 3162 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 367–375, October 1997.

[6] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004. Special issue on mathematics and image analysis.

[7] P. Charbonnier. *Reconstruction d'image: Regularisation avec prise en compte des discontinuitis*. PhD thesis, University of Nice-Sophia Antipolis, 1994.

[8] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[9] K. Ito and K. Kunisch. Bv-type regularization methods for convoluted objects with edge, flat and grey scales. *Inverse Problems*, 16:909–928, 2000.

[10] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.

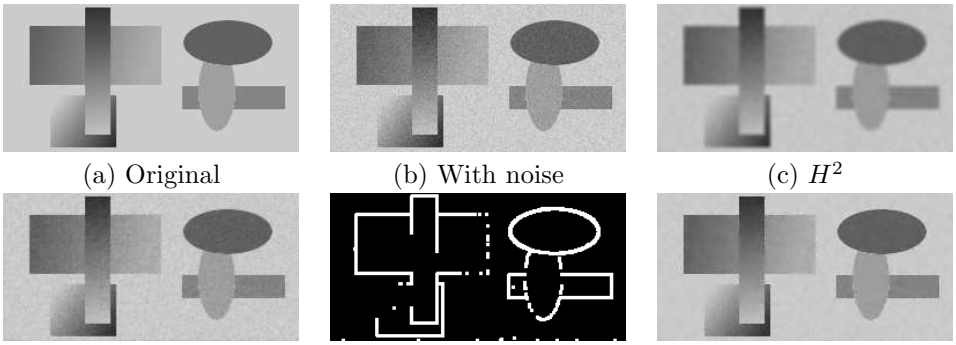
(a) Original (b) With noise

(c) H^2 (d) BV

FIG. 1.1. *Image reconstruction of the classical benchmark picture Lenna. (a) The original picture. (b) Added 5 % white noise. (c) Reconstruction using the H^2 regularization. The L^2 difference of the normalized H^2 reconstruction from the normalized original is 0.03954. (d) Reconstruction using the BV regularization. The L^2 difference of the normalized BV reconstruction from the normalized original is 0.02415.*

(a) Mask (b) Mixed

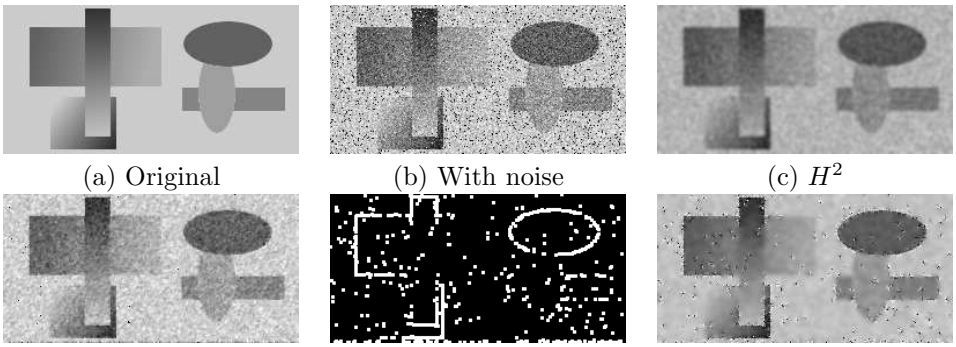
FIG. 1.2. *Image reconstruction with the mixed BV-H^2 regularization. (a) The decomposition of the domain into two regions: black part corresponds to smooth domains, white color indicates the regions with high jumps. (b) Reconstructed image. The L^2 difference of the normalized reconstruction from the normalized original is 0.02231.*



(a) Original (b) With noise (c) H^2

(d) BV (e) Mask (f) Mixed

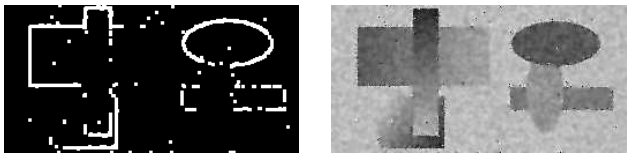
FIG. 3.1. *Artificial test image with 5% noise. The detection of the edges was satisfactory and the resulting BV-H^2 reconstruction is good. The L^2 difference of the normalized H^2 reconstruction from the normalized original is 0.03707. For the BV reconstruction it is 0.01667 and for the mixed reconstruction it reduces to 0.01052.*



