# Fast Augmented Lagrangian Method for Image Smoothing with Hyper-Laplacian Gradient Prior

Li Chen[1], Hongzhi Zhang[1], Dongwei Ren[1], David Zhang[1, 2], and Wangmeng Zuo[1]

[1] Computational Perception and Cognition Center, School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
{lichen.cs.hit,zhanghz0451,rendongweihit,cswmzuo}@gmail.com
[2] Biometrics Research Centre, Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
csdzhang@comp.polyu.edu.hk

**Abstract.** As a fundamental tool, $L_0$ gradient smoothing has found a flurry of applications. Inspired by the progress of research on hyper-Laplacian prior, we propose a novel model, corresponding to $L_p$-norm of gradients, for image smoothing, which can better maintain the general structure, whereas diminishing insignificant texture and impulse noise-like highlights. Algorithmically, we use augmented Lagrangian method (ALM) to efficiently solve the optimization problem. Thanks to the fast convergence rate of ALM, the speed of the proposed method is much faster than the $L_0$ gradient method. We apply the proposed method to natural image smoothing, cartoon artifacts removal, and tongue image segmentation, and the experimental results validate the performance of the proposed algorithm.

**Keywords:** Image smoothing, augmented Lagrangian method, hyper-Laplacian gradient prior.

## 1 Introduction

Noise and blur usually are inevitable in real world images. In many image processing applications, e.g., edge detection [1], object segmentation [2], etc., it is necessary to enhance the well-structured components, whereas suppressing noise and unnecessary texture. As a conventional approach for image denoising and enhancement, image smoothing is a well-studied problem with quite a number of methods in literatures, such as bilateral filtering (BLF) [3], weighted least squares (WLS) [4], total variation (TV) [5], $L_0$-norm smoothing [6].

Recent studies on natural image statistics have shown that heavy-tailed image gradient distribution [7] is an effective prior and can be well modeled by hyper-Laplacian with $0 < p < 1$ [8, 9, 10], which has been applied in several applications, e.g., image deburring, leading to superior results. At the same time, many efforts have been devoted to research on the $L_p$ optimization, e.g., iteratively reweighted least squares (IRLS) [11, 12], iteratively reweighted $L_1$-minimization (IRL1) [13]. Recently, Zuo *et al*. proposed a generalized iterative shrinkage algorithm (GISA) [14].

S. Li et al. (Eds.): CCPR 2014, Part II, CCIS 484, pp. 12–21, 2014.

In this paper, we propose to model image prior as an $L_p$-norm of gradients, which can better maintain the general structure, whereas suppressing insignificant texture. Furthermore, we adopt the augmented Lagrangian method (ALM) to efficiently solve the optimization problem, resulting in an ALM-Lp algorithm. Compared with $L_0$-norm smoothing, ALM-Lp algorithm is more effective in removing more highlights and avoiding color distortion, leading to a better visual perception.

(a) Input (b) BLF (c) WLS (d) TV (e) $L_0$-norm (f) our result

**Fig. 1.** The smoothing results on image *pflower*

Fig. 1 compares the image smoothing results obtained using different approaches on a natural image *pflower*. From Fig. 1(b)-(d), one can see that BLF, WLS, and TV are valuable in wiping off noises, but are limited in removing detailed textures and preserving the salient edges. And the result of $L_0$ smoothing is much better in the above aspects, but has color distortion to some extent. Besides, it performs poor in removing impulse noise-like highlights in the close-up. Compared with the competing methods, our result can obtain better smoothing result, which is effective in preserving the salient edges and removing noises and detailed textures. Moreover, thanks to the fast convergence rate of ALM, ALM-Lp is faster than $L_0$ smoothing.

The remainder of this paper is organized as follows. Section 2 presents a brief review of $L_0$ smoothing and the introduction of generalized shrinkage/thresholding (GST) algorithm. In Section 3, we present the $L_p$-norm smoothing model and the ALM-based optimization. Section 4 provides the experiment results. Finally, we end this paper with some concluding remarks in Section 5.

## 2 Related Work and Prerequisites

In this section, we first briefly review the $L_0$ smoothing method, and then summarize the generalized shrinkage/thresholding (GST) function used in this paper.

