SEMANTIC IMAGE CONTENT FILTERING VIA EDGE-PRESERVING SCALE-AWARE FILTER

Wei Ye and Kai-Kuang Ma

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
ye0003ei@e.ntu.edu.sg, ekkma@ntu.edu.sg

ABSTRACT

In this paper, we highlight a new filtering concept and methodology, called the *semantic image content filtering* (SICF), which aims to remove insignificant small details from the image while preserving its main structure. Such image *content* separation is not possible to achieve by using any conventional linear filter as it is essentially designed to perform *frequency* separation. To realize an effective SICF, a novel image filtering algorithm, called the *edge-preserving scale-aware filter* (ESF), is proposed in this paper. Our proposed ESF yields a significant improvement over a recently-developed scale-aware filter, called the *rolling guidance filter* (RGF). The key success of our ESF lies in the developed *adaptive relative total variation filter* (ARTVF), which replaces the RGF's Gaussian filter for generating a much improved *initial* guidance image. Extensive simulation results obtained from various test images have clearly demonstrated that the proposed ESF outperforms other state-of-the-art methods on conducting SICF task. That is, the semantically-important large-scale image structure has been better preserved, while the insignificant small details have been removed more effectively.

Index Terms— Semantic image content filtering, edge-preserving, scale-aware, rolling guidance filter.

1. INTRODUCTION

In many image processing applications, such as noise suppression, edge detection, segmentation, and so on, it is often highly desirable to separate *semantically* meaningful image contents (such as image's main structure) from insignificant ones (such as small details). In this paper, this task is coined as *semantic image content filtering* (SICF). Unfortunately, any conventional *linear shift-invariant* (LSI) filter (e.g., a Gaussian filter), which conducts filtering from the viewpoint of *frequency* separation, is unable to effectively accomplish this goal, since both main structure and small details can have both low and high frequencies.

Over the past decades, edge-aware or edge-preserving filters (e.g., [1]–[9]), such as the *bilateral filter* (BF) [2], the *guided filter* (GF) [4], and the *weighted least square filter* (WLS) [3], have emerged and can be exploited to conduct SICF. However, these filters judge the 'semantic importance' of image contents mainly based on the information of image *contrast*. As a result, they could fail to suppress insignificant small details that are in *high* contrast. Likewise, they could also fail to preserve those important main structures that are in *low* contrast. Recently, it has been pointed out in [10] that it is the *scale* information (rather than the *contrast*) that should be exploited for conducting SICF. Inspired by this insight, several works (e.g., [10]–[17]) have been proposed. Among them, the *rolling guidance filter* (RGF) [15] has been proven as a superior approach on conducting SICF task, compared with existing LSI filters and edge-preserving filters, as demonstrated in Fig. 1.

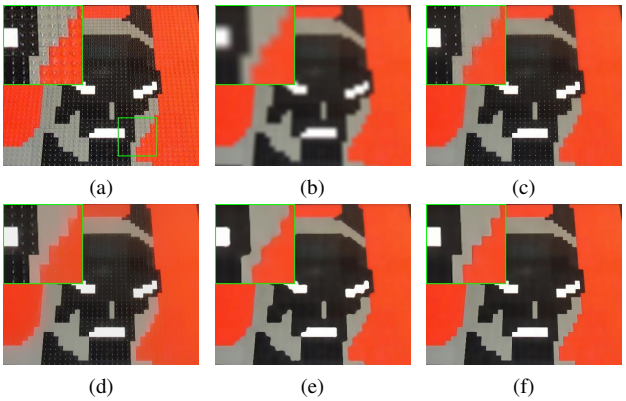


Fig. 1. Filtered results obtained from a test image "Batman" as shown in (a) by using: (b) the Gaussian filter, (c) the BF [2], (d) the WLS [3], (e) the RGF [15], and (f) the proposed ESF, respectively.

