Research Article

Weighted Nuclear Norm Minimization Based Tongue Specular Reflection Removal

Zhenchao Cui, Dongyu Zhang, Kuanquan Wang, Hongzhi Zhang, Naimin Li, and Wangmeng Zuo

School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Wangmeng Zuo; cswmzuo@gmail.com

In computational tongue diagnosis, specular reflection is generally inevitable in tongue image acquisition, which has adverse impact on the feature extraction and tends to degrade the diagnosis performance. In this paper, we proposed a two-stage (i.e., the detection and inpainting pipeline) approach to address this issue: (i) by considering both highlight reflection and subreflection areas, a superpixel-based segmentation method was adopted for the detection of the specular reflection areas; (ii) by extending the weighted nuclear norm minimization (WNNM) model, a nonlocal inpainting method is proposed for specular reflection removal. Experimental results on synthetic and real images show that the proposed method is accurate in detecting the specular reflection areas and is effective in restoring tongue image with more natural texture information of tongue body.

1. Introduction

In traditional Chinese medicine (TCM), the practitioners observe the color, shape, texture, and coating characteristics of tongue to evaluate the healthy condition of a person. Because of its convenience and effectiveness, for thousands of years, traditional tongue diagnosis has been very popular in the countries of East Asia, especially in China, Korea, Japan, and so forth [1–3].

Traditional tongue diagnosis, however, is a skill that requires years of training to master, and the diagnosis result is highly dependent on the practitioners' personal experience. That is, for a specific patient, the diagnosis results given by several practitioners may be distinctly different. In summary, the limitations of traditional tongue diagnosis greatly restrict its applications in modern medicine.

Recently, with the progress in image processing and pattern recognition, computational tongue diagnosis has received considerable interests to make tongue diagnosis more objective and automated. Many tongue diagnosis systems have been developed, as hyperspectral-based tongue diagnosis system [4], 3D tongue acquisition system [5], and so forth [6–10]. Different features of tongue, for example,

shape [11], color [8, 12, 13], texture, coating [14, 15], and sublingual vein [16], have been suggested for the representation of tongue images. Many methods, such as Bayesian network [7], support vector machine [17], decision trees, and naive Bayes [18], have been proposed for tongue image analysis.

However, because of the saliva on the tongue body, specular reflection generally is inevitable for the existing tongue image acquisition systems [19]. Figure 1 shows one typical semiclosed system for tongue image acquisition [10] and several tongue images with specular reflections. The specular reflection on the tongue image would adversely affect many tongue texture analysis results, for example, tongue fur, tongue texture, and tongue color detection. To alleviate the effect of specular reflection, one may develop some approaches to improve the robustness of the existing feature extraction and classification methods, but the natural strategy is to detect and repair the specular reflection areas (i.e., the detection and inpainting pipeline).

By far, several methods have been proposed for the detection and repairing of the specular reflection areas in tongue images. These methods, however, are limited in both detection and inpainting. In the detection stage, as shown in Figure 2, a typical specular reflection region in
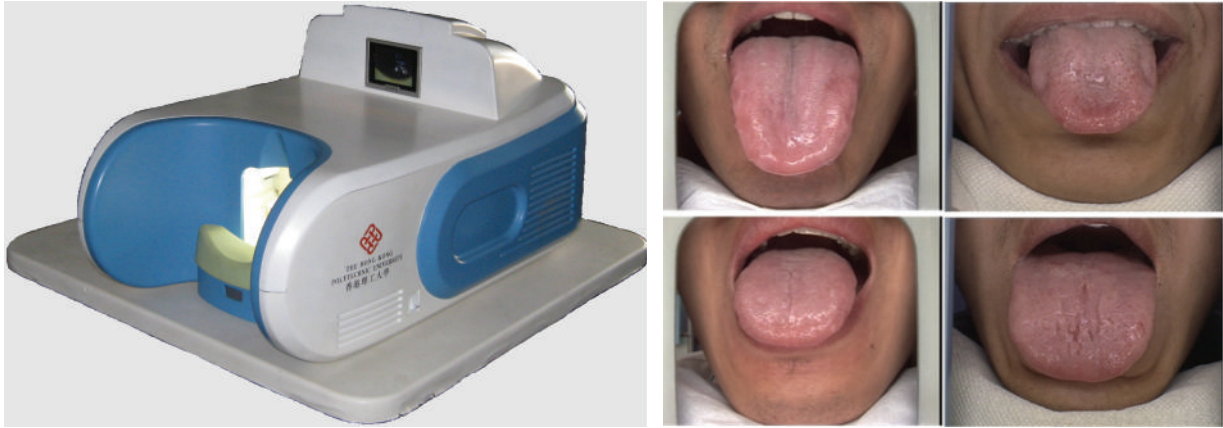
FIGURE 1: Tongue image acquirement device and typical tongue images with specular reflections.