### 2.1 $L_0$ Smoothing Method

Xu *et al*. [6] proposed a smoothing model based $L_0$ norm of gradients:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 + \mu\|\mathbf{D}\mathbf{x}\|_0, \tag{1}$$

where **y** and **x** denote the input and smoothed image, respectively, $\mu$ is a smoothing weight, $\|\bullet\|_0$ denotes the $L_0$ norm that counts the number of non-zero entries, and $\mathbf{D}=[\mathbf{D}_h,\mathbf{D}_v]$ denotes the gradient operator that includes the horizontal component $\mathbf{D}_h$ and vertical component $\mathbf{D}_v$, respectively. Then, Xu *et al*. presented an alternating optimization strategy with half-quadratic splitting to tackle this problem.

The proposed ALM-Lp method is different from $L_0$ smoothing at two aspects. First, we adopt the $L_p$-norm gradient prior while $L_0$ smoothing adopts the $L_0$-norm gradient prior. Previous studies [7] indicated that image gradients typically follow hyper-Laplacian distribution with $0.5 \le p \le 0.8$, which makes ALM-Lp more suitable for image smoothing. Second, ALM-Lp uses the augmented Lagrangian method (ALM) to solve the optimization, and is more efficient.

### 2.2 The Generalized Shrinkage/Thresholding (GST) Function

$L_p$-norm minimization problem is the key of many sparse coding problems. Here, we discuss the simplest $L_p$-minimization problem as follows,

$$\min_x \frac{1}{2}(x-y)^2+\lambda|x|^p. \tag{2}$$

Zuo *et al*. [14] introduced the generalized shrinkage/thresholding (GST) function which is a generalization of the soft-thresholding operator. Compared with the existing solvers for $L_p$-norm minimization, e.g., IRLS, LUT [7], and ITM-$L_p$ [15], GST is very efficient and converges to the correct solution. The GST function is defined as,

$$T_p^{GST}(y;\lambda)=\begin{cases}0, & \text{if } |y|\le \tau_p^{GST}(\lambda)\\ \operatorname{sgn}(y)S_p^{GST}(|y|;\lambda), & \text{else}\end{cases}, \tag{3}$$

where $\tau_p^{GST}(\lambda)=(2\lambda(1-p))^{\frac{1}{2-p}}+\lambda p(2\lambda(1-p))^{\frac{p-1}{2-p}}$ which stands for the thresholding value. Generally, if $|y|\le\tau_p^{GST}(\lambda)$, the generalized soft-thresholding operator uses the thresholding rule to assign $T_p^{GST}(y;\lambda)$ to 0; otherwise, uses the shrinkage rule to assign $T_p^{GST}(y;\lambda)$ to $\operatorname{sgn}(y)S_p^{GST}(|y|;\lambda)$. $S_p^{GST}(|y|;\lambda)$ can be obtained by iteratively performing the following operation :

$$S_p^{GST}(y;\lambda)=y\text{-}\lambda p\left(S_p^{GST}(y;\lambda)\right)^{p-1}. \tag{4}$$

Then the overall algorithm was summarized as,

$$T_p^{GST}(y;\lambda)=GST(y,\lambda,p,J), \tag{5}$$

where $J$ is the number of iterations, generally 1 or 2.

## 3 Model and Algorithm

In this section, we first present the proposed model, formulated as an $L_p$ norm minimization problem. For the non-convexity of the $L_p$ norm, it is not trivial to directly optimize the proposed model. Thus, by adopting the variable splitting strategy, we employ augmented Lagrangian method (ALM) to solve it efficiently.

The smoothing model with hyper-Laplaican gradient prior is formulated as,

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \mu\|\mathbf{D}\mathbf{x}\|_p^p. \tag{6}$$

By introducing auxiliary variable **d**= **Dx**, problem (6) can be reformulated as,

$$\min_{\mathbf{x},\,\mathbf{d}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \mu\|\mathbf{d}\|_p^p \quad s.t.\ \ \mathbf{d}=\mathbf{D}\mathbf{x}, \tag{7}$$

where $\mathbf{d}=[\mathbf{d}_h^T,\mathbf{d}_v^T]^T$ with $\mathbf{d}_h=\mathbf{D}_h\mathbf{x}$ ,$\mathbf{d}_v=\mathbf{D}_v\mathbf{x}$ . Then the augmented Lagrangian (AL) function of the problem in Eq. (7) is defined as,

