Denoising of Digital Radiographic Images with Automatic Regularization Based on Total Variation

Mirko Lucchese and N. Alberto Borghese

Applied Intelligent Systems Laboratory (AIS-Lab), Department of Computer Science, University of Milano, Via Comelico 39, 20135 Milano, Italy
{mirko.lucchese,alberto.borghese}@unimi.it

Abstract. We report here a principled method for setting the regularization parameter in total variation filtering, that is based on the analysis of the distribution of the gray levels on the noisy image. We also report the results of an experimental investigation of the application of this framework to very low photon count digital radiography that shows the effectiveness of the method in denoising such images. Total variation regularization leads to a non-linear optimization problem that is solved here with a new generation adaptive first order method. Results suggest a further investigation of both the convergence criteria and/or the scheduling of the optimization parameters of this method.

Keywords: Digital radiography, total variation filtering, regularization, Bayesian filtering, gradient descent minimization.

1 Introduction

Radiographic images are produced by converting the number of X-ray photons that hit the sensor inside the area of each pixel, into a gray level. Thanks to the sensitivity of modern detectors, radiation dose is getting lower and lower: digital panoramic images are produced by a maximum photon count around 10,000 on 14 bits: an almost one to one correspondence between the number of photons and the gray levels. This greatly increases the resolution of the imaging system, but it requires also a careful noise elimination to make the images most readable to the clinicians.

In radiographic images noise is mainly introduced by errors in photon ount and it is therefore modeled through a Poisson distribution [1, 2]. Traditional denoising approaches like linear filtering are inadequate to remove this kind of noise, as they generally tend to smooth significantly the edges in the image and different approaches have been investigated. Among these, regularization theory is particularly suitable as it clearly expresses the two goals of a denoising process: obtain a function as close as possible to the measured data and penalize solutions that have undesirable properties. In this framework a cost function is written as a weighted sum of a distance between the measured and the true data and a function that penalizes "bad" solutions [3]. In the original formulation the squared function was used to express the difference between the true and the measured data and the squared of the gradient was used as a penalization term. This formulation leads to a squared cost function that is convex and calls

for an easy minimization. However, it was soon recognized that the weighting coefficient, called regularization parameter, was critical for the quality of the result.

More recently, a clear connection between regularization theory and Bayesian estimate has been described [4]. In the Bayesian framework, the distance between the true and the measured image is evaluated through the negative logarithm of the likelihood of the measured image given the true one, and the penalizing term is associated to a suitable prior distribution of the solutions, often in terms of an Gibbs distributions. The filtered image, $\mathbf{g_f}$, is therefore obtained as:

$$\begin{cases} \mathbf{g_f} = \arg\min_{\mathbf{g}}\left[J\left(\mathbf{g},\mathbf{g_n}\right)\right] \\ J\left(\mathbf{g},\mathbf{g_n}\right) = L\left(\mathbf{g},\mathbf{g_n}\right) + \lambda \cdot R\left(\mathbf{g}\right) \end{cases} \tag{1}$$

where, $i=1\ldots N$ are the pixels in the image, $\mathbf{g} = \{g_i\}_{i=1..N}$ is the unknown, noiseless (true) radiographic image and $\mathbf{g_n} = \{g_{n,i}\}_{i=1..N}$ is the measured, noisy image. R(.) is the regularization term, function of $\mathbf{g}$ (often of its derivatives). λ is the regularization parameter and balances closeness to the data and reliability of prior distribution.

The statistical view allows incorporating an adequate noise model and a principled regularizer inside the cost function J(.). In particular, in X-ray radiography, Poisson noise is considered inside the first term. This leads to the following L(.) function:

$$L(\mathbf{g},\mathbf{g_n}) = \sum_{i=1}^{N}\left[-g_i + g_{n,i}\ln(g_i) - \ln\left(g_{n,i}!\right)\right] \tag{2}$$

