

Adaptive Image Decomposition by Improved Bilateral Filter

Liu Qiegen
Department of BME,
Shanghai Jiaotong University.
Shanghai, China

Luo Jianhua
Professor
Department of BME,
Shanghai Jiaotong University.
Shanghai, China y

Zhu Yuemin
Professor
Centre National de la
Recherche Scientifique, INSA
Lyon

ABSTRACT

This paper presents an adaptive image decomposition method based on the newly proposed texture descriptor, which combines contrast variation and structure correlation. By incorporating it into the classical edge-preserving bilateral filter, we construct an adaptive nonlinear filter capable of preserving edges and removing texture parts effectively. We derive the variational formulation of the proposed filter and explore its relationship with robust statistics. Comparative experiments against other state-of-the-art decomposition approaches show that better separation results of geometric structure from texture are obtained.

Keywords

Image decomposition, Bilateral filter, Texture descriptor, Robust statistics.

1. INTRODUCTION

Image decomposition is a common first step to the solution of many image processing. Suppose an image f is separated as the sum of two independent components $f = u + v$: a piecewise smooth function u with quasi-flat intensity plateaus and jump discontinuities, the ‘cartoon’, that contains main large-scale structure features of the image and a small-scale oscillatory function v , capturing texture and possibly noise, usually with some periodicity and oscillatory nature. Then the structure component can be used for feature detection, segmentation and object recognition, while the texture component for solving various texture-depended problems e.g. classification, surface analysis, shape/orientation from texture. It has motivated solutions to image processing like inpainting [1], demosaicing [2] and registration [3] by adapting algorithms for the two components.

Model based approaches and adaptive filtering are two main approaches that address the image decomposition problem. The first category of methods is often related to total variation (TV) minimization. They perform decomposition by modeling the cartoon u with TV semi-norm and appropriate norms for modeling oscillating features v . The original formulation of such approaches is due to Meyer [4]. He proposed starting from the ROF algorithm [5]:

$$\inf_{u \in BV(\Omega)} \int_{\Omega} |Du| + \frac{\lambda}{2} \int_{\Omega} |(f - u)(x)|^2 dx \quad (1)$$

where $\Omega \subset \mathbb{R}^2$ denotes an open and bound domain and $\int_{\Omega} |Du| = \sup \{ \int_{\Omega} u(x) \operatorname{div}(\zeta(x)) \mid \zeta \in C_c^1(\Omega; \mathbb{R}^2), \|\zeta\|_{L^\infty(\Omega)} \leq 1 \}$. Meyer showed that this model rejects the textures and then

proposed to use a new function space, G , by replacing the L^2 -norm by the G -norm. He proved that this space corresponds to a space of oscillating functions that are useful to model textures. This idea inspired several works. Some approximated norm like $\operatorname{div}(L^p)$ -norm [6], H^{-1} -norm [7], and homogeneous Besov space $\dot{B}_{p,q}^s$, $-2 < s < 0$, $1 \leq p, q \leq \infty$ [8] are developed. A deficiency of these models is that the cartoon part exhibits staircasing.

The second type of decomposition methods exploits the spatial filtering directly. Bilateral filters [9] and feature-preserving smoothing operators have been used to decompose images [10], [11]. However, textures often have high-contrast edges and bilateral filters are not effective to handle such textures. Farbman et al. introduced a weighted least squares (WLS) based filter [12], which preserves multi-scale features better than bilateral filtering. However, their method also cannot smooth out textures of high contrast. Currently most of the edge-preserving image decomposition filters assume detail to be low contrast variation. As a result, they are impossible to preserve salient edges with lower contrast than the texture that is to be smoothed. To tackle this problem, we propose a new spatial adaptive-filtering algorithm (TDBF), which combines Texture Descriptor with the classical Bilateral Filter. The presented filter is adaptive to the local image information. At locations with high degree of textures, the filter approaches a Gaussian filter for getting the cartoon u . While at locations with low degree of textures such as the edges or homogeneous regions, the filter approaches the original bilateral filter, thus the decomposition is implemented adaptively.

