SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

March 2016, Vol. 59 032112:1–032112:7
doi: 10.1007/s11432-015-5375-x

Joint sparsity and fidelity regularization for segmentation-driven CT image preprocessing

Feng LIU[1,2] & Huibin LI[1,2*]

[1]*School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an* 710049*, China;*
[2]*Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing* 100037*, China*

Received March 31, 2015; accepted May 6, 2015; published online January 20, 2016

Abstract In this paper, we propose a novel segmentation-driven computed tomography (CT) image preprocessing approach. The proposed approach, namely, joint sparsity and fidelity regularization (JSFR) model can be regarded as a generalized total variation (TV) denoising model or a generalized sparse representation denoising model by adding an additional gradient fidelity regularizer and a stronger gradient sparsity regularizer. Thus, JSFR model consists of three terms: intensity fidelity term, gradient fidelity term, and gradient sparsity term. The interactions and counterbalance of these terms make JSFR model has the ability to reduce intensity inhomogeneities and improve edge ambiguities of a given image. Experimental results carried out on the real dental cone-beam CT data demonstrate the effectiveness and usefulness of JSFR model for CT image intensity homogenization, edge enhancement, as well as tissue segmentation.

Keywords CT image, homogenization, enhancement, tissue segmentation, gradient fidelity, gradient sparsity

Citation Liu F, Li H B. Joint sparsity and fidelity regularization for segmentation-driven CT image preprocessing. Sci China Inf Sci, 2016, 59(3): 032112, doi: 10.1007/s11432-015-5375-x

1 Introduction

Segmentation is an important tool in both medical imaging and medical image analysis, and it has been useful for many applications [1]. However, it is a very challenging task mainly due to the large intensity non-uniformities within the same tissues or structures (i.e. intensity inhomogeneities) and high intensity similarities across different tissues or structures (i.e. boundary ambiguities) [1]. This is particularly true for CT images that suffer from data shortcomings, such as sampling artifacts, partial volume effects, noise and remained shading artifacts after metal artifact reduction. A reconstructed CT image usually contains very low contrast and fuzzy edges, and also has serious intensity non-uniformity.

A possible way to reduce the aforementioned difficulties is performing image preprocessing techniques before the desired segmentation. The image preprocessing techniques usually associate to image denoising and enhancement. Methods like nonlinear total variation [2], wavelet shrinkage [3], nonlinear anisotropic diffusion [4], and sparse representation [5] have been successfully used for denoising of both natural and medical images. Medical image enhancement approaches include histogram equalization, edge sharping [1] and wavelet-based enhancement [6], etc. These methods can efficiently reduce image noise and improve

* Corresponding author (email: huibinli@mail.xjtu.edu.cn)

image quality. However, denoising and enhancement have limited effects on overcoming the intensity non-uniformities and edge ambiguities of CT images.

For above reasons, we investigate the segmentation-driven CT image preprocessing, including the intensity homogenization and edge enhancement. Here, intensity homogenization aims to smooth homogeneous region as much as possible so that the intensities within homogeneous tissue to be as similar as possible. While edge enhancement aims to reduce the edge ambiguities as much as possible so that edges across tissues look as steep as possible. Thus, our preprocessing can be regarded as an over-smoother or an over-enhancer since it not only removes noise, but also may change true gray-values and lose some subtle structure information.

Inspired by the successful applications of sparse regularization [5] and TV regularization [2] for image preprocessing, we propose to model the behaviors of our over-smoother and over-enhancer with joint sparsity and fidelity regularization (JSFR). For sparsity regularization, we propose to use $l_p(0 \leqslant p \leqslant 1)$ norm of image gradient as a penalty of image structure sparsity. Moreover, along with the classical intensity fidelity term, we suggest to add an additional image gradient fidelity term. This leads to a flexible balance between intensity and gradient fitting errors. Thus, JSFR can be regarded as a generalized total variation (TV) model or a generalized sparse representation model. Benefitting from both models, we will show that JSFR model has good ability to remove intensity inhomogeneities and improve edge ambiguities, both of which are useful for the subsequent segmentation procedure.