This expression contains a factorial term, difficult to compute; therefore, (2) is often substituted by the ratio between $L(\mathbf{g},\mathbf{g_n})$ and $L(\mathbf{g_n},\mathbf{g_n})$, that is also called normalized Kullback-Liebler divergence:

$$KL\left(\mathbf{g},\mathbf{g_n}\right) = \frac{L\left(\mathbf{g},\mathbf{g_n}\right)}{L\left(\mathbf{g_n},\mathbf{g_n}\right)} = -\sum_{i=1}^{N}\left\{g_{n,i}\cdot\ln\frac{g_i}{g_{n,i}} + g_{n,i} - g_i\right\} \tag{3}$$

The expression of the penalization term, as the squared sum of the gradients, has been also questioned as it tends to over smooth the images. For this reason regularizers that increase less smoothly with the edge amplitude have been proposed. One of the most used is based on a total variation measure, defined as:

$$R(\mathbf{g}) = \sum_{i=1}^{N}\left\|\nabla g_i\right\| \tag{4}$$

where ∇g_i is the gradient of $\mathbf{g}$ computed in position i and $\|.\|$ indicates its l_1 or l_2 norm. This approach has become recently popular also in the image processing field [5].

J(.) in (1) becomes strongly non linear with respect to $\mathbf{g}$, and iterative algorithms have been adopted for its minimization. Very recently improved first order methods have been proposed, and in particular the Scaled Gradient Projection Method [6] has shown superior convergence speed and it has been adopted here.

The most critical issue of a regularization approach is an adequate setting of the regularization parameter, λ. One possibility is to use the Mozorov discrepancy principle [7] that states that: the regularization parameter should be increased as much as possible until the difference between the filtered and original data match the noise covariance. This principle can be adopted in a straightforward way when noise is supposed Gaussian, as the covariance matrix, in this case, does not depend of the gray levels distribution. The adaptation of the discrepancy principle to the Poisson case is more difficult. Only very recently a first attempt in this direction has been recently proposed in [6], with an approximation of the likelihood function. We derive here a different method to set λ that is based on the properties of the Poisson distribution and does not require any approximation. We show that it behaves very similarly to [6] for most photon counts, while it is more accurate when photon count is very low.

2 Method

We have generated a set of simulated radiographic images of 512 x 512 pixels, with 16 bpp. First, an absorption coefficients map is created, with coefficients increasing from 0% for the left-most pixels to 100% for the right-most ones. Then, 50 different geometrical figures (circles and rectangles) were randomly positioned inside the map, and added to the background previously created. The circles have a random radius between 1 and 512 pixels; the rectangles sides have a random length between 1 and 512 pixels. Each time a circle or a rectangle is added to the map, all the absorption coefficients covered by the figure are modified as follows: either they are substituted by their complements with respect to 100%, or they are multiplied by a random value between 0 and 1, or a random value between 0% and 100% is added to them. In the latter case, the resulting absorption coefficients are always clipped to 100%. The choice among the three modalities is random.

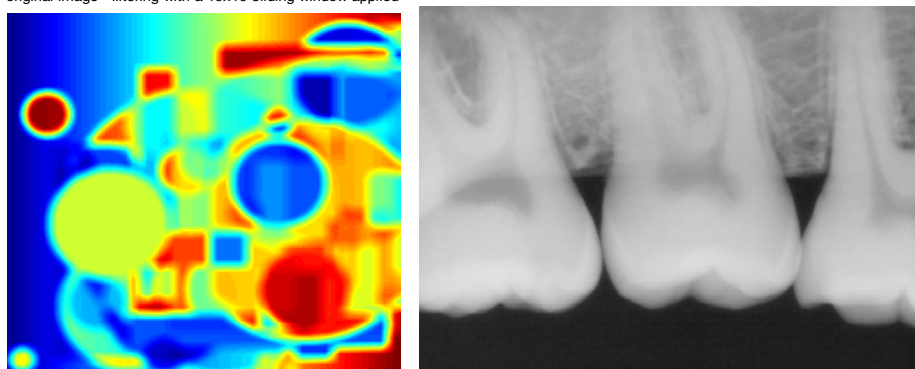original image - filtering with a 19x19 sliding window applied

Fig. 1. A typical synthetic digital image used for testing is shown in panel (a). A typical intra-oral image on which total variation filtering was applied is shown in panel (b). It has dimension 1000 x 1387 on 14 bits. The image is shown applying a γ-correction with an exponent equal to 1.5 as usual in clinical practice.