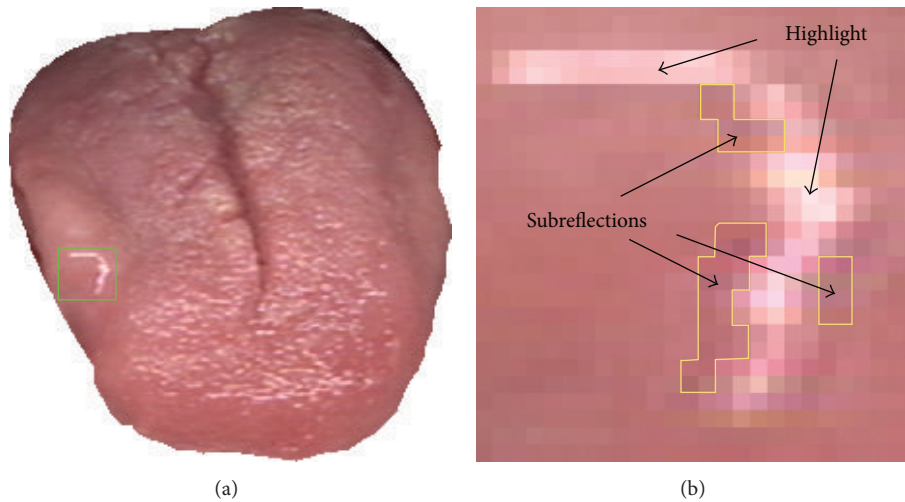(a)   (b)

FIGURE 2: A tongue image with specular reflection area: (a) one of the highlight areas is marked by the green box and (b) the enlarged version of the subimage in the green box. It is observed that the pixels along the highlight area are usually darker than the usual pixels.

tongue image includes both the highlight reflections and the subreflections, while the existing detection methods usually only considered the highlight reflections [16, 20, 21] or just adopted some trivial strategies (e.g., morphological operators) to cope with the subreflections [21]. In the repairing stage, bilinear interpolation [20] and total-variation- (TV-) based methods [21] were applied for inpainting the tongue specular reflection regions. These methods, however, only considered the local smoothness of the image and were not effective when the reflection areas are large.

In this paper, we adopted the detection and inpainting pipeline and proposed a novel method for the removal of specular reflection areas in tongue image. In the detection stage, adaptive thresholding and superpixel segmentation methods were applied for the highlight reflections and subreflections detection. Referring to the location of the initial highlight reflections obtained via thresholding, subreflections were defined as the surrounding superpixels with lower illumination than that of normal pixels of tongue body. The highlight reflections together with the subreflections were

regarded as the final detection result of specular reflection areas.

In the inpainting stage, based on the homogenous property of tongue image, we proposed a nonlocal inpainting approach, that is, weighted nuclear norm minimization- (WNNM-) based tongue specular reflections removal. For a single small patch of a tongue image, there were many nonlocal similar patches across the whole image. If the patch and its nonlocal similar ones are stacked into a matrix, it was reasonable to assume that the matrix should be of low rank. Thus by solving a low rank matrix completion problem with WNNM framework, we could fill the specular reflection areas with the texture information of tongue.