$$\mathcal{L} = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \mu\|\mathbf{d}\|_p^p + \boldsymbol{\lambda}^T(\mathbf{D}\mathbf{x}-\mathbf{d}) + \frac{\delta}{2}\|\mathbf{D}\mathbf{x}-\mathbf{d}\|_2^2, \tag{8}$$

where $\boldsymbol{\lambda}$ is the Lagrangian vector, and $\delta$ is a positive penalty parameter. With minor algebra, the AL function can be rewritten as,

$$\mathcal{L}(\mathbf{x},\mathbf{d},\mathbf{q},\delta) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \mu\|\mathbf{d}\|_p^p + \frac{\delta}{2}\|\mathbf{D}\mathbf{x}-\mathbf{d}+\mathbf{q}\|_2^2, \tag{9}$$

where $\mathbf{q}=\boldsymbol{\lambda}/\delta$ and $\mathbf{q}=[\mathbf{q}_h^T,\mathbf{q}_v^T]^T$. The AL function can be optimized using the alternating direction method of multipliers, i.e, updating **x** while **d** is fixed, and vice-versa.

### 3.1 The x-subproblem

By fixing the variable **d**, the subproblem w.r.t. **x** is a quadratic optimization problem and the solution to **x** is

$$\mathbf{x} = \left(\mathbf{1}+\delta\left(\mathbf{D}_h^T\mathbf{D}_h+\mathbf{D}_v^T\mathbf{D}_v\right)\right)^{-1}\left(\mathbf{y}+\delta\left(\mathbf{D}_h^T(\mathbf{d}_h-\mathbf{q}_h)+\mathbf{D}_v^T(\mathbf{d}_v-\mathbf{q}_v)\right)\right), \tag{10}$$

where **1** means the delta function, and the inversion operation can be efficiently computed in the Fourier domain. Assuming circular boundary conditions, we can apply 2D fast Fourier transform (FFT) which diagonalizes the derivative operators, and the close-form solution of **x** is

$$\mathbf{x} = \mathcal{F}^{-1}\frac{\mathcal{F}(\mathbf{y})+\delta\mathcal{F}\left(\mathbf{D}_h^T(\mathbf{d}_h-\mathbf{q}_h)\right)+\mathcal{F}\left(\mathbf{D}_v^T(\mathbf{d}_v-\mathbf{q}_v)\right)}{\mathcal{F}\left(\mathbf{1}+\delta\left(\mathbf{D}_h^T\mathbf{D}_h+\mathbf{D}_v^T\mathbf{D}_v\right)\right)}, \tag{11}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote FFT and inverse FFT, respectively. The plus, subtract, multiplication and division are all component-wise.

### 3.2 The d-subproblem

The **d** subproblem can be formulated as,

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}} \frac{1}{2}\left\|\mathbf{d} - (\mathbf{D}\mathbf{x} + \mathbf{q})\right\|_2^2 + \mu/\delta \left\|\mathbf{d}\right\|_p^p, \tag{12}$$

which can be solved by GST [14],

$$\mathbf{d}_h = GST\left(\mathbf{D}_h\mathbf{x} + \mathbf{q}_h, \mu/\delta, p, J\right), \mathbf{d}_v = GST\left(\mathbf{D}_v\mathbf{x} + \mathbf{q}_v, \mu/\delta, p, J\right). \tag{13}$$

Once obtaining **x** and **d**, **q** can be updated as follows

$$\mathbf{q}^{(t+1)} = \mathbf{q}^{(t)} + \mathbf{D}\mathbf{x}^{(t+1)} - \mathbf{d}^{(t+1)}. \tag{14}$$

For the updating of penalty parameter $\delta$, we adopt the adaptive updating strategy proposed by Lin *et al*. [16] to accelerate the convergence speed,

$$\delta^{(t+1)} = \min\left(\delta_{\max}, \rho\delta^{(t)}\right), \tag{15}$$

where $\delta_{\max}$ is the upper bound of $\delta$, and the values of $\rho$ is defined as,

$$\rho = \begin{cases} \rho_0, & if\ \delta\left\|\mathbf{d}^{(t+1)} - \mathbf{d}^{(t)}\right\| \Big/ \left\|\mathbf{D}\mathbf{x}^{(t+1)}\right\| < \varepsilon \\ 1, & otherwise \end{cases}. \tag{16}$$

where $\rho_0 > 1$ is a positive constant.