The image is then filtered with a moving average (MA) filter, whose window size was set to 19 x 19 to obtain frequency content similar to that of X-ray digital images and panoramic images in particular. A maximum number of photons equal to 15,000 is considered here. A typical image of this type is shown in Fig. 1a (cf. [2]). A sensor gain equal to one and an offset equal to zero were adopted to guarantee that the gray levels coincide with photon counts. Noise is then added independently for each pixel, according to Poisson statistics (Fig. 1a). Sensor parameters and maximum number of photon counts were chosen such that sensor does not saturate.

These images were then filtered minimizing the cost function (1) through the Scaled Projected Gradient method introduced in [6]. This method adapts at each step the update amplitude analyzing the cost function value in two consecutive steps, applying the Barzilai and Borwein adaptation rule. Moreover, at each step, the cost function is compared to that of the previous step and the update can be decreased in order to keep the cost function decreasing (Armijo rule, [8]).

The following parameters were used according to [6]: the maximum absolute value of the scaling factor is set to $L = 1000$, the range of admissible step sizes, α_{min}, α_{Max} to: $[10^{-3} \ 10^{3}]$, the number of consecutive values of α stored is $M_\alpha = 10$, the rate of updating used to evaluate if the scale size has to be reduced is $\gamma = 10^{-4}$ and the reduction factor of the step size introduced by Armijo rule is $\theta = 0.4$. The α value used in the first iterate was set to $\alpha_0 = 1.3$. The gradient in (4) was computed as the difference between the gray level of the pixel and that of the north and east pixels. To avoid possible singularity in the derivative of the gradient a value of small quantity is added to the absolute value of the gradient in (4), according to [4]. This small quantity has been chosen equal to one here.
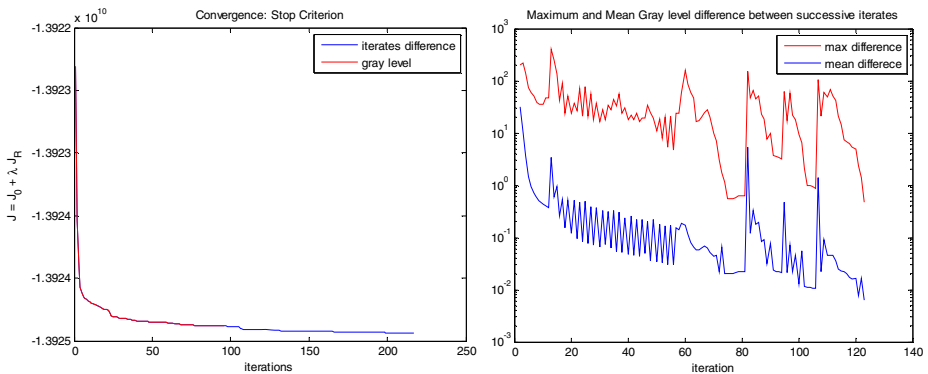

Fig. 2. The cost function (1) is plotted as iterations progresses on the left. The iterations in which the maximum difference of gray levels between two consecutive iterations is above half a gray level are plotted in red. On the right, the mean and maximum variation of gray levels between two iterates.

The following convergence criterion is adopted:

$$\frac{\left|J_k - J_{k-1}\right|}{J_k} < \tau \tag{5}$$

with k number of iteration. $\tau = 10^{-9}$ was set: we verified that the maximum variation of the grey levels when the optimization algorithm was stopped was never larger than half a gray level.

As shown in Fig. 2, the cost function (1) decreases smoothly as the iterations progresses. In particular condition (5) is met at 140 iterations, while a maximum increase between two consecutive iterates below half a gray level is achieved already after 62 iterations.