This paper focused on the tongue image specular reflections removal, whose contributions are of twofold: (i) detection of highlight reflection and subreflection areas, and (ii) WNNM-based inpainting. First, we analyzed the adverse impact of subreflections, which to the best of our knowledge receive little attention in tongue specular reflection removal. To address this, we proposed a superpixel based specular reflection detection method which can effectively detect

the highlight reflection together with subreflection areas, while affecting little the normal pixels around the subreflection areas. Second, we proposed a WNNM-based nonlocal inpainting method for the removal of specular reflections. The proposed method converts the inpainting problem to an unconstrained optimization problem which could be iteratively solved by using standard WNNM. Compared with other inpainting approaches, our method can obtain more natural textures information of tongue, especially for large reflection area, which could promote both the PSNR value and the visual effect of the restored tongue image.

The remainder of the paper is organized as follows: Section 2 describes the superpixel based specular reflections detection method. Section 3 presents the WNNM-based nonlocal inpainting method for specular reflection removal of tongue image. The experimental results are provided in Section 4, and finally Section 5 concludes this paper.

2. Tongue Image Specular Reflection Detection

In this section, we firstly described the characteristics of specular reflections of tongue image and then proposed our superpixel-based specular reflections detection method.

Figure 2(a) shows a typical example of tongue image with the specular reflection areas, while its enlarged subimage is shown in Figure 2(b). As shown in Figure 2(b), the pixels of highlight reflection areas always have higher illumination and less saturation values than other pixels, while the subreflections consist of many abnormal dark pixels with lower illumination around the highlight reflection areas. In [21], Jia et al. proposed to use morphological dilation to detect the subreflection pixels around the highlight reflection area. However, for large area of highlight reflection, because the surrounding abnormal pixels of subreflection area are not uniform, the morphological dilation often ended with inaccurate detection result by including much more normal pixels of tongue body, as shown in Figures 5(c) and 5(d).

To solve this problem, we proposed a two-stage method for the detection of specular reflection areas. First, utilizing hue-saturation-illumination (HSI) color model, an initial detection result was obtained for highlight reflection area. Then, the subreflection areas were further detected via superpixel segmentation.

2.1. Detection of Highlight Reflection Area.
As for our observation, the highlight reflection areas of tongue image often have higher illumination values and less saturation values. Thus, it is natural to transform the tongue image into the HSI color space, and thus the hue and saturation were adopted to describe color. The following equations were adopted to convert the tongue image into the HSI color space:

$$I = \frac{1}{3}(R + G + B),$$

$$H = \begin{cases} \theta, & G > B, \\ 2\pi - \theta, & G \leq B, \end{cases}$$

$$S = 1 - \frac{3 \times \min(R, G, B)}{(R + G + B)},$$

(1)

where $\theta = \cos^{-1}\{(1/2)[(R - G) + (R - B)]/\sqrt{(R-G)^2 + (R-B)(G-B)}\}$, R, G, and B are the three channels of RGB color space, respectively.

Figure 3(a) shows a typical tongue image with specular reflection, while Figures 3(b) and 3(c) show the I channel and the S channel of Figure 3(a) in the HSI color space, respectively. For each channel, a threshold was used to check whether a pixel should belong to the highlight reflection area or not. The thresholds T_I for I channel and T_s for the S channel were adaptively obtained by

$$T_I = i_{\max} + \frac{(i_{\max} - i_{\mean})^2}{(i_{\max} - i_{\min})},$$

$$T_s = s_{\min} + \frac{(s_{\mean} - s_{\min})^2}{(s_{\max} - s_{\min})},$$

(2)

where $i_{\max}$, $i_{\min}$, and $i_{\mean}$ denote, respectively, the maximum, minimum, and mean value of all illumination values in a tongue image, and $s_{\max}$, $s_{\min}$, and $s_{\mean}$ denote, respectively, the maximum, minimum, and mean value of all saturation values in the tongue image. Finally, the highlight reflection areas image HR could be obtained by using the following criterion:

$$\begin{aligned} & \text{HR}(x, y) \\ & := \{(x, y) \mid (I(x, y) > T_I) \cap (S(x, y) < T_s)\}, \end{aligned}$$

(3)

where $I(x, y)$ and $S(x, y)$ denote the I channel and S channel of pixel (x, y), respectively. Figure 3(d) shows an example of the detection result of highlight reflection areas in a typical tongue image.