**Algorithm 1.** ALM-Lp

1. **Input:** image $\mathbf{y}$, smoothing weight $\mu$, parameters $\delta_0, \delta_{\max}$
2. **Initialize:** $\mathbf{x} = \mathbf{y}$, $\mathbf{d}^{(0)}$, $\mathbf{q}^{(0)}$, $t$=0
3. **Precompute:** $\mathcal{F}(\mathbf{y}), \mathcal{F}(\mathbf{H}) = \mathcal{F}\left(\mathbf{I} + \delta\left(\mathbf{D}_h^T\mathbf{D}_h + \mathbf{D}_v^T\mathbf{D}_v\right)\right)$
4. **While** not converged
5. $\mathbf{s}_h = \mathbf{d}_h^{(t)} - \mathbf{q}_h^{(t)}, \mathbf{s}_v = \mathbf{d}_v^{(t)} - \mathbf{q}_v^{(t)}$
6. $\mathbf{x}^{(t+1)} = \mathcal{F}^{-1}\left(\left(\mathcal{F}(\mathbf{y}) + \delta\left(\mathcal{F}\left(\mathbf{D}_h^T\mathbf{s}_h + \mathbf{D}_v^T\mathbf{s}_v\right)\right)\right) \oslash \mathcal{F}(\mathbf{H})\right)$
7. $\mathbf{d}^{(t+1)} = GST\left(\mathbf{D}\mathbf{x}^{(t)} + \mathbf{q}^{(t)}, \mu/\delta, p, J\right)$
8. $\mathbf{q}^{(t+1)} = \mathbf{q}^{(t)} + \mathbf{D}\mathbf{x}^{(t+1)} - \mathbf{d}^{(t+1)}$
9. Update $\delta^{(t+1)}$
10. $t$=$t$+1
11. **End while**

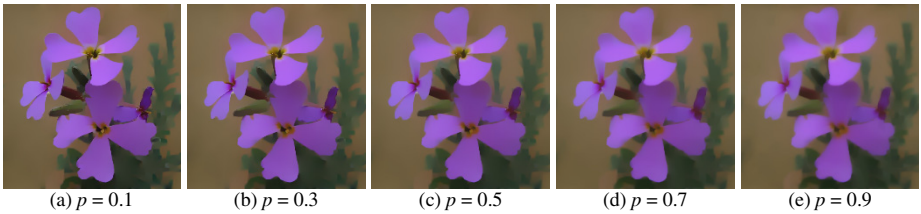
We summarize the proposed ALM-Lp algorithm in Algorithm 1. Since the Fourier transform of $\mathcal{F}(\mathbf{y}), \mathcal{F}\left(\mathbf{D}_h^T\right), \mathcal{F}\left(\mathbf{D}_v^T\right)$, and $\mathcal{F}\left(\mathbf{I}+\delta\left(\mathbf{D}_h^T\mathbf{D}_h+\mathbf{D}_v^T\mathbf{D}_v\right)\right)$ can be pre-computed, the proposed algorithm requires 2 FFT operations per iteration. ALM-Lp involves several parameters, we empirically fix $\delta_0 = 4\mu$, $\delta_{\max} = 10^5$ and $\varepsilon = 10^{-2}$ in our experiments.

## 4 Experimental Results

In this section, we evaluate the proposed ALM-Lp method. We first evaluate the smoothing results of ALM-Lp obtained using different $p$ values, and further compare ALM-Lp with several state-of-the-art smoothing methods, i.e., BLF, WLS, TV, and $L_0$ smoothing. Then, we apply ALM-Lp to cartoon artifacts removal and tongue image segmentation. The programs in our experiments are all coded in MATLAB and ran on a computer with Intel(R) Xeon(R) CPU E3-1230 V2@3.30GHz and 16GB memory.

### 4.1 ALM-Lp with Different $p$ Values

From Fig. 2, we can draw the similar conclusion with [7] that the gradient distributions of real-world images are well modeled by hyper-Laplacian prior, typically with $0.5 \le p \le 0.8$. Particularly, when $p = 0.7$, better tradeoff can be achieved in removing noise-like highlights and detailed textures and preserving salient edges. Thus we adopt $p = 0.7$ for ALM-Lp in the following experiments.