2 Related work

2.1 TV regularization based models

Consider an observed image $u_0 = u + n$ defined on $\Omega \subset \mathbb{R}^2$, where u denotes the desired clear image of u_0, and n denotes various non-stationary noises. General regularization model to recover u is as follows:

$$\min_u ||u - u_0||_{L_2}^2 + \lambda J(u). \tag{1}$$

The first term (i.e. data term) implies the intensity fidelity, which forces the intensities of the recovered image and the observed image as close as to each other. The second term (i.e. regularization term) hints some kinds of roughness penalty that depends on the potential noises, and λ is a positive parameter used to balance the effects of these two terms. Perhaps the TV regularization term first introduced by Rudin et al. in [2] is the most popular choice of $J(u)$ in the wide field of natural and medical image deionising. This model assumes that the desired image u belongs to a special function space named bounded variation (BV) and thus this TV regularization model can be expressed as follows,

$$\min_u ||u - u_0||_{L_2}^2 + \lambda \int_\Omega |\nabla u|, \tag{2}$$

where $u \in BV(\Omega)$. As pointed in [2], the TV regularization term can recover the discontinuities of an image, thus edges can been well persevered and noise can be also effectively removed.

2.2 Sparsity regularization based models

Sparsity regularization based image denoising model assumes that the desired clear image u can be represented by a few atoms of a suitable over-completed dictionary [5]. Once the dictionary matrix is given, the denoising problem is transferred to find the optimal sparse representation coefficients, i.e.,

$$(l_0) \quad \min_\alpha \|u_0 - A\alpha\|_2^2 + \lambda \|\alpha\|_0, \tag{3}$$

where $\|\cdot\|_0$ denotes the l_0 norm, and $\|\alpha\|_0$ counts the number of non-zero components of α. $A \in \mathbb{R}^{m \times n} (m \ll n)$ represents the over-completed dictionary which can be learned from the prior information. Sparsity is one of the fundamental property of images. It is well known that the above l_0 optimization is

a non-convex and non-smooth problem and is NP hard. In practice, it is equivalent to solving the relaxed l_1 minimization problem [5]. More recently, people find that there will be more exciting and desirable properties when enhancing the sparsity from l_1 to l_p $(0 < p < 1)$. l_1 and l_p models can been expressed by the following single formula:

$$\min_{\alpha} ||u_0 - A\alpha||_2^2 + \lambda||\alpha||_p^p, \quad (0 < p \leqslant 1). \tag{4}$$

Although l_p $(p < 1)$ model is still a non-convex optimization problem, efficient algorithms are available [5].

3 The proposed method

3.1 Joint sparsity and fidelity regularization (JSFR) model

Following the above notations, the proposed joint sparsity and fidelity regularization (JSFR) model is defined as follows:

$$\min_{u} \|u - u_0\|_2^2 + \lambda_1\|\nabla u - \nabla u_0\|_2^2 + \lambda_2\|\nabla u\|_p^p. \tag{5}$$

In JSFR model, $\|\nabla u - \nabla u_0\|_2^2$ implies the gradient fidelity regularization term (structure fidelity), and $\|\nabla u\|_p^p$ $(0 \leqslant p \leqslant 1)$ hints the sparsity regularization term (structure sparsity). The regularization parameters $\lambda_1 \geqslant 0$ and $\lambda_2 > 0$ imply the importance weights of the later two terms. When $\lambda_1 = 0$, the JSFR model can be regarded as a generalized TV regularization model, with enhanced structure sparsity penalty from l_1 to l_p. Here, we define $\|\nabla u - \nabla u_0\|_2^2 = \|u_x - u_{0x}\|_2^2 + \|u_y - u_{0y}\|_2^2$, and $\|\nabla u||_p^p = ||u_x||_p^p + ||u_y||_p^p$.

3.2 Fidelity and sparsity analysis of JSFR model

Different from the TV model in Eq. (2), our JSFR model contains two terms for data fidelity, namely, the intensity fidelity and the gradient fidelity. The gradient fidelity term is designed particularly for handling the inhomogeneities of CT images. As mentioned in Section 1, CT images reconstructed by using classical reconstruction algorithms, e.g. FDK (Feldkamp, Davis and Kress), FBP (filtered backprojection), not only contain non-stationary noise, but also exist significant intensity inhomogeneities even in the regions of the same kind of tissue. We will show in our numerical experiments that the gradient fidelity term has a powerful effect to reduce these kinds of non-uniformity of intensity in the homogenous regions. In particular, the larger λ_1 means stronger intensity homogenization and weaker intensity fidelity guarantee. Thus, the gradient fidelity can be regarded as a regularization effect for the intensity fidelity term.