2.2. Superpixel Based Detection of Subreflection Area.
For a tongue image, by applying oversegmentation, a large number of small regions, that is, superpixels, were obtained. Subreflections were defined as the superpixels that surround the highlight reflection and which had lower illumination values than that of normal pixels of tongue body. Utilizing the isotropic characteristics of superpixels [22, 23], we could naturally find the subreflections and avoid bringing in much normal pixels of tongue body.

Superpixels can be obtained by oversegmenting a tongue image with any reasonable existing segmentation algorithms. In this paper, a graph-based segmentation method [22] was adopted. For each tongue image, an undirected graph was defined with the pixels/regions as nodes, connected by the edges of its neighborhood. A nonnegative weight $w(e)$ was used to edge e to measure the dissimilarity of its corresponding nodes. In the beginning, each pixel was a node. The graph-based segmentation algorithm gradually merged similar regions/nodes to the same superpixel. The merging process was motivated by the internal variation int(R) of region R defined as

$$\text{int}(R) = \max_{e \in \text{MST}(R)} w(e) + \frac{k}{|R|},$$

(4)

where $|R|$ is the size of region R, MST(R) is the minimum spanning tree of R, and k is a nonnegative parameter related
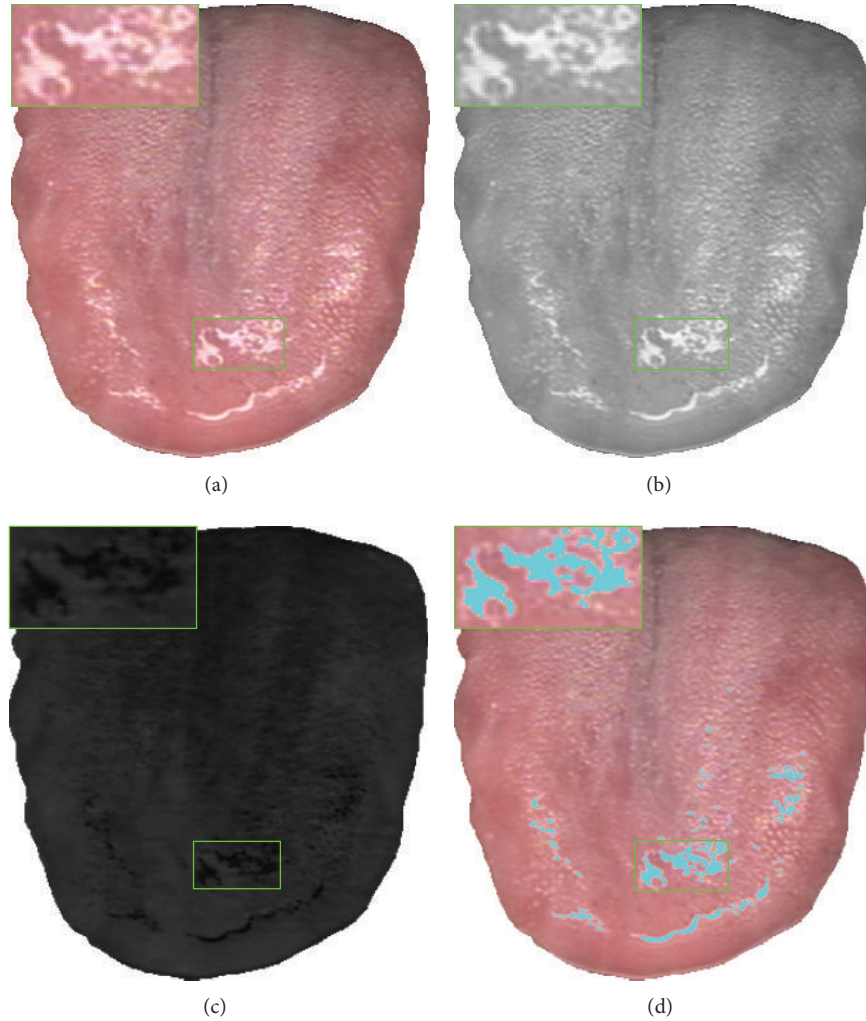
(a)