In this paper, a new SICF algorithm is proposed, called the *edge-preserving scale-aware filter* (ESF), which is a significantly improved version of the RGF, particularly on the aspect of edge-preserving capability (see Fig. 1). The key success of our approach lies in the developed *adaptive relative total variation filter* (ARTVF), which replaces the RGF's Gaussian filter for generating a much improved *initial* guidance image. Consequently, large-scale main structure can be more accurately preserved, while insignificant small-scale image details

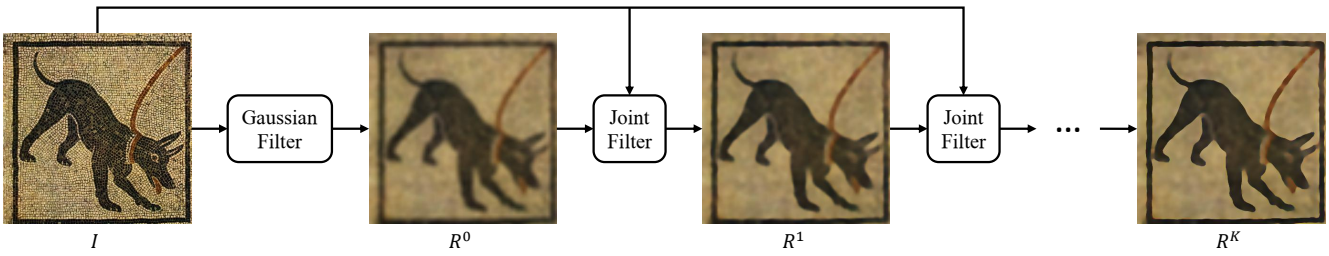


Fig. 2. Block diagram of the scale-aware RGF algorithm [15].

can be more effectively suppressed. Extensive simulations conducted on a variety of test images have clearly demonstrated that the proposed ESF is able to deliver superior SICF results, compared to that of existing state-of-the-art methods for performing SICF task.

The rest of this paper is organized as follows. Section 2 provides a succinct review of the RGF algorithm. Section 3 describes the proposed ESF method in detail. Section 4 conducts performance evaluations of our proposed ESF against several comparable filters. Section 5 concludes this paper.

2. ROLLING GUIDANCE FILTER (RGF)

The block diagram of the RGF [15] is demonstrated in Fig. 2. It is an iterative filtering process with two inputs and one output. One of the inputs is the original image I, and the other is a detail-removed version of I, denoted as R^k. The process aims to recover the main structure of the original image I under the latest-generated guidance of R^k in each iteration. Note that the filtered image R^k is an enhanced version of R^{k-1}, in the sense that the main structure presented in R^k is closer to that in the original image I. In [15], the initial detail-removed image R^0 is obtained by applying a Gaussian filter with the standard deviation σ to the input image I. As expected, the main structure of the image will be inevitably blurred as demonstrated in Fig. 2.

In [15], the *joint bilateral filter* (JBF) [18] is employed as the joint filter as depicted in Fig. 2, where the filtered value R_p^k at each pixel p in the k-th iteration is computed by

$$R_p^k = \frac{1}{J_p} \sum_{q \in \Omega_p} \exp\left(-\frac{||p-q||^2}{2\sigma_s^2} - \frac{||R_p^{k-1} - R_q^{k-1}||^2}{2\sigma_r^2} \right) I_q, \tag{1}$$

where Ω_p denotes a local image region centered at the pixel p, the pixel q denotes a neighboring pixel within Ω_p, and $J_p = \sum_{q \in \Omega_p} \exp\left(-\frac{||p-q||^2}{2\sigma_s^2} - \frac{||R_p^{k-1} - R_q^{k-1}||^2}{2\sigma_r^2} \right)$ is a normalization factor. One can see that the JBF's kernel is a product of two Gaussian kernels; i.e., a *spatial* kernel with the standard deviation σ_s and a *range* kernel with the standard deviation σ_r. The setting of these two parameters will be discussed in Section 4.