Given an image u with size of $m \times n$, reshaped it to a column vector of dimension $l = mn$. It is clear that there exist invertible derivative operators $D_x \in \mathbb{R}^{l\times l}$ and $D_y \in \mathbb{R}^{l\times l}$, such that $u_x = D_x u, u_y = D_y u$. When $\lambda_1 = 0$, JSFR model (5) can be expressed as

$$\min_{u} \frac{1}{2}\{\|u_0 - D_x^{-1}u_x\|_2^2 + \|u_0 - D_y^{-1}u_y\|_2^2\} + \lambda_2\{\|u_x\|_p^p + \|u_y\|_p^p\}. \tag{6}$$

Compared with the l_p sparse model (4), the last term imposes a sparsity penalty on u_x and u_y. Thus the proposed JSFR model implies the gradient sparsity of u.

3.3 Algorithm for JSFR model

For convenience, we define the image gradient operators u_x and u_y of image $u \in \mathbb{R}^{m\times n}$ as

$$u_x(i,j) = \begin{cases} u(i+1,j) - u(i,j), & 1 \leqslant i \leqslant m-1, \\ -u(m,j), & i = m, \end{cases} \quad u_y(i,j) = \begin{cases} u(i,j+1) - u(i,j), & 1 \leqslant j \leqslant n-1, \\ -u(i,n), & j = n, \end{cases} \tag{7}$$

and the derivative operators $D_x = \mathrm{diag}(d_1^x, d_2^x, \ldots, d_n^x) \in \mathbb{R}^{l\times l}$ and $D_y \in \mathbb{R}^{l\times l}$ of $u \in \mathbb{R}^{l\times 1}$ as

$$d_k^x(i,j) \in \mathbb{R}^{m\times m} = \begin{cases} -1, & j = i, 1 \leqslant i \leqslant m, \\ 1, & j = i+1, 1 \leqslant i \leqslant m-1, \\ 0, & \text{otherwise}, \end{cases} \quad D_y(i,j) = \begin{cases} -1, & j = i, 1 \leqslant i \leqslant l, \\ 1, & j = i+m, 1 \leqslant i \leqslant l-m, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

for all $1 \leqslant k \leqslant n$. Following the Iterative Reweighed Least Squares (IRLS) algorithm [5], let $W_1 = \mathrm{diag}(1/|u_x^1|^{2-p}, 1/|u_x^2|^{2-p}, \ldots, 1/|u_x^l|^{2-p})$, and $W_2 = \mathrm{diag}(1/|u_y^1|^{2-p}, 1/|u_y^2|^{2-p}, \ldots, 1/|u_y^l|^{2-p})$. Here u_x^i and u_y^i denote the i-th component of $u_x \in \mathbb{R}^{l\times 1}$ and $u_y \in \mathbb{R}^{l\times 1}$, respectively. Then the JFSR model in Eq. (5) can be expressed as

$$\min_u \|u-u_0\|_2^2 + \lambda_1(\|D_x u - D_x u_0\|_2^2 + \|D_y u - D_y u_0\|_2^2) + \lambda_2(u^\top D_x^\top W_1 D_x u + u^\top D_y^\top W_2 D_y u). \quad (9)$$

Let $B = I + \lambda_1(D_x^\top D_x + D_y^\top D_y)$, $C = D_x^\top W_1 D_x + D_y^\top W_2 D_y$, then solving (9) is equivalent to solving

$$\min_u \frac{1}{2} u^\top (B + \lambda_2 C)u - u^\top B u_0. \quad (10)$$

Please refer to Algorithm 1 for the details of JSFR algorithm.