2. TEXTURE DESCRIPTOR BASED BILATERAL FILTER

2.1 Proposed Texture Descriptor

In order to design a spatial adaptive filter for image decomposition, we first develop a spatially varying quality map, which delivers quality measure of texture contained in the image, and then integrate it to the filter. There are many models to design the texture-related map, the main feature of these texture estimators involves local total variation [13, 14]. In this paper, we introduce a novel texture estimator including local variation and structure correlation, which is motivated by Structure SIMilarity (SSIM) [15]. Formally, let $G_\sigma * f$ denotes the degraded image filtered by linear Gaussian filter. \mathbf{p} and \mathbf{q} denote two local image patches, whose associated central pixel is x , extracted from the original and degraded images respectively. Then the proposed texture descriptor scheme is defined as follow:

$$TD(x) = 1 - |c(\mathbf{p}, \mathbf{q})s(\mathbf{p}, \mathbf{q})|$$

$$= 1 - \frac{2\sigma_p\sigma_q + C_1}{\sigma_p^2 + \sigma_q^2 + C_1} \left| \frac{\sigma_{pq} + C_2}{\sigma_p\sigma_q + C_2} \right| = 1 - \left| \frac{2\sigma_{pq} + C_1}{\sigma_p^2 + \sigma_q^2 + C_1} \right| \quad (2)$$

Where local statistics μ_p , σ_p and σ_{pq} are defined accordingly

$$\text{as } \mu_p = \sum_{i=1}^N p_i \quad ; \quad \sigma_p = \left(\sum_{i=1}^N (p_i - \mu_p)^2 \right)^{\frac{1}{2}} \quad ;$$

$$\sigma_{pq} = \sum_{i=1}^N (p_i - \mu_p)(q_i - \mu_q). \text{ The constants } C_1 \text{ and } C_2 = C_1/2$$

are introduced for numerical stability. When this local measurement is applied to an entire image using a sliding window approach, a texture quality map is obtained. The texture measure contains two components: the contrast variation comparison $c(\mathbf{p}, \mathbf{q})$ calculates the reduction of contrast variation due to the degraded operation; the correlation $s(\mathbf{p}, \mathbf{q})$ quantifies the structural correction between the original and degraded images. The local variation does not decrease by low pass filtering in cartoon region, while in texture region it decreases drastically. Besides, structure correlation is also a useful tool to describe the local oscillatory changed by the low pass filter.

The TD value is in the interval $[0, 1]$, measuring the degree of the pixel belonging to a textural region. The more it is close to 1, the more the considered point belongs to a textured region. The more it is close to 0, the more the considered point belongs to cartoon region. The rationality of the combination of contrast and structure components accounting for texture measure is illustrated in Fig. 1: the reference image shown in plot (a) is composed of solid circle and parallel stripes, where the texture and cartoon parts are interlaced highly. As can be seen from plot (b), the contrast comparison between the reference and its corresponding degraded one basically reflects where the textured region is (the blacker the pixel is, the more possible it belongs to the texture region). On the other hand, combined with the structure correlation displayed in plot (c) as a supplementary criterion, the compound texture map TD indicates the texture parts substantially.

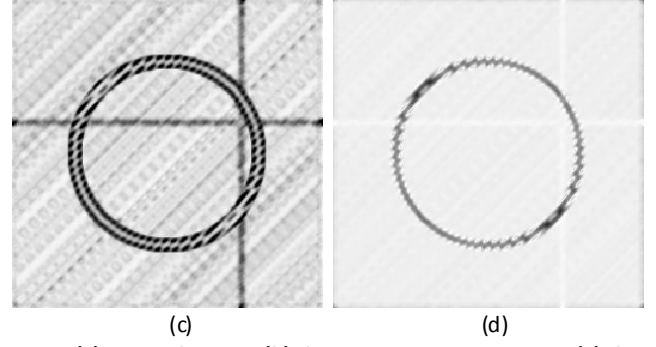
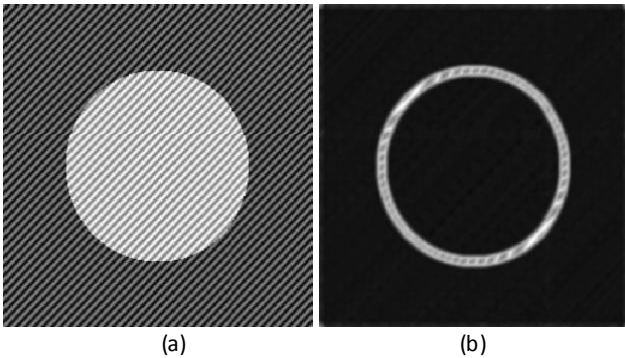


Fig 1: (a) Original image. (b) the contrast component c . (c) the structural similarity component $|s|$. (d) the function $TD = 1 - |c \cdot s|$.