Since most the small-scale details have been effectively removed by the Gaussian filter, the resultant initial guidance image R^0 tends to be *flat* in the corresponding local regions. In this case, the JBF's range kernel will *not* take effect as $||R_p^0 - R_q^0|| \approx 0$. Consequently, the JBF becomes the same Gaussian filter that was exploited on the generation of R^0; in this case, $\sigma_s = \sigma$. Hence, the JBF is unable to recover those small-scale details that have been removed by the Gaussian filter. On the contrary, the range kernel *will* take effect on the main structure, since it has not been greatly destroyed by the Gaussian filter and thus possibly being recovered in R^K.

However, we have observed that the RGF is not capable on recovering those delicate parts of the main structure, such as sharp corners and thin edges. To demonstrate, one can refer to the R^K in Fig. 2 and observe that the dog's tail is not well recovered. As another example, the sharp corners appeared on the input image I in Fig. 1 are no longer sharp in the recovered R^K. This concern on the "edge-preserving" aspect will be addressed in our proposed method, as described in the following section.

3. PROPOSED EDGE-PRESERVING SCALE-AWARE FILTER (ESF)

In this paper, a new SICF algorithm is proposed, called the *edge-preserving scale-aware filter* (ESF), which is a much improved version of the RGF [15]. The improvement is achieved on the stage of generating the initial guidance image R^0, which is considered not trivial but of great importance on such guidance-based image filtering approach (Fig. 2), since the generated R^0 is used as one of two starting inputs for conducting iterative recovery of the original image's main structure. Ideally, the R^0 contains no small-scale details, but only with well-preserved main structure of the input image.

By examining the R^K generated in the RGF [15], it has been observed that many small-scale details are *not* removed satisfactorily. Moreover, distinct distortions have incurred on the main structure. The root of these problems mainly lies in the RGF's initial guidance image R^0, which was generated by exploiting an *isotropic* Gaussian filter (Fig. 3(a)) with a fixed

and large standard deviation (normally, σ = 5 or even larger). Although the Gaussian kernel can help to remove small-scale details, but this is achieved at the expense of yielding distorted main structure. To address this issue, we replace RGF's Gaussian filter by our developed *adaptive relative total variation filter* (ARTVF), which not only effectively remove small-scale details but also accurately preserve the main structure of the input image.

The developed ARTVF is inspired by the *relative total variation filter* (RTVF) [12], which has the following objective functional:

$$\mathcal{F}(S) = \sum_p \left\{ (S_p - I_p)^2 + \lambda \left(\frac{|(\partial_x S)_p|}{\mathcal{L}_{x,p} + \epsilon} + \frac{|(\partial_y S)_p|}{\mathcal{L}_{y,p} + \epsilon} \right) \right\}, \tag{2}$$

where I and S denote the input image and the detail-removed output image, respectively.[1] The first term in (2) is a *data fidelity* term, and the second term is a *regularization term*. The constant λ balances the contributions inserted from these two terms. In the regularization term, the partial differentiation operators ∂_x and ∂_y compute the partial derivatives of S along the horizontal (i.e., x-axis) and the vertical (i.e., y-axis) directions, respectively. The small constant $\epsilon = 10^{-3}$ is employed to avoid any possible numerical instability.

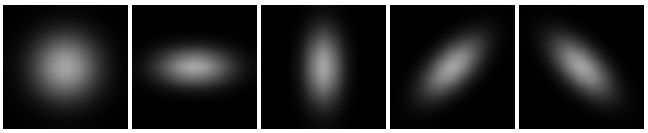
In (2), $\mathcal{L}_{x,p}$ and $\mathcal{L}_{y,p}$ are computed for each pixel p within its local window Ω_p centered on the pixel p, as follows [12]:

$$\mathcal{L}_{x,p} = \left| \sum_{q \in \Omega_p} w_{p,q} \cdot (\partial_x S)_q \right|, \ \mathcal{L}_{y,p} = \left| \sum_{q \in \Omega_p} w_{p,q} \cdot (\partial_y S)_q \right|, \tag{3}$$

where $w_{p,q}$ is a Gaussian weighting factor between the current pixel p and its neighboring pixel q within Ω_p. Note that $\mathcal{L}_{x,p}$ and $\mathcal{L}_{y,p}$ play an important role on discriminating unwanted small-scale image details from important large-scale main structure. These values tend to be small, if the local region Ω_p surrounding the current pixel p only contains small-scale textures. This is because the directions of gradients measured in such kind of regions are more *inconsistent* or *scattered*, thus the gradients will counteract each other when being summed up. With smaller $\mathcal{L}_{x,p}$ and $\mathcal{L}_{y,p}$ in (2), the regularization term will impose *larger* penalties to those regions with small details only than to those ones containing large-scale main structure.