Algorithm 1 Joint sparsity and fidelity regularization (JSFR)

Require: u_0, λ_1, λ_2, tolerance ε, and the number of maximum iterations K;
Ensure: u;
1: Initialization: $u_0 = u_0/\max(|u_0|)$, $k = 1$ and $u^k \in \mathbb{R}^{l\times 1}$ be an arbitrary vector;
2: Compute u_x^k, u_y^k, W_1^k and W_2^k by $D_x u^k$ and $D_y u^k$;
3: Compute u^{k+1} by solving (10);
4: **if** $k \leqslant K$ or $\|u^{k+1} - u^k\| \geqslant \varepsilon$ **then**
5: $k \Leftarrow k+1$, go to step 2;
6: **else**
7: Stop;
8: **end if**

4 Experimental results

4.1 Dental cone-beam CT images

In our experiments, we employed 256 slices of real dental cone-beam CT images with size of 512×512 reconstructed by using the FDK algorithm. For the sake of display, all slices were down-sampled to 256×256. As shown in the first row of Figure 1 and at the top of Figure 2, column (a), the CT images contain non-stationary noise, and exist significant non-uniformity of intensity (i.e. inhomogeneities) within the soft tissue region in particular near the teeth. Moreover, the contrast ratios among some tissues are also very low (e.g. the cervical vertebra region and its nearby tissues).

4.2 Parameter setting

As mentioned in Section 3, λ_1 and λ_2 in Eq. (5) are used to balance the importance weights of the fidelity and the sparsity regularization terms. The basic rule to choice appropriate values largely depends on the intensity inhomogeneities and the contrast ratios among tissues. Generally, larger λ_1 means stronger intensity homogenization and larger λ_2 hints stronger smoothing effect. According to our numerical experiments, the effectiveness of homogenization by using JSFR model is not sensitive to λ_1, i.e., we can always find a satisfied solution for any λ_1 belonging to a certain interval. Thus we can specify λ_1 in advance. In our experiments, we take λ_1 to be greater than or equal to 1 based on the edge contrast ratios between the tooth and cervical vertebra. What we really need to do is to tune λ_2. It depends on the signal-to-noise ratio and λ_1. The parameter p is related to gradient sparsity. Smaller p hints stronger gradient sparsity and steeper edge enhancement. In all our experiments, p is taken to be 0.2. The other fixed values are the tolerance ($\varepsilon = 10^{-7}$) and the number of maximum iterations ($K = 20$).

4.3 Effectiveness for homogenization and enhancement

In this experiment, we compare the effect of the proposed JSFR model to TV model for CT image homogenization and enhancement. Three original and preprocessed slices of dental CT images under