(b)

(c)

(d)

Figure 3: Tongue image in the HSI color space: (a) the typical tongue image, (b) I channel of tongue image, (c) S channel of tongue image, and (d) the initial detection of reflection.

to the size or the numbers of superpixels. For any two different regions R_1 and R_2, there will be at least one edge between them, with its weight higher than $\text{Int}(R_1)$ and $\text{Int}(R_2)$ (if there is no edge between the two regions, the edge weight is regarded as $+\infty$). Otherwise, the two regions were merged into one new region, and the internal variation of this compound region was updated. Finally, for a tongue image, we could obtain a series of superpixels. For the detail of this algorithm, please refer to [22].

After oversegmentation, we could obtain many super-pixels of tongue image which could be roughly grouped into three categories, highlight reflections superpixels, sub-reflection superpixels, and normal tongue body superpixels. As the subreflections were commonly around the highlight reflection, we could use the following strategy to locate them. First, as shown in Figure 4, with the initial highlight reflection detected in Section 2.1, we could obtain an interest area, a rectangle box with green border, and a circle band with red dash line border around each highlight reflection using

morphological dilation operation. Then, for each superpixel in the circle band, we determinate its category with

$$
\text{superpixel} \in \begin{cases} \text{Sub-reflection} & m_{\text{sp}} < m_{\text{tb}} \times \gamma_{\text{min}}, \\ \text{Highlight reflection} & m_{\text{sp}} \geq m_{\text{tb}} \times \gamma_{\text{max}}, \\ \text{Tongue body} & \text{otherwise}, \end{cases} \tag{5}
$$