We repeated the minimization criterion for different values of the regularization parameter, λ, for this image and found that for $\lambda = 0.01$ the distance, measured through the normalized KL, between the filtered and the true image, KL(g, g_{true}), was the smallest as shown in Figs. 3. Different values of λ are found for different images although the values are generally quite similar.
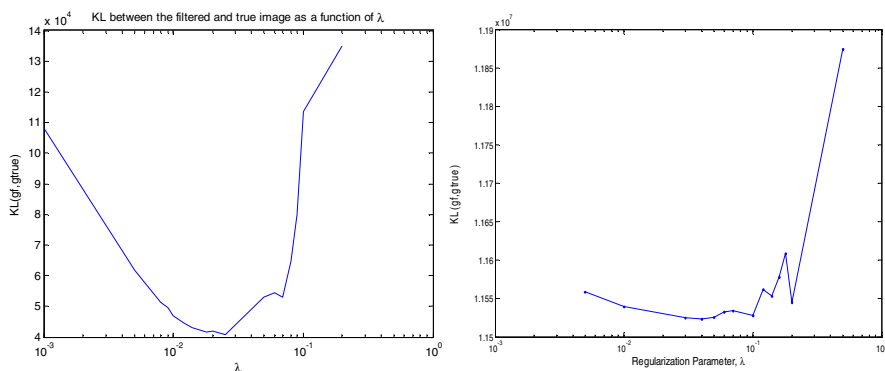

Fig. 3. KL(g,g_{true}) as a function of the regularization parameter λ for two different images of the data set. The minimum is achieved for $\lambda = 0.025$ in the first image, for $\lambda = 0.04$ in the second one. The stopping condition of Equation (5) is adopted here.

However, in real life, we usually do not have g_{true} (otherwise, there would be no need of filtering!) and the need of a suitable criterion to compute the optimal value of λ arises. Very recently, [6], an approach based on the Mozorov principle has been proposed: the KL in (1) is linearized around the true value g, obtaining the following Taylor expansion:

$$KL(g_n, g) = \frac{(g_n - g)^2}{2g} - \frac{(g_n - g)^3}{6g^2} + \frac{(g_n - g)^4}{12g^3} - \frac{(g_n - g)^5}{2g^4} + \ldots \tag{6}$$

The expected value of (6) can be computed from the gray value of the pixels in the image. Taking into account that the variance and the curtosis of a Poisson distribution are equal to the mean value [9], we obtain:

$$E\{KL(g_n,g)\}=\frac{1}{2}\left\{\frac{\sigma^2}{g}+\frac{1}{6g}-\frac{29}{6g^2}+....\right\} \tag{7}$$

The approximation gets better and better as photon count, g, increases, therefore, in [6], the expansion in (5) arrested at the first order derivative is considered:

$$E\{KL(g_n,g)\}=\frac{1}{2}\left\{\frac{\sigma^2}{g}+O\left(\frac{1}{g}\right)\right\} \tag{8}$$

Considering that in the real case, we have a finite number of pixels with the same gray level, but that this number can be sufficiently large to get a reliable estimate of KL(.), the following discrepancy principle has been proposed [6]:

$$D(g_n,g)=\frac{2}{n*m}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(g_{n_{i,j}}\ \ln\left(\frac{g_{n_{i,j}}}{g_{i,j}}\right)+g_{i,j}-g_{n_{i,j}}\right) \tag{9}$$

This can be viewed as twice the average value of the normalized KL, averaged over all the pixels. The true value of D(.) is larger than 1 because of the term O(1/g) in (8), and it approaches 1 with the increase in the photons number as shown in Fig. 4. For the images containing very few photon, the possibility of measuring $g_n(.) = 0$ can be meaningful, in this case we assume that $g_n\log(g_n/g) = 0$ for $g_n = 0$ and therefore set the value of the generalized KL to 0 for those pixels for which the noisy image contains zero. As can be seen in Fig. 4, D(.) becomes very little reliable for very low photon counts.
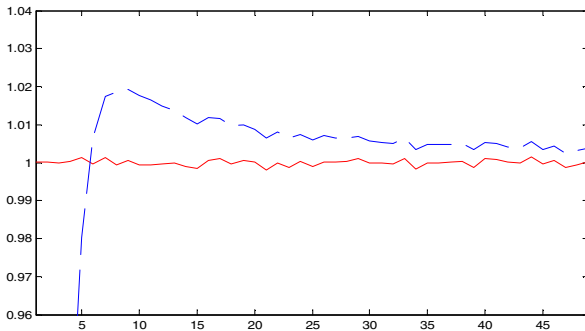

Fig. 4. The value of the discrepancy function in (9) as a function of the pixel counts is plotted in dashed line. Data are obtained from empty simulated digital images for which the same number of photons arrives in all the pixels. Results are averaged over 40 different images and are a good approximation of the expected value of D(.) for photon counts in the range between 0 and 100. The value of the discrepancy function D'(.) in (11) as a function of photon counts is plotted as a continuous line and it is almost constant and equal to 1.