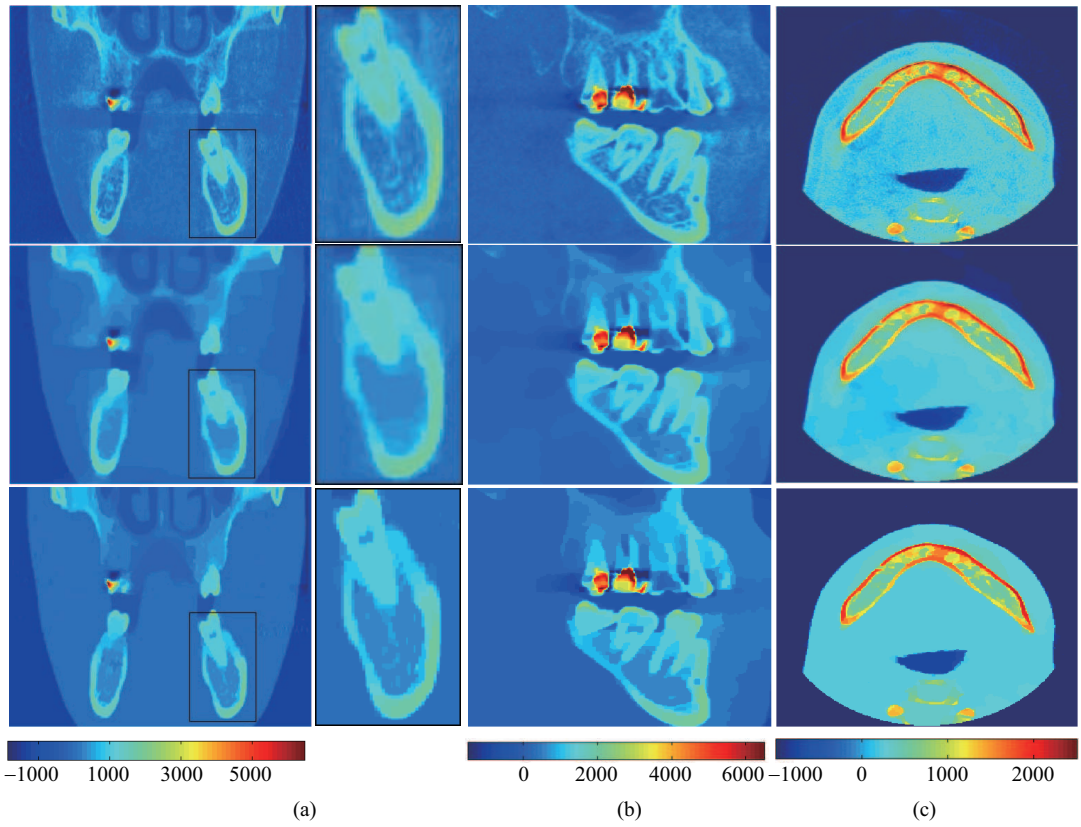

Figure 1 The pseudo-color illustration of the homogenized and enhanced CT images: from top row to bottom, raw CT images, results from TV model (λ: 0.05, 0.015 and 0.06) and JSFR model (λ_1: 100, 1 and 1; λ_2: 0.01, 0.0005 and 0.01). (a) Coronal ($x = 300$); (b) sagital ($y = 177$); (c) axial ($z = 50$).

different views (i.e. coronal, sagital, axial) are displayed in Figure 1. As shown in this figure, the edges in the coronal slice has higher contrast, whereas teeth root edges in the sagital slice and vertebra edges in axial slice has lower contrast. Thus we take λ_1 to be 100 for the coronal slice and 1 for the other slices. We observed that compared with TV model, JSFR model performs better to eliminate the intensity inhomogeneities of regions with same tissues (e.g. the soft tissue), and enhance the fussy edges and preserve the fine structure (e.g. the zoomed cone shape structure of tooth root in the right hand of Figure 1(a)).

4.4 Usefulness for tissue segmentation

In this experiment, we show that both intensity homogenization and edge enhancement performed by JSFR model can simplify the subsequent tissue segmentation procedures and achieve better segmentation results than TV model. As shown at the top left of Figure 2, given an axial section ($z = 154$) of the real dental cone-beam CT data, we want to extract the teeth and the cervical vertebra. This is a very challenging task since the CT image contains metal artifacts, inhomogeneous intensity, and ambiguous vertebra edges. Thus we first manually specify two rectangular regions of interest (ROIs) on this test CT slice. They are $\mathrm{ROI}_{\mathrm{teeth}} = u(1:150, 1:256)$ and $\mathrm{ROI}_{\mathrm{vertebra}} = u(192:256, 1:256)$. Then, we applied the TV model and JSFR model to $\mathrm{ROI}_{\mathrm{teeth}}$ and $\mathrm{ROI}_{\mathrm{vertebra}}$, respectively. λ in TV model is set to be 0.15 and 0.055 for $\mathrm{ROI}_{\mathrm{teeth}}$ and $\mathrm{ROI}_{\mathrm{vertebra}}$, respectively. For JSFR model, λ_1 is set to be 10 for $\mathrm{ROI}_{\mathrm{teeth}}$ due to the high contrast edges of teeth, whereas λ_1 is set to be 1 for $\mathrm{ROI}_{\mathrm{vertebra}}$ due to the low contrast vertebra. λ_2 is taken to be 0.035 and 0.055 for $\mathrm{ROI}_{\mathrm{teeth}}$ and $\mathrm{ROI}_{\mathrm{vertebra}}$, respectively. The homogenized and enhanced ROIs are shown in Figure 2 column (b). Finally, simple threshold method is used to extract the teeth and cervical vertebra from the homogenized and enhanced ROIs. The segmented results are shown in Figure 2 column (c). Here, the thresholds are selected according to the intensity histograms of