The main difference between our objective functional defined in (2) and that in the RTVF [12] lies in the ways on computing the $w_{p,q}$ in (3). In [12], $w_{p,q}$ is computed based on an *isotropic* Gaussian kernel with the standard deviation σ_v (denoted as G_{360°) as demonstrated in Fig. 3(a). In our approach, four additional *anisotropic* Gaussian kernels are included and denoted as $G_{0^\circ}, G_{90^\circ}, G_{45^\circ}$, and G_{135°, respectively (Fig. 3(b)-(e)). For each anisotropic Gaussian kernel,

[1] For ease of paper development, both images I and S are grayscale. For color images, the described detail-removal filtering process will be applied to its three color channels individually.



(a) G_{360° (b) G_{0° (c) G_{90° (d) G_{45° (e) G_{135°

Fig. 3. Five Gaussian weighting kernels employed in our proposed ARTVF.

(a) (b) (c) (d)

Fig. 4. Filtered results of different detail-removal filters: (a) input image; (b)-(d) filtered images by using (b) the Gaussian filter, (c) the RTVF [12], and (d) the proposed ARTVF.

the standard deviations along its major and minor axes are σ_v and $\sigma_v/2$, respectively. Further define the *total response* resulted by exploiting the Gaussian kernel G_θ as $\mathcal{L}_\theta$, where $\mathcal{L}_\theta = \mathcal{L}_{x,p}(\theta) + \mathcal{L}_{y,p}(\theta)$. For each pixel p, its $w_{p,q}$ will be computed by exploiting each of these five Gaussian kernels individually. The values of $\mathcal{L}_{x,p}$ and $\mathcal{L}_{y,p}$ to be used in (2) are identified among all the computed $\mathcal{L}_\theta$ by checking which one yields the largest response; that is, $\mathcal{L}^*_\theta = \max \mathcal{L}_\theta$. Meanwhile, the following condition is also required to be satisfied: $\mathcal{L}^*_\theta > \eta \cdot \mathcal{L}_{360^\circ}$. This condition ensures that the pixel p is indeed an 'edge' pixel; otherwise, the $\mathcal{L}_{x,p}$ and $\mathcal{L}_{y,p}$ obtained in $\mathcal{L}_{360^\circ}$ will be used in (2) instead. The threshold η is empirically set to 2 and used throughout all of our experiments.

Since the objective functional as defined in (2) is non-convex, an iterative numerical approach is adopted in [12] to approximate the optimal solution; this is similar to the iterative re-weighted least square algorithm [19]. In Fig. 4, the results obtained by exploiting different filters for removing small details of image content are demonstrated. One can see that the proposed ARTVF clearly outperforms the RTVF method and the Gaussian filter used in the RGF algorithm, not only on removing small-scale details but also on preserving large-scale main structure, especially for the parts that have thin edges or in small size, as indicated by the red and blue arrows in Figs. 4(c) and 4(d), respectively.

Despite the promising filtering results yielded by the developed ARTVF, the proposed method could still leave some small details unfiltered. These unremoved details are denoted as *residues* in our paper, and they could be easily incurred in the locations nearby the main structure. This is mainly due to the fact that both $\mathcal{L}_{x,p}$ and $\mathcal{L}_{y,p}$ are computed over a local window Ω_p. Consequently, when the small-scale details are nearby an edge of the main structure, the measured $\mathcal{L}_{x,p}$ and $\mathcal{L}_{y,p}$ could be fairly close to that of edge pixels. In this

case, the small-scale details will be treated as 'edge' pixels and thus preserved. To solve this issue, we found out that by simply applying a Gaussian filter G to S (generated by our ARTVF), the above-mentioned residues of small-scale details can be effectively removed; i.e., $R^0 = G * S$, where $*$ denotes the convolution operation. Note that the value of the standard deviation used at this stage must be small for avoiding significant distortion incurred to the main structure. For that, the standard deviation is empirically determined and set as 2. Although the initial guidance image becomes slightly more blurred after the above-mentioned convolution, however it is a much better starting point for conducting the recovery process of the main structure.