We observe that D(.) derived in (7) is based on an approximation (Eq. (6)). A different discrepancy function is derived here that does not require any approximation. Let us consider the properties of the Poisson distribution, namely that the variance is equal to the expected value. From this observation, for each gray level, we can derive the simple relationship:

$$\frac{\sigma_{g_{true}}^{2}}{g_{true}} = 1 \tag{10}$$

which expresses the well known property that Poisson noise variance increases with the gray levels. We can specialize this property to the case in which we have many, although in finite number, of pixels for each gray level: we can write (10) as:

$$D'(g_n, g_f) := \frac{1}{n*m} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\left(g_{n,i} - g_{f,i}\right)^2}{g_{f,i}} = 1 \tag{11}$$

The time course of D'(.) is very similar to that of D(.) in the examined images and the value computed for λ, in most of the gray levels range, is almost the same. Eq. (11) may produce a different, more appropriate, value of λ when images have a very low photon count as shown in Fig. 4.

We also notice that D'(.) is coincident with the second order approximation of the normalized KL. Therefore it seems that KL is equal to the normalized variance plus other terms: these additional terms constitute the difference between D(.) and D'(.) and tend to vanish increasing the photon count (cf. Equation (8)).

3   Results

Equation (11) can be used to automatically set the value of λ in (1) as follows: starting from a reasonable low value, λ is increased until the function D'(.) in (11), computed between the filtered and the noisy image, crosses the value of 1 ([6] and cf. Figs. 5).
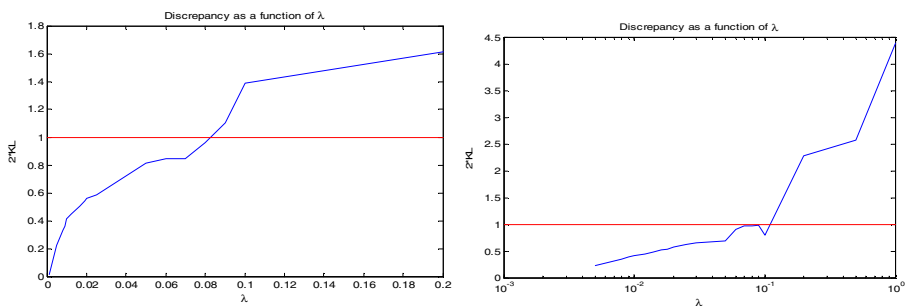

Fig. 5. The value of the function D(.) in (9) is reported as a function of λ for the same image of the leftmost panel of Fig. 3. Notice that the curve intersects the axis D(.) = 1 λ = 0.09, larger than the optimal one defined in Fig. 3.

As it can be seen the curve is quite smooth but the value of λ is slightly overestimated being set to $\lambda = 0.09$ versus the value of $\lambda = 0.025$ that results optimal from Fig. 3.

We have first investigated the sensitive of λ to the stopping conditions. If we stop the iterations when the maximum increment of the gray levels falls below half a gray level, KL(g_f, g) follows the curve reported in Fig. 6. This curve suggest a value $\lambda = 0.2$, and it is quite different from that reported in Fig. 3, where a maximum
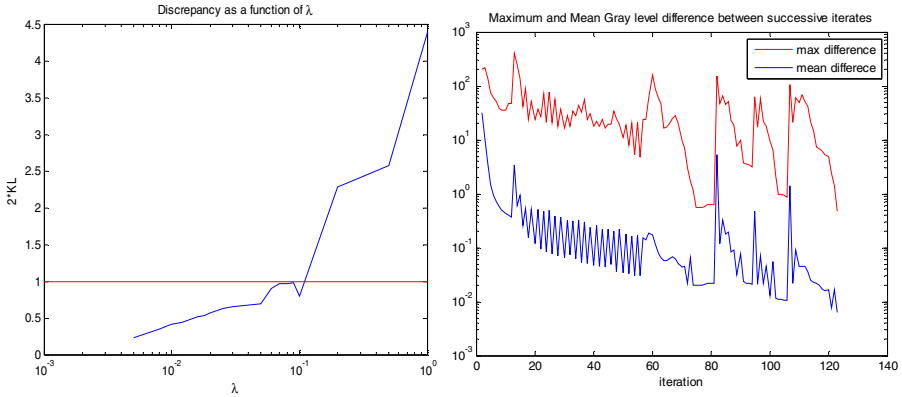
Fig. 6. The KL($\mathbf{g_f}$, $\mathbf{g_{true}}$) as a function of λ for the same image in the upper panel of Fig. 3. Convergence was stopped when the maximum increment of the gray levels between two iterates falls below half a gray level. Notice the different shape of the curve, the minimum is achieved for $\lambda = 0.05$, about twice that found in Fig. 3.