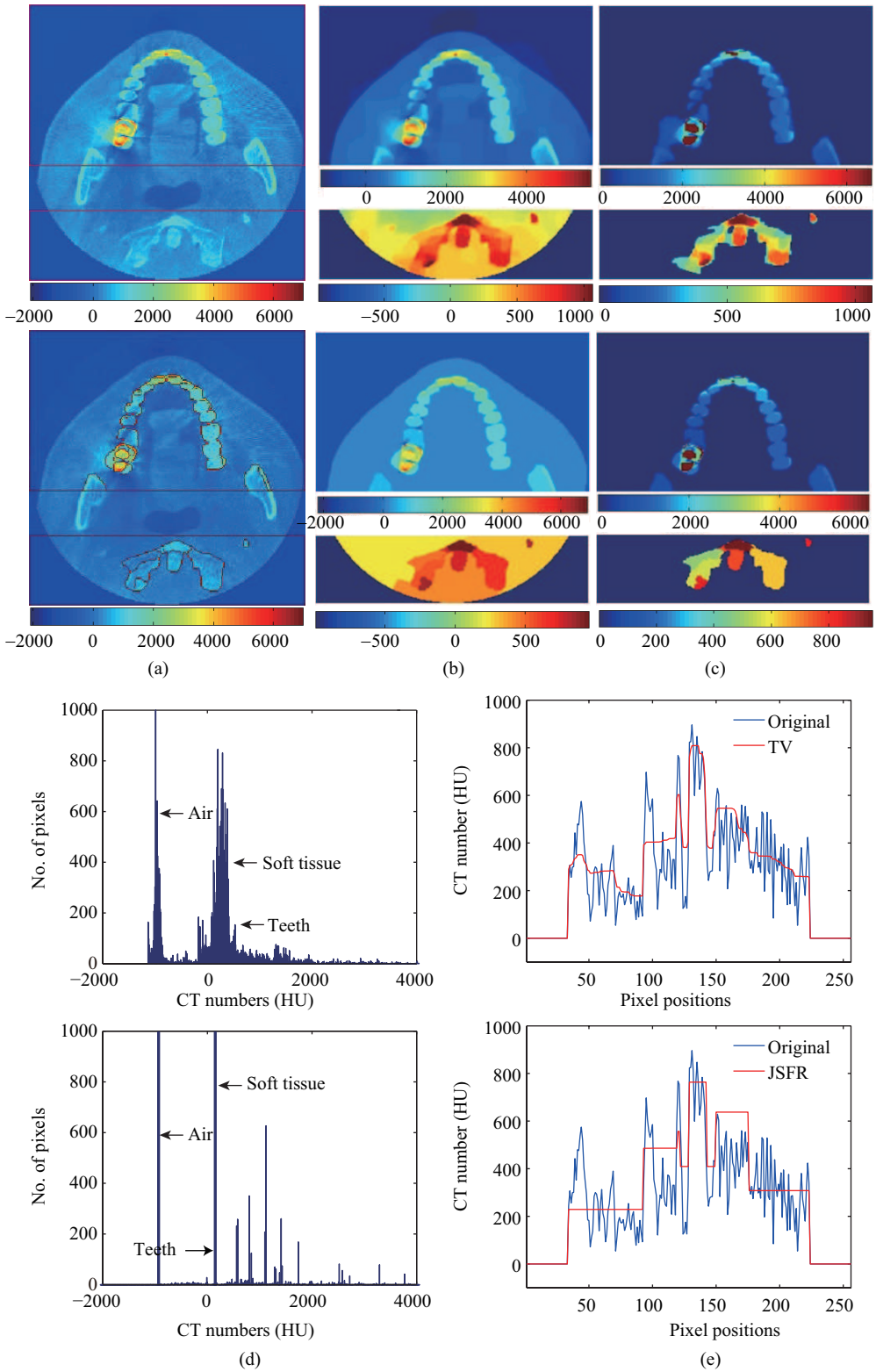

Figure 2 Illustration of TV model and JSFR model for CT image homogenization, enhancement, and tissue segmentation. From top to bottom: column (a), a real dental cone-beam CT image and its two manually cropped ROIs: $\mathrm{ROI}_{\mathrm{teeth}}$ and $\mathrm{ROI}_{\mathrm{vertebra}}$; the boundaries of ROIs extracted by JSFR; column (b), the homogenized and enhanced ROIs by TV and JSFR; column (c), the segmented ROIs by TV and JSFR; column (d), the intensity histograms of $\mathrm{ROI}_{\mathrm{teeth}}$ in column (b); column (e), intensity variations of the 21th row sampling of $\mathrm{ROI}_{\mathrm{vertebra}}$ before and after preprocessing by TV and JSFR.