where m_{sp} means pixels value of single superpixel in the band, m_{tb} is the mean value of normal pixels of tongue body in the interest area, and γ_{min} and γ_{max} are two nonnegative parameters which were empirically set to 1 and 1.15 in this paper. The width of the circle band was adaptively adjusted according to the size of highlight reflection, while for large highlight reflection the width of corresponding circle was wider. By classifying all the candidate superpixels in the circle
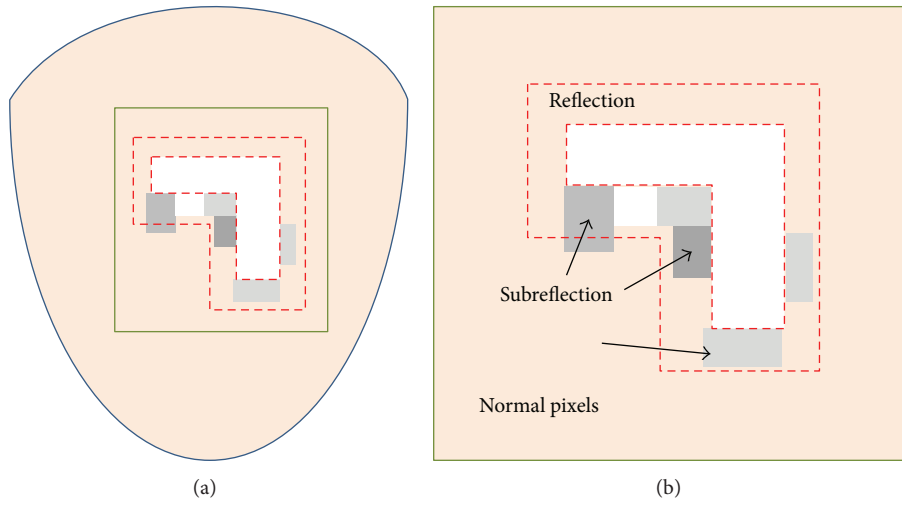
FIGURE 4: Sketch map of subreflection detection: (a) tongue image with specular reflections where green box was the select interest area, and (b) the enlarged interest area in (a) where the candidate superpixels were searched in the circle band with red dash line border.
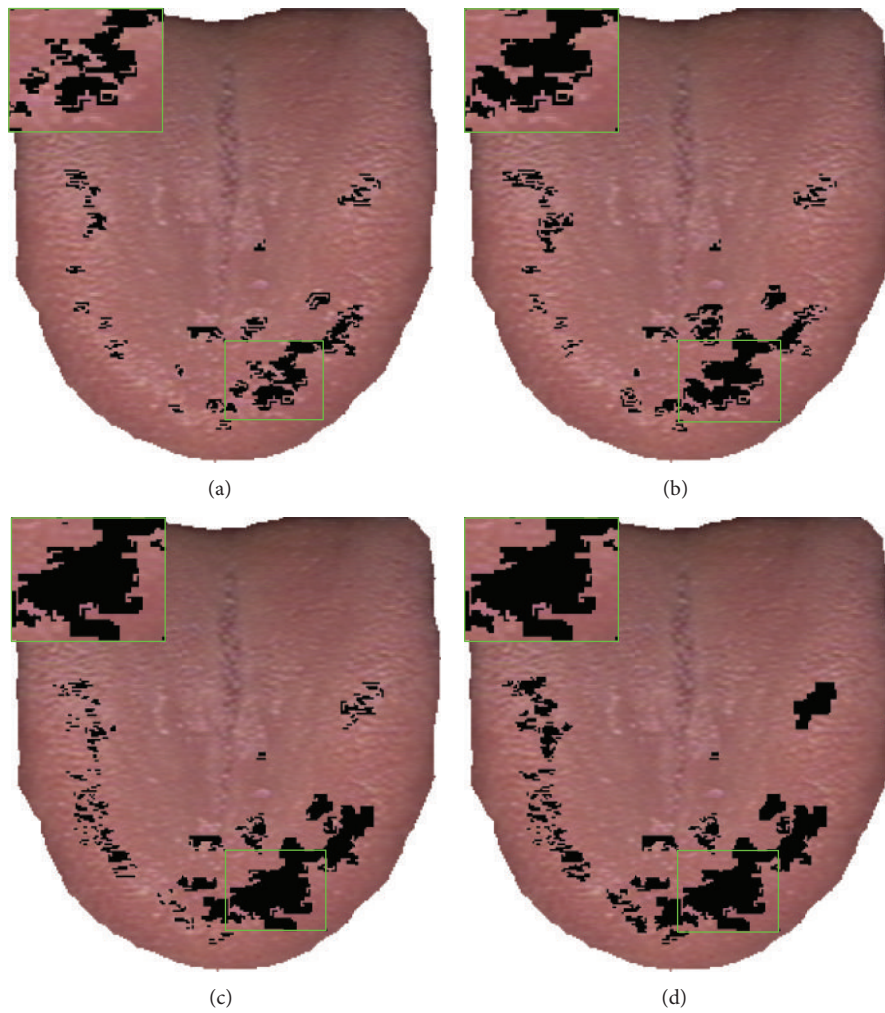