2.2 Texture Descriptor based Bilateral Filter

In the successful edge-preserving bilateral filter [9], the intensity value at each pixel is replaced by a weighted mean of the intensity values from its neighbors, with the weights decreasing both with spatial distance and with difference in intensity value. For more precision, let $\Omega_x(N)$ gives the set of pixels in $(2N+1) \times (2N+1)$ neighborhood of central pixel x . The weight of each pixel $y \in \Omega_x(N)$ with respect to x is the multiplication of spatial and radiometric component:

$$w(x, y) = w_s(x, y)w_r(x, y) \quad (3)$$

where $w_s(x, y) = e^{-|x-y|^2/2\sigma_s^2}$ and $w_r(x, y) = e^{-|f(x)-f(y)|^2/2\sigma_r^2}$. σ_s^2 and σ_r^2 adjust the spatial similarity and the intensity similarity respectively. The spatial component of the weight decreases the influence of pixels far away from x to generally reduce blurring, while the radiometric component diminishes the influence of pixels with significantly different intensities to keep the edges of distinct image regions sharp.

When coming to the image decomposition, since this purely gradient-dependent approach assumes detail is low-contrast, it is unable to distinguish between high-contrast, fine-scale features and edges of similar contrast that should to be preserved. Intuitively, in the texture-like region the radiometric weight should be posed more slightly so that radiometric differences become irrelevant, thus we incorporate the adaptive texture descriptor into the original bilateral filter. By adding a spatial-adaptive variable $J(x, y)$ on radiometric weighting function w_r . The new modified filter is:

$$w(x, y) = w_s(x, y)w_r(x, y)^{J(x, y)} \quad (4)$$

where $J(x, y) = e^{-(TD(x)+TD(y))/2\sigma_j^2}$ assumes value in $[0, 1]$, and σ_j^2 controls the behavior of the weights. After the weights are normalized, we have the decomposition method as follows:

$$u(x) = \frac{\sum_{y \in \Omega_x(N)} w(x, y)f(y)}{\sum_{y \in \Omega_x(N)} w(x, y)}, \quad v(x) = f(x) - u(x) \quad (5)$$

where u stands for the target cartoon part and v for the texture part. We would like to use the radiometric weight more heavily when $J(x, y) \approx 1$ to cartoon region and less heavily when $J(x, y) \approx 0$. Conversely, we would like to use the texture descriptor weight less heavily when $J(x, y) \approx 1$ and more heavily when $J(x, y) \approx 0$. In this way, the appropriate weighting function is applied on a pixel-by-pixel basis. The geometry adaptive ability makes better weighting between the centered pixel and the other nearby pixels, thereby integrating the weighted average of all the considered pixels will lead to wonderful estimate.

Aided by the accurate texture descriptor, the proposed filter has adaptive weight distribution in different regions. Fig. 2 depicts the weight distributions of four pixels in the Barbara image. The former two pixels locate in cartoon-like region, one locates around the edge region and the other locates around the flat

region; the latter two pixels locate in texture-like region. As can be seen, the filter's weights with respect to the first pixel in cartoon-like region assemble around the direction of edge, and the weight is almost to zero in other regions. This indicates that our filter inherits the perfect edge-preserving property of the original bilateral filter.

For the third and fourth pixels located in texture-like region, we can observe from the bottom row of Fig. 2(b) that both of the weight distributions are almost identical to each other. This is because that the texture descriptor indeed impacts the weight distributions. Seen from Fig. 2(c), the texture descriptor TD in both neighborhoods of the third and fourth pixels are high and almost the same. So the weight distributions are very similar to the Gaussian spatial weighting function $w_s(x, y)$.