After the initial guidance image R^0 is generated as above-mentioned, the same structure recovery process as depicted in Fig. 2 will be conducted to generate the final output image R^K. Since the structure distortion incurred in R^0 is much slighter in our proposed ESF than in the original RGF, the iterative structure recovery process generally requires fewer iterations to converge; typically, 3 iterations is sufficient while the RGF requires 5 iterations or more. Ideally, the recovered image R^K should deliver the same main structure of the original image I without showing any unwanted small detail.

4. EXPERIMENTAL RESULTS

Extensive simulations have been conducted to evaluate the performance of our proposed ESF. The filtered results of the proposed ESF for conducting SICF are compared to several state-of-the-art methods, including two representative edge-aware filters (i.e., the *bilateral filter* (BF) [2] and the *weighted-least-square* (WLS) filter [3]) and two recently-developed scale-aware filters (i.e., the *relative total variation filter* (RTVF) [12] and the *rolling guidance filter* (RGF) [15]).

Four parameters are required to be set in our ESF—two are set for removing small details (i.e., the constant λ in (2) and the standard deviation σ_v of the Gaussian kernels as depicted in Fig. 3) and two are used in the follow-up main structure recovery process (i.e., the spatial standard deviation σ_s and the range standard deviation σ_r of the bilateral filter in (1)). Note that the larger the λ and σ_v, the more small details will be removed. In contrast, σ_s and σ_r control the main structure recovering ability of the follow-up joint filtering stage. To prevent the removed small details from being recovered back, a sufficiently large value of σ_s should be used; typically, $\sigma_v \leq \sigma_s \leq 2\sigma_v$ is used in our simulations. As for σ_r, it controls how well the main structure can be recovered. Note that a larger value of σ_r tends to make the recovered main structure more blurred. Finally, it should be noted that the value-setting of these four parameters should be determined based on the desired degree of small detail removal, and this decision intimately depends on each image processing application.

Fig. 5 demonstrates the filtered images obtained by using different filters. It can be observed that both BF and WLS fail to completely remove those high-contrast textures as shown in Fig. 5(b) and (c). Although two state-of-the-art scale-aware filters RTVF and RGF can yield much improved SICF results, but some residues (i.e., small textures) and some degradations on the main structure are quite noticeable as shown in Fig. 5(d)-(e). The proposed ESF, in contrast, has successfully avoided all the above-mentioned concerns and delivered the best visual quality of the SICF-filtered image (see Fig. 5(f)). Due to the attractive edge-preserving scale-aware filtering ability, the proposed ESF could benefit a variety of practical image processing applications, such as image enhancement, texture replacement, image de-halftoning, edge detection, to name a few[2].