normalized variation in the cost function was considered as a stopping criterion. When this second condition was adopted as stopping criterion, the value of $\lambda = 0,025$ was suggested, almost 8 times smaller.

Notice that, in Fig. 6, the behavior of D'(.) is not monotone and the risk of a double crossing of the horizontal line D'(.)=1 does exist. This would make the choice of λ not well behaving.

We have further analyzed the behavior of the minimization algorithm as the number of iterations progresses: we have plotted the mean and maximum variation of gray levels between two successive iterates in Fig. 2. Data are referred to the image in the bottom panel of Fig. 4. As it can be appreciated, although the cost function is monotonically decreasing (Fig. 2), there are large variations in the gray levels, of up to 100 gray levels, also close to convergence. This regards only a very few pixels and therefore it does not change significantly the overall cost function. A similar behavior can be noticed also in the normalized variation of the cost function. These data may suggest a more in depth investigation of the convergence criterion and/or of the scheduling of the parameters in the optimization algorithm.

Although the computation of λ has to be further investigated, the filtering result is quite good as it can be appreciated in Fig. 7. In particular the filter is able to remove most of the Poisson noise from the image, leaving the edges almost unchanged (Fig. 7c). We have compared the behavior of such filter with a popular non-linear edge preserving filter, the bilateral filter, implemented according to [10]. Results with this filter are reported in Fig. 7b and in Fig. 7d; they show that this filter does reduce the edges amplitude significantly. The better capability of removing Poisson noise by total variation can be appreciated by computing the standard deviation in a homogeneous region like the one delimited by a dashed square in Figs. 7a and 7b. This is 3.37 for the image obtained through total variation filtering and 9.34 for bilateral filtering. This result indirectly confirms the suitability of the method proposed here to set the value of λ.
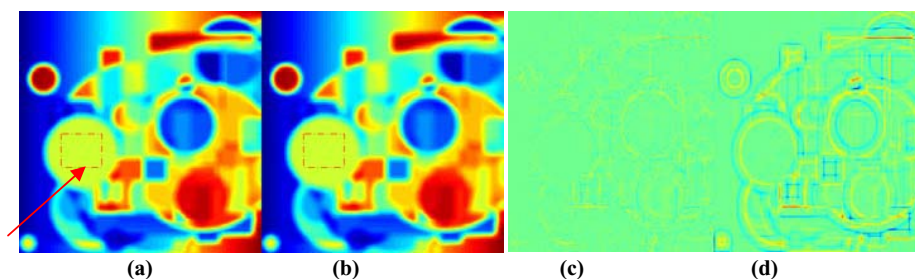
Fig. 7. The image in Fig. 1b, filtered through total variation is shown in panel (a) and filtered through the bilateral filter in panel (b), of 11x11 mask, a spatial standard deviation $\sigma_d = 3$, and an intensity standard deviation, $\sigma_r = 0,1$, that proved experimentally the best results. The difference between filtered images and the original noise-less image in Fig. 1a is shown in panels (c) and (d). Colors have been assigned by the colormap "jet" of Matlab software.
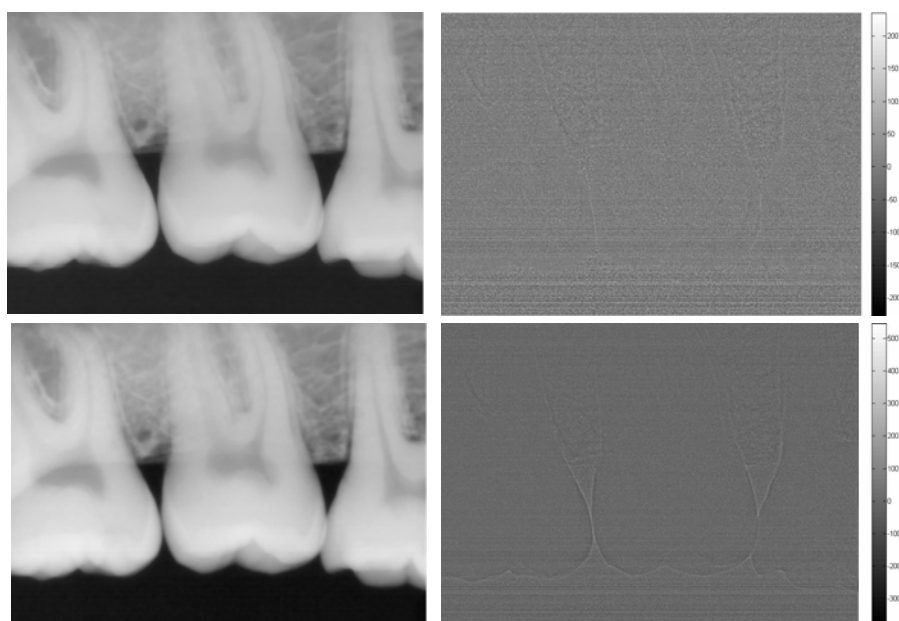