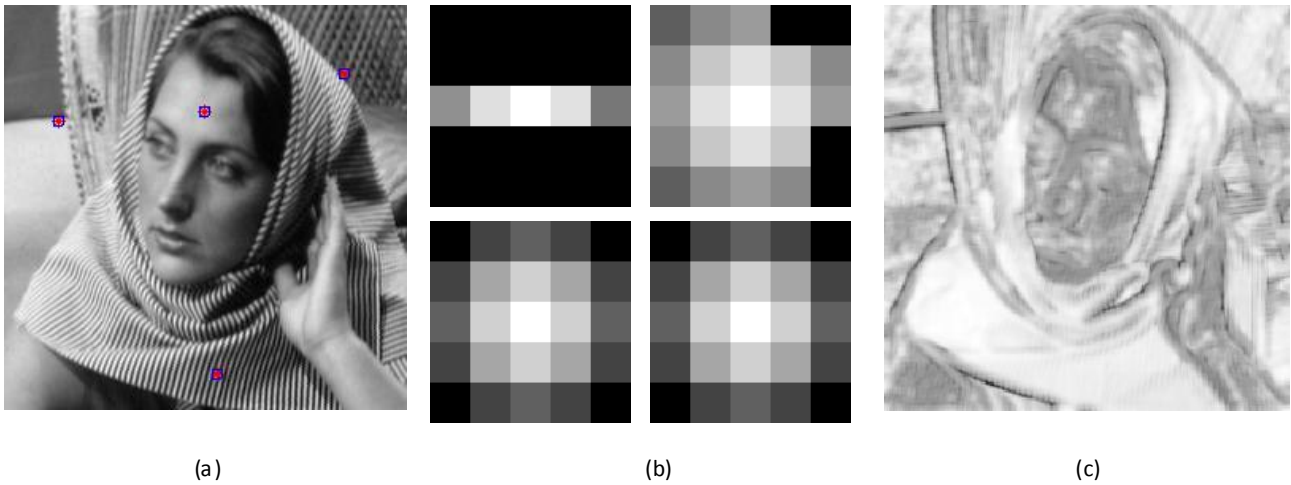


Fig 2: (a) Image Barbara. (b) weight distributions of four interested pixels. The block size is 5×5 . (c) texture descriptor TD .

2.3 Variational Formulation Interpretation

In this subsection, we derive the filter's variational formulation and demonstrate its relation to robust statistics. We consider minimizing the following energy:

$$\text{Min}_{u(x)} \rho(u(x)) \quad (6)$$

where

$$\rho(u(x)) = \sum_{y \in \Omega_x(N)} w_s(x, y) \left[1 - \frac{1}{J(x, y)} \left(e^{-\frac{|u(x)-u(y)|^2}{2\sigma_r^2}} \right)^{J(x, y)} \right]$$

Minima of the minimization problem (6) are yielded as solutions of the Euler-Lagrange equation,

$$\sum_{y \in \Omega_x(N)} w_s(x, y) \left(e^{-\frac{|u(x)-u(y)|^2}{2\sigma_r^2}} \right)^{J(x, y)} (u(x) - u(y)) = 0 \quad (7)$$

the equation can be solved by gradient descent method, namely:

$$u^{n+1}(x) = u^n(x)$$

$$+ \lambda_{EL} \sum_{y \in \Omega_x(N)} w_s(x, y) \left(e^{-\frac{|u^n(x)-u^n(y)|^2}{2\sigma_r^2}} \right)^{J(x, y)} (u^n(x) - u^n(y)) \quad (8)$$

where λ_{EL} is the step size.

or a fixed-point iteration, resulting in

$$u^{n+1}(x) = \frac{\sum_{y \in \Omega_x(N)} w_s(x, y) \left(e^{-\frac{|u^n(x)-u^n(y)|^2}{2\sigma_r^2}} \right)^{J(x, y)} u^n(y)}{\sum_{y \in \Omega_x(N)} w_s(x, y) \left(e^{-\frac{|u^n(x)-u^n(y)|^2}{2\sigma_r^2}} \right)^{J(x, y)}} \quad (9)$$

Compared (5) with (9), it can be found that our proposed filter defined in (5) is essentially one iteration output of the formulation (9) under initial value $u^0 = f$. Let

$t = (u(x) - u(y)) / \sigma_r$, $g(t) = e^{-t^2/2}$, we obtain robust penalty function $\rho(t, J) = 1 - g(t)^J / J$ and influence function $\psi(t, J) = \rho'(t, J) = g(t)^J t$ from (6) and (8) respectively:

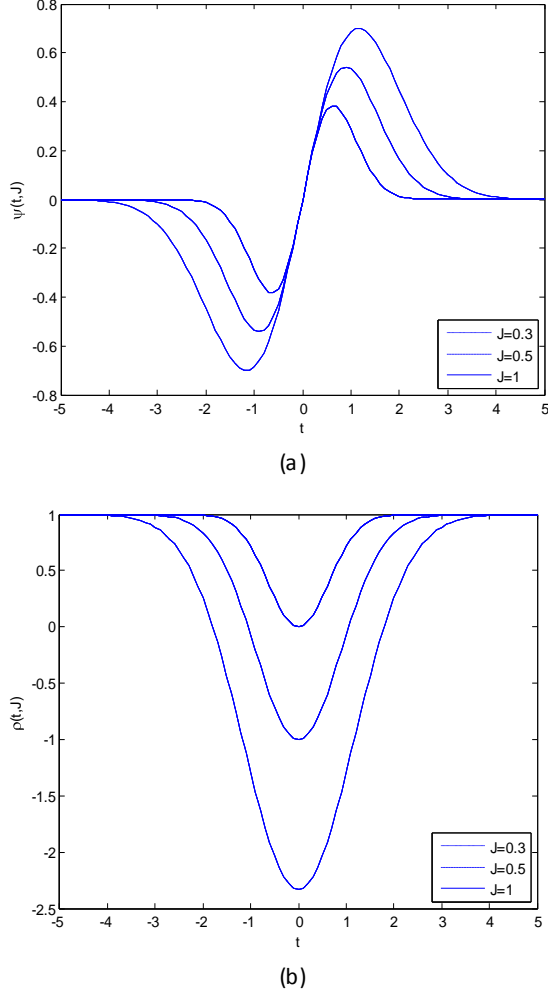


Fig 3: (a) Influence functions $\psi(t, J)$. (b) energy functionals $\rho(t, J)$, $\sigma_r = 0.6$, $J = 0.3, 0.5, 1$.

1) In the extreme case $J(x, y) \equiv 1$, formulation (6) reduces to the variational formulation of the original bilateral filter, which has been proven to be a robust statistical estimator [16]. Furthermore, the iterative formulation (8) and (9) solve the same energy minimization problem (6). The so-called influence function ψ characterizes the influence of a sample on the estimate. It depends on a parameter σ_r that provides the notion of scale in the intensity domain, and controls where the influence function becomes redescending, and thus which values are considered outliers. The point where the influence function becomes redescending is called “threshold point”. A display of the influence function ψ is presented in Fig. 3(a). As can be seen, ψ gives much less weight to outliers and is therefore robust.

2) Our contribution in this work is that we adaptively minimize the energy function by integrating the texture descriptor based variable $J(x, y)$, therefore $\psi(t)$ and $\rho(t)$ are extended to be $\psi(t, J)$ and $\rho(t, J)$. As displayed in Fig. 3(a), given the predefined parameter $\sigma_r = 0.6$, the curve line of influence function $\psi(t, J)$ with intensity difference increase as variable $J(x, y)$ decreases. This indicates that the more pixel x or y seems texture-like, the bigger the “threshold point” is, and thus only these pixels with very big intensity difference will be considered as outliers. This can also be validated from the corresponding energy functional $\rho(t, J)$, which is shown in Fig. 3(b). The bigger variable $J(x, y)$ is, the more penalty the function is, thus prefers to cartoon component.

3. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our filter, we compare our results with two of the most representative state-of-the-art algorithms: gradient-based edge preserving smoothing technique WLS filtering¹ and the variation minimization OSV model ($TV - H^{-1}$) [7]. Implementing the OSV model amounts to the minimization of the following energy

$$\inf_{u \in BV(\Omega)} \int_{\Omega} |Du| + \frac{\lambda}{2} \int_{\Omega} |L_{\sigma} * (f - u)(x)|^2 dx \quad (10)$$

where the smoothing kernel L_{σ} is defined by $\hat{L}_{\eta}(\xi) = 1 / (\|\xi\|^2 + \eta)$ in the Fourier domain, $\eta = 0.05$. The detailed implementation of the scheme is followed from Peyre’s numerical tour².