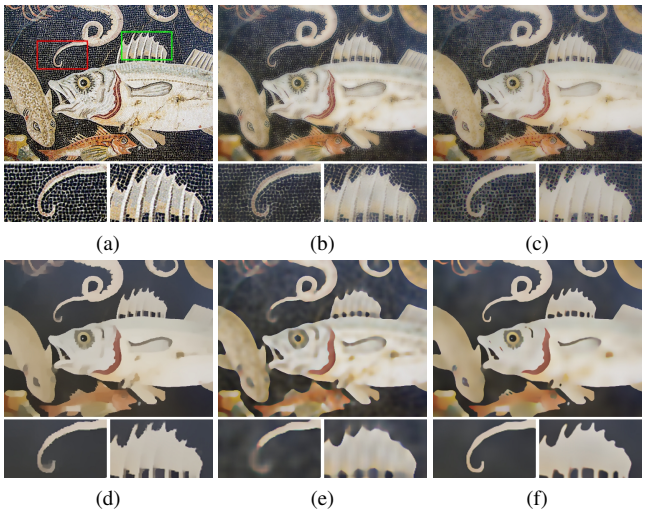


Fig. 5. Filtered results of the test image as presented in (a) by using: (b) BF [2] ($\sigma_s = 4$, $\sigma_r = 0.5$), (c) WLS [3] ($\lambda = 4$, $\alpha = 1$), (d) RTVF [12] ($\lambda = 0.01$, $\sigma = 5$), (e) RGF [15] ($\sigma_s = 5$, $\sigma_r = 0.1$), and (f) the proposed ESF (in which $\lambda = 0.05$, $\sigma_v = 5$, $\sigma_s = 10$, $\sigma_r = 0.05$), respectively.

5. CONCLUSION

In this paper, a novel *semantic image content filtering* (SICF) method is developed, called the *edge-preserving scale-aware filter* (ESF). The proposed ESF is inspired by a recently introduced scale-aware filter, called the *rolling guidance filter* (RGF), but with a significant improvement on its edge-preserving capability. This is achieved due to a much improved initial guidance image generated by our proposed *adaptive relative total variation filter* (ARTVF) method and the follow-up residue removal filtering strategy. Extensive simulations have demonstrated that the proposed ESF is able to deliver much superior SICF results compared to several comparable state-of-the-art filtering methods.

[2]For more examples, please refer to the supplementary material.

6. REFERENCES

[1] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 12, no. 7, pp. 629–639, 1990.

[2] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *IEEE Int. Conf. on Comp. Vis.*, 1998, pp. 839–846.

[3] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," in *ACM Trans. on Graphics*, 2008, vol. 27, pp. 67:1–67:10.

[4] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

[5] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l0 gradient minimization," in *ACM Trans. on Graphics*, 2011, vol. 30, pp. 174:1–174:12.

[6] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. on Graphics*, vol. 30, no. 4, pp. 69:1–69:12, 2011.

[7] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[8] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. Do, "Fast global image smoothing based on weighted least squares," *IEEE Trans. on Image Process.*, vol. 23, no. 12, pp. 5638–5653, 2014.

[9] J. Lu, K. Shi, D. Min, L. Lin, and M. Do, "Cross-based local multipoint filtering," in *IEEE Conf. on Comp. Vis. and Pattern Reco.*, 2012, pp. 430–437.

[10] K. Subr, C. Soler, and F. Durand, "Edge-preserving multiscale image decomposition based on local extrema," *ACM Trans. on Graphics*, vol. 28, no. 5, pp. 147:1–147:9, 2009.

[11] L. Karacan, E. Erdem, and A. Erdem, "Structure-preserving image smoothing via region covariances," *ACM Trans. on Graphics*, vol. 32, no. 6, pp. 176:1–176:11, 2013.

[12] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. on Graphics*, vol. 31, no. 6, pp. 139:1–139:10, 2012.

[13] L. Bao, Y. Song, Q. Yang, H. Yuan, and G. Wang, "Tree filtering: Efficient structure-preserving smoothing with a minimum spanning tree," *IEEE Trans. on Image Process.*, vol. 23, no. 2, pp. 555–569, 2014.

[14] H. Cho, H. Lee, H. Kang, and S. Lee, "Bilateral texture filtering," *ACM Trans. on Graphics*, vol. 33, no. 4, pp. 128:1–128:8, 2014.

[15] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Proceedings of the European Conference on Computer Vision*, 2014, pp. 815–830.

[16] B. Ham, M. Cho, and J. Ponce, "Robust image filtering using joint static and dynamic guidance," in *IEEE Conf. on Comp. Vis. and Pattern Reco.*, 2015, pp. 4823–4831.

[17] F. Zhang, L. Dai, S. Xiang, and X. Zhang, "Segment graph based image filtering: Fast structure-preserving smoothing," in *IEEE Int. Conf. on Comp. Vis.*, 2015, pp. 361–369.

[18] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. on Graphics*, vol. 23, no. 3, pp. 664–672, 2004.

[19] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, 2008, pp. 3869–3872.