(a) Original (b) With noise (c) H^2

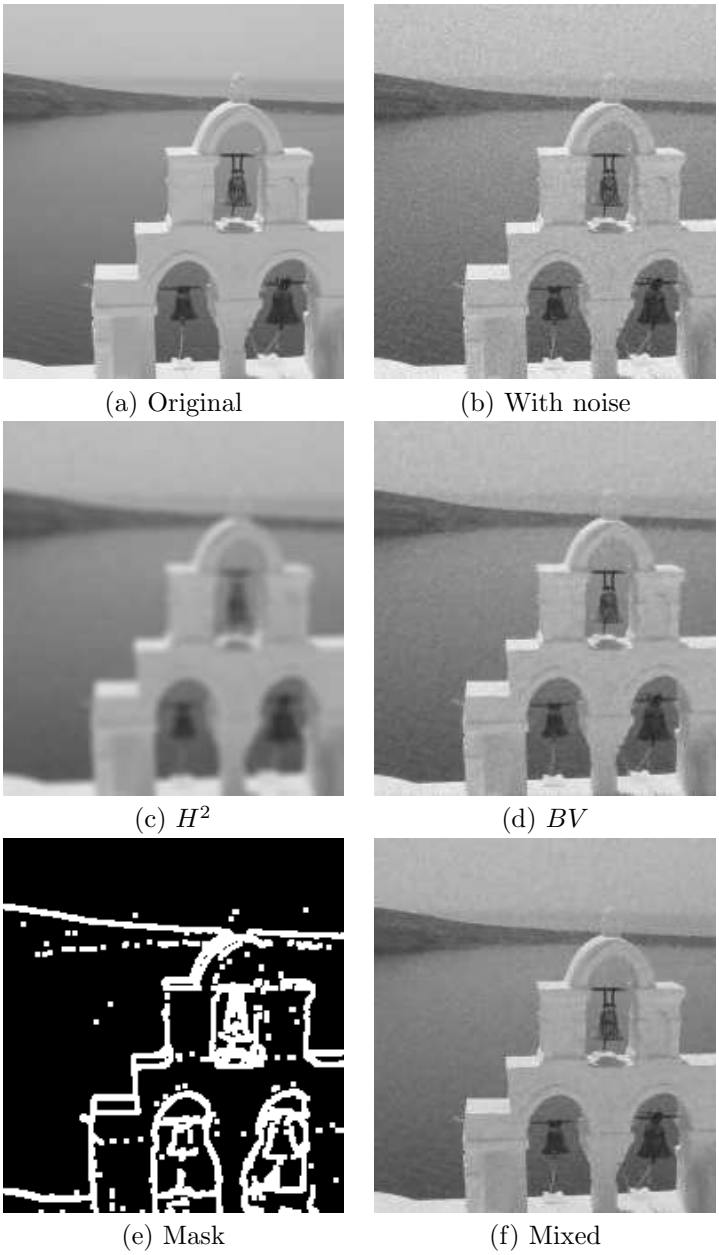
(d) BV (e) Mask (f) Mixed

FIG. 3.2. *Artificial test image with a high level of noise of 20%. The reconstruction already suffers from the bad detection of the edges. The L^2 difference of the normalized H^2 reconstruction from the normalized original is 0.10318. For the BV reconstruction it is 0.04306 and for the mixed reconstruction it reduces to 0.02612*



(a) Mask (b) Mixed

FIG. 3.3. *Artificial test image with a high level of noise of 20%. Noisy data f was first processed by the H^2 regularization resulting in partly annihilation of the noise. Afterwards the BV regularization has reconstructed the edges and from the result, the mask has been obtained. The L^2 difference of this reconstruction from the normalized original is 0.02153.*



(a) Original (b) With noise

(c) H^2 (d) BV

(e) Mask (f) Mixed

FIG. 4.1. *Real picture taken in Greece. Level of the noise is 3%. The L^2 difference of the normalized H^2 reconstruction from the normalized original is 0.05521. For the BV reconstruction it is 0.02713 and for the mixed reconstruction it reduces to 0.01719.*