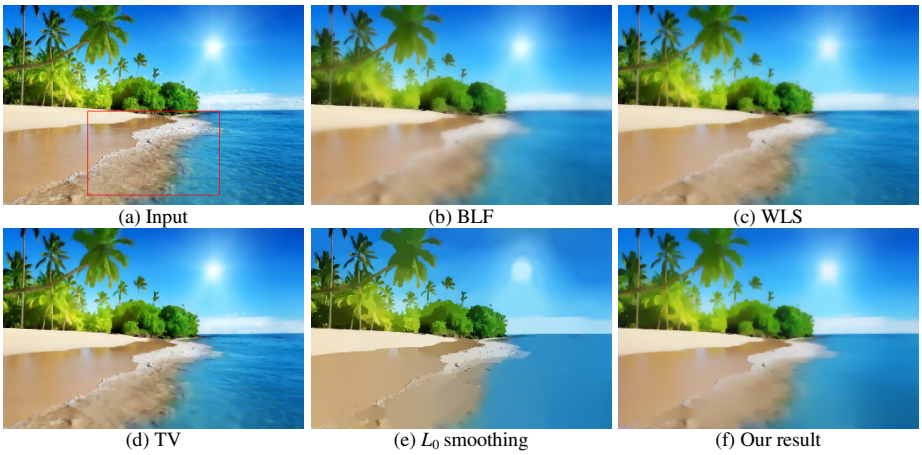
(a) $p = 0.1$ (b) $p = 0.3$ (c) $p = 0.5$ (d) $p = 0.7$ (e) $p = 0.9$

**Fig. 2.** The smoothing result of ALM-Lp with different $p$ values. This figure is best viewed in electronic form and zoomed.

### 4.2 Comparison with Other Methods

Using six natural images, we compare ALM-Lp with several state-of the art image smoothing methods, i.e., bilateral filter, weighted least-squares, TV, and $L_0$ smoothing. Fig. 3 and Fig. 4 show the results on images *beach* and *scenery*. On the whole, ALM-Lp can maintain the general structure and is effective in smoothing insignificant details, whether the gravel in Fig. 3 or the grass in Fig. 4.

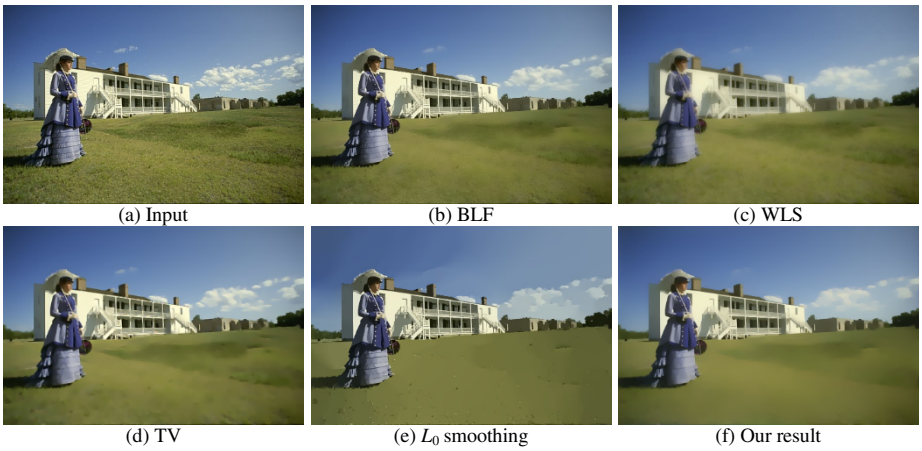
We further compare the running time of ALM-Lp with that of $L_0$ smoothing on the six natural images, as listed in Table 1. One can see ALM-Lp is more efficient than $L_0$ smoothing.



(a) Input (b) BLF (c) WLS

(d) TV (e) $L_0$ smoothing (f) Our result

**Fig. 3.** The smoothing results on the image *beach*. As the red boxes show, our method can better smoothing insignificant textures.