FIGURE 5: The detection results of specular reflections by using different values of k, (a) $k = 0.01$, (b) $k = 0.05$, (c) $k = 0.10$, and (d) $k = 0.15$.
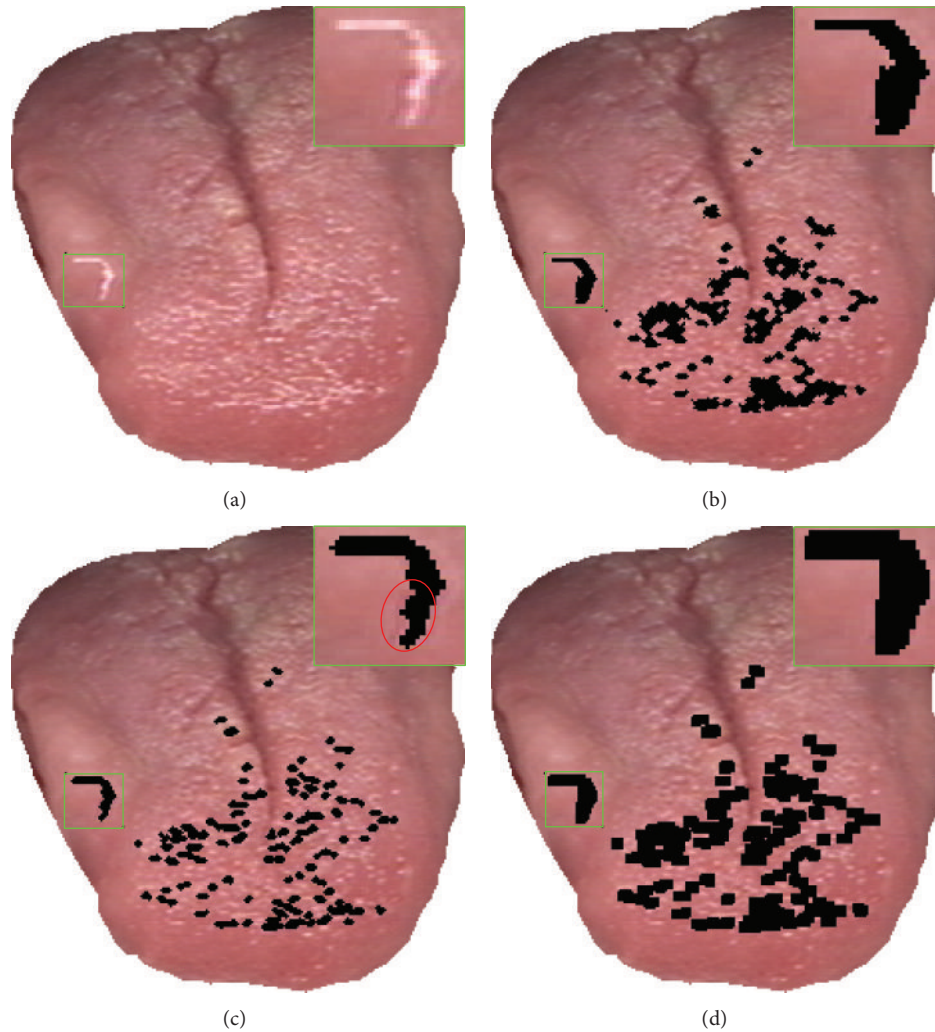
Figure 6: Comparison of specular reflection area detection results: (a) the original tongue image, (b) the detection result by using the proposed superpixel-based method, (c) the detection result of morphological dilation method by using disk-shaped structuring element with 1×1 pixel, and (d) 5×5 pixels.

band using Formula (5), we could detect all the subreflections and update the highlight reflections at the same time.

Parameter k in (4) is one parameter that indirectly affects the granularity of the final segmentation. With a larger k, it usually leads to larger regions but also reduces a higher likelihood of missing segmentation boundaries, while a smaller k often leads to a consistent oversegmentation. Figure 5 shows the detection results by setting k to 0.01, 0.05, 0.10, and 0.15, respectively. The mask map in Figure 5(b) can cover the highlight reflections and subreflections, while excluding the normal pixels around subreflections. In Figure 5(a), the subreflections areas are not detected, while in Figures 5(c) and 5(d) many normal pixels are misclassified to the subreflection area. Our experiments show that, for most of tongue images, satisfactory subreflection detection results are obtained with $k \in [0.05, 0.07]$.

Finally, the highlight reflections and subreflections were combined together as the final detection results of specular reflection areas for inpainting. Figure 6(b) shows the final detection result of the specular reflection area of Figure 6(a).

For comparison, Figures 6(c) and 6(d) show the results of morphological dilation based method [21] by using disk-shaped structuring element with 1×1 pixel and 5×5 pixels, respectively. Compared with the enlarged subimages in Figures 6(b) with 6(c) and 6(d), we can see that the mask map in Figure 6(c) does not cover all the subreflections, especially in the part inside the red circle, while the mask map in Figure 6(d) eroded too much normal pixels. Generally, the proposed superpixel-based method could accurately detect the highlight reflection and subreflection areas while compared with morphological dilation.