the preprocessed ROIs as shown in Figure 2 column (d). In particular, thresholds are set to be 541 HU and 414 HU for $\mathrm{ROI}_{\mathrm{teeth}}$ and $\mathrm{ROI}_{\mathrm{vertebra}}$ preprocessed by TV model, and to be 500 HU and 480 HU for $\mathrm{ROI}_{\mathrm{teeth}}$ and $\mathrm{ROI}_{\mathrm{vertebra}}$ preprocessed by JSFR model. Furthermore, to extract the embedded metals, the threshold is set to be 3000 HU for $\mathrm{ROI}_{\mathrm{teeth}}$ preprocessed by both TV and JSFR. To better illustrate the different effects of intensity homogenization and edge enhancement between TV model and JSFR model, we also sample and plot the 21th row intensity of the preprocessed $\mathrm{ROI}_{\mathrm{vertebra}}$ in Figure 2(e).

From column (b) of Figure 2, we can see that the proposed JSFR model can be regarded as a stronger over-smoother and an over-enhancer than TV model. For example, even the strong metal artifacts around the leftmost wisdom tooth in $\mathrm{ROI}_{\mathrm{teeth}}$ are almost removed by JSFR model, and the extreme low contrast edges of the leftmost and rightmost cancellous bone in $\mathrm{ROI}_{\mathrm{vertebra}}$ are also significantly enhanced. The detail effects can be found in column (e) of Figure 2, from which we can find that the homogenized regions within the same kind of tissues by JSFR model look more flat, and the enhanced edges across different tissues by JSFR model look steeper (i.e. quasi-step edges). Column (d) of Figure 2 shows that our proposed JSFR model is more suitable for segmentation-driven preprocessing than TV model. In particular, the intensity histograms of soft tissue and teeth are still overlapped together after the preprocessing of TV model, while they can be largely separated by JSFR model. Thus, based on the preprocessed ROIs by JSFR model, the teeth and vertebra are perfectly segmented out by simple threshold method as shown in column (c) of Figure 2. It worth noticing that the metals embedded in the teeth are also perfectly extracted. This result will be very useful for metal artifacts reduction.

5 Conclusion

This paper focuses on segmentation-driven CT image preprocessing. To this end, we proposed a novel homogenization and enhancement model, namely JSFR. Experimental results demonstrated on the real dental cone-beam CT data show that: i) JSFR model can effectively homogenize intensities within the same tissue and enhance the ambiguous edges across different tissues; ii) JSFR model can be used as a useful preprocessing solution to extract tissues of interest from CT image. As shown in Subsection 4.4, simple threshold method can achieve perfect segmentation results benefitting from the proposed JSFR model. For the future work, we will study more detail rules to select the parameters λ_1, λ_2 and p for homogenization and enhancement, and assign adaptive thresholds for segmentation. Moreover, we will test our model and algorithm for 3D dental reconstruction [7] and to other real CT data.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 11401464) and the China Postdoctoral Science Foundation (Grant No. 2014M560785).

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Bankman I N. Handbook of Medical Image Processing and Analysis. 2nd ed. Burlington: Academic Press, 2009
2 Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Phys D, 1992, 60: 259–268
3 Donoho D L. De-noising by soft-thresholding. IEEE Trans Inf Theory, 1995, 41: 613–627
4 Kim H S, Yoon H S, Trung K N, et al. Automatic lung segmentation in CT images using anisotropic diffusion and morphology operation. In: Proceedings of the 7th IEEE International Conference on Computer and Information Technology, Dhaka, 2007. 557–561
5 Elad M. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. New York: Springer, 2010
6 Hai N T. Wavelet-based image fusion for enhancement of ROI in CT image. J Biomed Eng Med Imag, 2014, 1: 1–13
7 Naumovich S S, Naumovich S A, Goncharenko V G. Three-dimensional reconstruction of teeth and jaws based on segmentation of CT images using watershed transformation. Dentomaxillofac Radiol, 2015, 44: 20140313