(a) Input (b) BLF (c) WLS

(d) TV (e) $L_0$ smoothing (f) Our result

**Fig. 4.** The smoothing results on the image *scenery*

**Table 1.** Speed (*sec.*) comparison of $L_0$ gradient smoothing and ALM-Lp method

| Image | Size | $L_0$ smoothing | ALM-Lp |
|---|---|---|---|
| *rock* | $800 \times 533$ | 4.19 | 1.58 |
| *flower* | $800 \times 533$ | 2.99 | 1.99 |
| *beach* | $800 \times 533$ | 3.02 | 1.70 |
| *pflower* | $475 \times 494$ | 1.45 | 0.97 |
| *scenery* | $481 \times 321$ | 1.11 | 0.54 |
| *basketball* | $270 \times 358$ | 0.72 | 0.31 |

### 4.3 Artifact Removal for Cartoon Pictures

Cartoon image compressed by conventional JPEG compression may contain some artifacts that severely damage the image structures. Thus, we need to remove artifacts while sharpening important salient edges. Fig. 5 shows the results obtained using ALM-Lp, from which we see the proposed method can effectively wipe off artifacts while preserving general structures, demonstrating its better property.


(a) Input (b) Our result (c) Input (d) Our result

**Fig. 5.** Cartoon pictures artifacts removal

### 4.4 Tongue Images Segmentation

For tongue body segmentation, the edge maps obtained by edge detectors, e.g., Canny, often contain unnecessary textures, shown as Fig. 6(c), which makes it hard to correctly segment the tongue body. Thus, we first used ALM-Lp to enhance the edge of tongue body, whereas diminishing unnecessary details, and then the well-known gradient vector flow (GVF) snake [17] was adopted to perform the segmentation. From Fig. 6(c)(e), one can see that, wispy textures in the edge map of the original image make the snake curve converge to a wrong segmentation result, while the smoothed image using ALM-Lp has a remarkable improvement, in which unnecessary textures are diminished, leading to the satisfactory segmentation results shown as Fig. 6(d)(f).

Furthermore, we validate the segmentation performance on three more tongue images, as shown in Fig. 7, from which we can draw the conclusion that it makes a significant segmentation improvement to apply ALM-Lp as a preprocessing step.

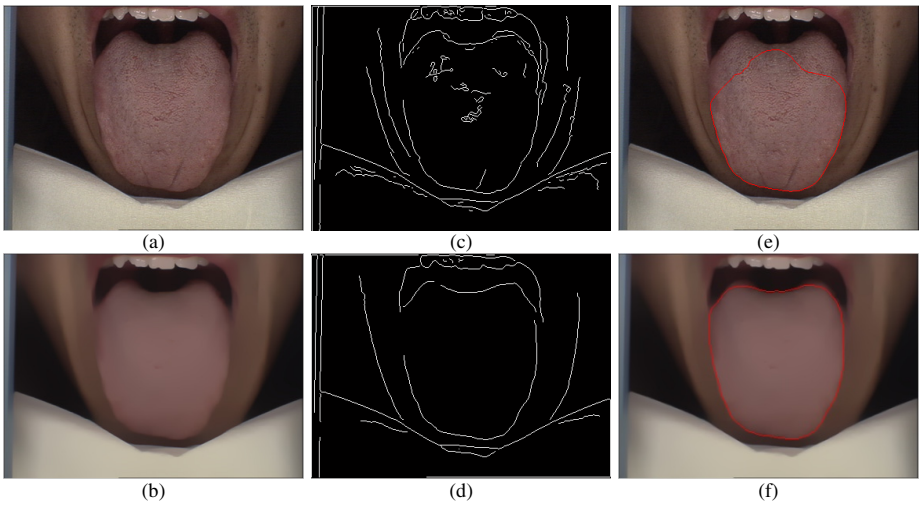



**Fig. 6.** Tongue image segmentation. (a) the original tongue image, (b) the result using ALM-Lp, (c), (d) are edge maps, and (e), (f) are segmentation results, respectively.