We tested textile texture decomposition by applying the three models to image “Barbara” as depicted in Fig 2(a), where there is a high presence of textures combined with non textured parts. Ideally, only the stripes on Barbara’s scarf should be extracted. As can be seen from Fig. 4, due to the high-contrast stripes details, the cartoon component of WLS filtering result still has texture information and its associated texture component has structural information. The cartoon component of OSV model has lost shading information and contains staircasing artifacts. On the other hand, due to the satisfied texture descriptor, our method gave remarkable results as little background appears in the v parts. It preserves subtle features and effectively smoothes the textured details. Furthermore, the edges along the scarf are better preserved in the cartoon parts by the new filter.

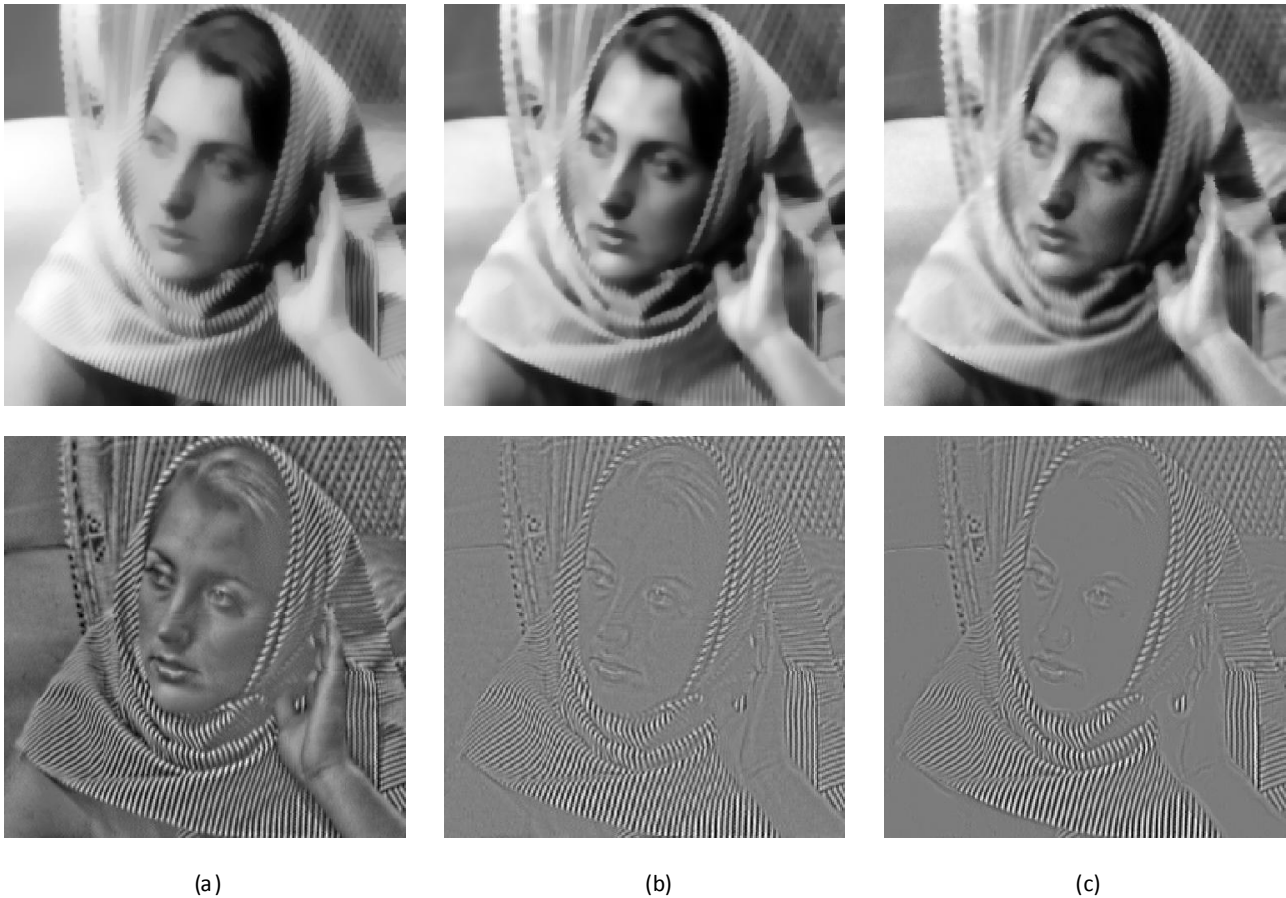
The structures + textures decomposition results of image “Lena” are shown in Fig. 5. The input image contains lots of structures such as edge, flat, and corner areas, as well as many texture details on the hat. Again, we can make similar observation: compared with WLS filter and OSV model, our method separates the textured details shown in v from non-textured images kept in u better. Particularly, on one hand, we still can see the background regions and the right arm in the results