Fig. 8. In panel (a) the image in Fig. 10 filtered with total variation, the difference between the filtered and the original image is shown in panel (b), notice that noise removed is distributed homogeneously over the image. The same image in Fig. 1b, filtered with a bilateral filter, with the same parameters reported for Fig. 7. The difference between the image in panel (c) and the original image is reported in panel (d), notice a dependence of the filtered signal on the edge amplitude. Notice also that the difference in the gray levels in the image filtered with the bilateral filter is more than twice that of total variation filtering; this variation is concentrated in the edges.

Total variation filtering has been applied to real digital radiographic images, and in particular to intra-oral images like the one reported in Fig. 1b. The image has been filtered with both total variation and bilateral filtering and the results are reported in Fig. 8. As it can be appreciated, results are very similar to those obtained in simulated images and total variation achieves a better removal of Poisson noise minimally reducing the amplitude of the edges.

4   Conclusion

From these preliminary results, the ability of total variation filtering of removing Poisson noise from digital radiographs with low photon counts is evident. This confirms also the suitability of the method described here to compute the optimal value of the regularization parameter λ. The experimental results suggest a further investigation of both the convergence criteria and/or the scheduling of the optimization parameter inside the scaled gradient projection method considered here for optimization.

References

[1] Webb, S.: The Physics of Medical Imaging. Adam Hilger, Bristol (1988)
[2] Frosio, I., Borghese, N.A.: Statistical Based Impulsive Noise Removal in Digital Radiography. IEEE Transactions on Medical Imaging 28(1), 3–16 (2009)
[3] Tikhonov, A.N., Arsenin, V.A.: Solution of Ill-posed Problems. Winston & Sons, Washington (1977)
[4] Bertero, M., Lanteri, H., Zanni, L.: Iterative image reconstruction: A point of view. In: Proc. Interdisciplinary Workshop on Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation, Therapy (IMRT), Pisa, Italy (2007)
[5] Hirakawa, K., Meng, X.L., Wolfe, P.J.: A Framework for Wavelet-Based, Analysis and Processing of Color Filter Array Images With Applications to Denoising and Demosaicking. In: ICASSP (2007)
[6] Zanella, R., Boccacci, P., Zanni, L., Bertero, M.: Efficient gradient projection methods for edge-preserving removal of Poisson noise, Inverse Problems (in press)
[7] Engle, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer, Dordrecht (1996)
[8] Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena (1999)
[9] Johnson, N.L., Katz, S., Kemps, A.W.: Univariate Discrete Distributions. Wiley and Sons, New York (1993)
[10] Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth International Conference on Computer Vision, 1998, pp. 839–846 (1998)