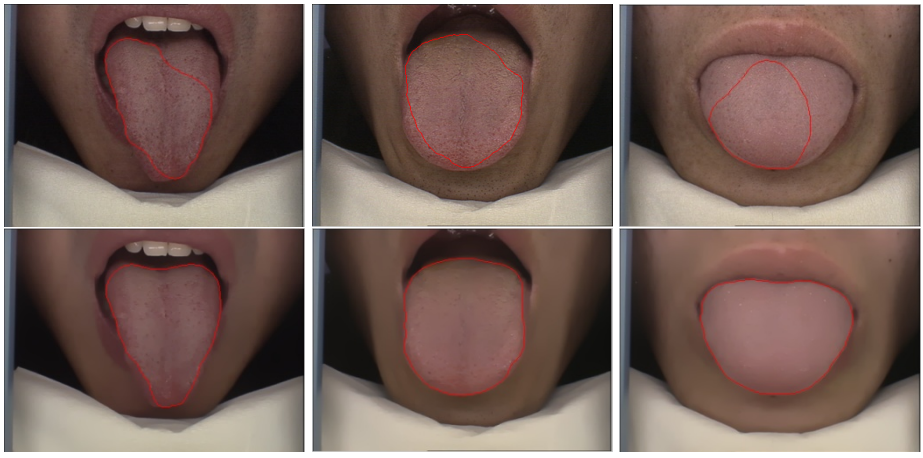


**Fig. 7.** More tongue images segmentation results. The top line presents the results on original images, and the bottom line presents the results on smoothed ones obtained using ALM-Lp.

## 5 Conclusion

In this paper, we studied the image smoothing problem and proposed an ALM-Lp method based on hyper-Laplacian gradient prior. An efficient augmented Lagrangian method (ALM) is developed to solve the proposed model. For natural image smoothing, ALM-Lp can obtain better smoothing results while compared with the state-of-the-arts. That is to say, ALM-Lp is effective in removing impulse noise-like highlights and detailed textures and preserving salient edges. Compared with $L_0$ smoothing, ALM-Lp is

more efficient. Moreover, ALM-Lp can also be applied to cartoon artifacts removal and tongue image segmentation. Since impulse noise-like reflection and highlights usually are inevitable in tongue images, ALM-Lp, as a preprocessing step, can significantly improve tongue image segmentation results.

## References


1. Park, J.M., Murphey, Y.L.: Edge detection in grayscale, color, and range images. Wiley Encyclopedia of Computer Science and Engineering (2008)
2. Carreira, J., Sminchisescu, C.: Constrained parametric min-cuts for automatic objects egmentation. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 3241–3248 (2010)
3. Durand, F., Dorsey, J.: Fast bilateral filtering for the display of high-dynamic-rangeimages. ACM Transactions on Graphics (TOG) 21, 257–266 (2002)
4. Farbman, Z., Fattal, R., Lischinski, D., Szeliski, R.: Edge-preserving decompositions for multi-scale tone and detail manipulation. ACM Transactions on Graphics (TOG) 27, 67 (2008)
5. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60(1), 259–268 (1992)
6. Xu, L., Lu, C., Xu, Y., Jia, J.: Image smoothing via L0 gradient minimization. ACM Transactions on Graphics (TOG) 30(6), 174 (2011)
7. Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-laplacian priors. NIPS 22, 1–9 (2009)
8. Field, D.J.: What is the goal of sensory coding? Neural Computation 6(4), 559–601 (1994)
9. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional-camera with a coded aperture. ACM Transactions on Graphics (TOG) 26(3), 70 (2007)
10. Simoncelli, E.P., Adelson, E.H.: Noise removal via bayesian wavelet coring. In: International Conference on Image Processing, vol. 1, pp. 379–382 (1996)
11. Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: IEEE international conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3869–3872 (2008)
12. Cho, T.S., Joshi, N., Zitnick, C.L., Kang, S.B., Szeliski, R., Freeman, W.T.: A content-aware image prior. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 169–176 (2010)
13. Candes, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted l1 minimization. Journal of Fourier Analysis and Applications 14(5-6), 877–905 (2008)
14. Zuo, W., Meng, D., Zhang, L., Feng, X., Zhang, D.: A generalized iterated shrinkage algorithm for non-convex sparse coding. In: IEEE International Conference on Computer Vision (ICCV) (2013)
15. She, Y.: An iterative algorithm for fitting non-convex penalized generalized linear models with grouped predictors. Computational Statistics & Data Analysis 56(10), 2976–2990 (2012)
16. Lin, Z., Liu, R., Su, Z.: Linearized alternating direction method with adaptive penalty for low-rank representation. NIPS 2, 6 (2011)
17. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE Trans. Image Processing 7(3), 359–369 (1998)