¹ Mat lab code of the algorithm is available at <http://www.cs.huji.ac.il/~danix/epd/wlsFilter.m>.

² The code is available in http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/variational_3_image_separation/#1

produced by WLS and OSV. These are not seen in the v component produced by our method. On the other hand, structure details like the eyes and nose are much better kept in the u component of our method. In other words, by taking advantage of the texture descriptor, the new filter performs much better in keeping the main contours in the u component,

instead of the v component. Our method integrates the texture descriptor into the bilateral filter directly, so it prefers to the edge-preserving property. It worth be noting that fewer edges boundaries are along the hat in the v component, thereby giving better visual results.



(a) (b) (c)
Fig 4: Decomposition results by (a) WLS ($\lambda_{WLS} = 0.5, \alpha_{WLS} = 1.2$), (b) OSV ($\lambda = 0.5$) and (c) TDBF ($N = 2, \sigma_s = N, \sigma_r = 1, \sigma_f = 0.017N$).

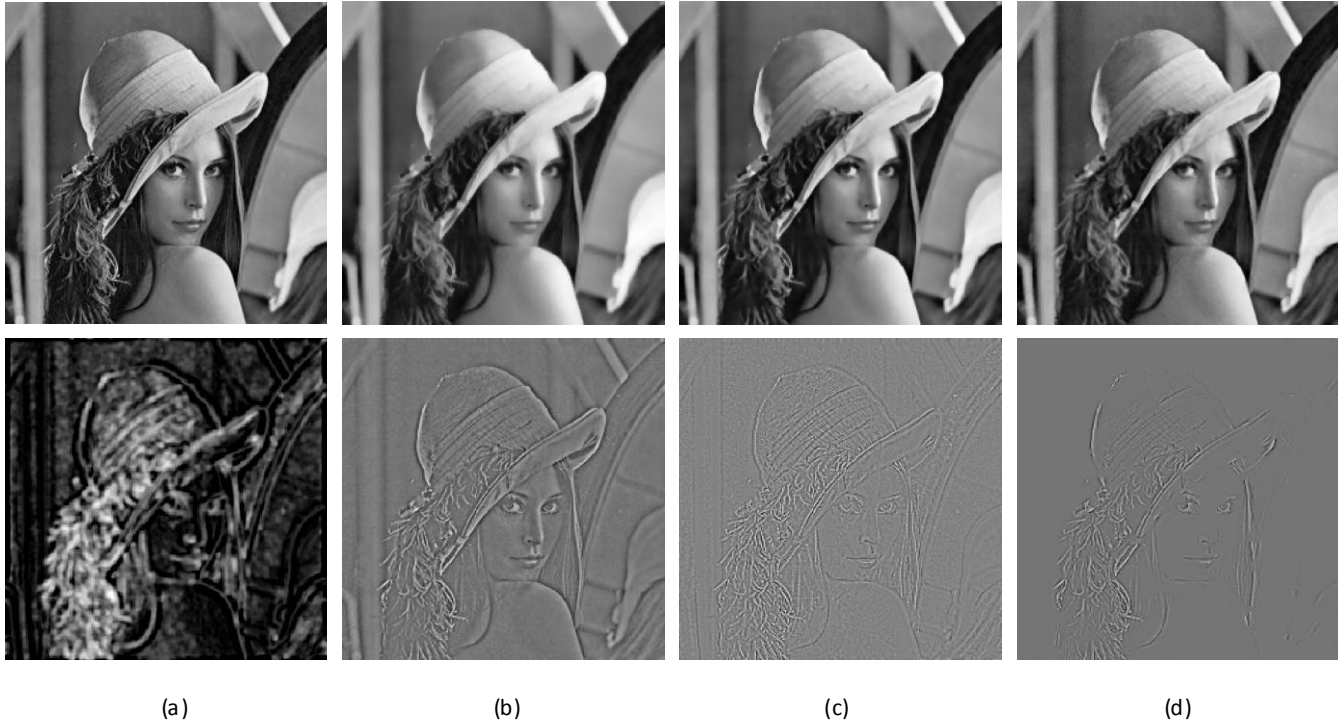


Fig 5: (a) the original image and TD . Decomposition results by (b) WLS ($\lambda_{wls} = 0.5$, $\alpha_{wls} = 0.5$). (c) OSV ($\lambda = 2$) and (d) TDBF ($N = 2$, $\sigma_s = 0.38N$, $\sigma_r = 0.1$, $\sigma_j = 0.015N$).

4. CONCLUSION

This work has presented a selected filter based approach for the image decomposition problem. By incorporating additional information of the texture into the classical bilateral filter, we derived an approximate Gaussian smoothing in the texture regions and an approximate bilateral filter in the flat and edge regions. We have studied the adaptive filter and illustrated this new method with numerical examples. In the current research, the filter developed involves four parameters: σ , σ_s^2 , σ_r^2 and σ_j^2 . How to reduce its parameters freedom for guiding the filter better is a future research direction.

5. ACKNOWLEDGMENTS

This work was partly supported by Shanghai International Cooperation Grant under 06SR07109, Region Rhône-Alpes of France under the project Mira Recherche 2008, and the joint project of Chinese NSFC (under 30911130364) and French ANR 2009 (under ANR-09-BLAN-037201).

6. REFERENCES

[1] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture inpainting" *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 882–889, 2003.
 [2] T. Saito, Y. Ishii, H. Aizawa, D. Yamada, and T. Komatsu, "Image-processing approach via nonlinear image decomposition for a digital color camera," in 15th IEEE

International Conference on Image Processing (ICIP), 2008, pp. 905-908.
 [3] D. Paquin, D. Levy, E. Schreibmann, and L. Xing, "Multiscale image registration," *Math. Biosci. Eng.*, vol. 3, pp. 389–418, 2006.
 [4] Y. Meyer, "Oscillating patterns in image processing and nonlinear evolution equations," Volume 22 of University Lecture Series. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.
 [5] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D.*, vol. 60, pp. 259–268, 1992.
 [6] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing*, vol. 19, no. 1, pp. 553–572, 2003.
 [7] S.J. Osher, A. Sole, and L.A. Vese, "Image decomposition and restoration using total variation minimization and the H-1 norm," *SIAM Multiscale Modeling and Simulation*, vol. 1, no. 3, pp. 349-370, 2003.
 [8] J. Garnett, T. Le, Y. Meyer, and L. Vese, "Image decompositions using bounded variation and homogeneous Besov spaces," *Tech. Rep. CAM Report 05-57*, UCLA, 2005.

- [9] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proc. IEEE Int. Conf. Computer Vision, 1998, pp. 839–846.
- [10] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," ACM Trans. Graphics, vol. 21, no. 3, pp. 257–266, 2002.
- [11] S. Bae, S. Paris, and F. Durand, "Two-scale tone management for photographic look," ACM Trans. Graphics, vol. 25, no. 3, pp. 637–645, 2006.
- [12] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," ACM Trans. Graphics, vol. 27, no. 3, pp. 1-10, 2008.
- [13] M.W. Tao, M.K. Johnson, and S. Paris, "Error-tolerant Image Compositing," Computer Vision, vol. 20, no. 63, pp. 31-44, 2010.
- [14] A. Baudes, T. Le, J.M. Morel, and L. Vese, "Fast cartoon + texture image filters," IEEE Trans. Image Process., vol. 19, no. 8, pp. 1978-1986, 2010.
- [15] Z. Wang, A.C. Bovik, H.R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Measurement to Structural Similarity," IEEE Trans. Image Process., Vol. 13, No. 4, pp. 600-613, April 2004.
- [16] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, "Robust statistics: The approach based on influence functions," Wiley, New York, 1986.