3. The WNNM-Based Nonlocal Inpainting Method

In this section, we proposed an examplar-based inpainting method, that is, WNNM-based inpainting, for tongue specular reflection removal. The examplar-based methods can be traced back to 1999 [24] and had been widely adopted

for image inpainting [25–31]. The arising of nonlocal self-similarity methods [32] further triggered the development of examplar-based inpainting approaches. Most existing methods, however, fill the holes based on the best matched patch or the mean of the nonlocal similar patches, while the second-order information usually is neglected in inpainting. To remedy this, a low rank based inpainting model [33] is proposed to synthesize the missing regions, where tensor trace norm is adopted as the low rank regularizer. Besides, the WNNM model [34] generally is superior to other low rank models, for example, trace norm or nuclear norm, for image denoising. This motivates us to employ WNNM for tongue specular reflection removal, resulting in the proposed WNNM-based inpainting model.

Given a small patch on the tongue image, there are many nonlocal similar patches across the whole image. If we stretch each patch to a vector and stack all the vectors in a matrix, it is intuitive that the matrix should be of low rank. Based on this assumption, the tongue image inpainting work could be regarded as a low rank matrix completion problem. By utilizing a newly proposed weighted nuclear norm model, we further showed that tongue image inpainting could be well solved by iteratively performing weighted nuclear norm minimization (WNNM).

In the following subsection, a brief review on matrix completion was firstly introduced. Then, WNNM was described. Finally, the WNNM-based nonlocal inpainting method was proposed for specular reflections removal of tongue image.

3.1. Matrix Completion. Matrix completion aims to recover a low rank matrix from incomplete samples of its entries, which has received considerable attention in many areas of engineering and science [35, 36], for example, the well-known Netflix problem [37]. Matrix completion can be cast as the following minimization problem:

$$\min_{X \in \mathfrak{R}^{m \times n}} \{ \text{rank}(X) : P_\Omega(X) = P_\Omega(B) \}, \qquad (6)$$

where $X \in \mathfrak{R}^{m \times n}$ is recovered from $B \in \mathfrak{R}^{m \times n}$, Ω denotes the set of position of known entries in matrix B, and $P_\Omega(\cdot)$ is a linear operator with the following definition: for any matrix $D \in \mathfrak{R}^{m \times n}$,

$$P_\Omega(D) = \begin{cases} D_{i,j} & (i,j) \in \Omega, \\ 0 & (i,j) \notin \Omega. \end{cases} \qquad (7)$$

The model in (6) is nonconvex and is NP-hard problem which is nontrivial to solve. Therefore, convex relaxation of (6) usually was adopted with the following formulation:

$$\min_{X \in \mathfrak{R}^{m \times n}} \{ \|X\|_* : P_\Omega(X) = B \}, \qquad (8)$$

where $\|X\|_* = \sum_i \sigma_i(X)$ is the nuclear norm of X, and $\sigma_i(X)$ are the ith singular values of X. The problem in (8) can be approximated by the following unconstrained convex minimization problem:

$$\min_{X \in \mathfrak{R}^{m \times n}} \left\{ \frac{1}{2} \|P_\Omega(X) - B\|_F^2 + \lambda \|X\|_* \right\}, \qquad (9)$$

where $\| \cdot \|_F$ denotes the Frobenius norm, and λ is a tradeoff parameter. When λ goes infinite, the solution of (9) will be the same as that of (8). As an unconstrained optimization problem, (9) could be solved by the iterative shrinkage algorithm [38] and the APG method [39].

3.2. Weighted Nuclear Norm Minimization. The iterative shrinkage algorithm of the model in (9) usually involves solving the nuclear norm minimization (NNM) problem as

$$X_{k+1} = \arg \min_X \left\{ \|X - Y^*\|_F^2 + \lambda \|X\|_* \right\}, \qquad (10)$$

where $X, Y^* \in \mathfrak{R}^{m \times n}$, $Y^* = (X_k - (1/L_f)\nabla f(X_k))$, $f(X_k) = (1/2)\|P_\Omega(X_k) - B\|_F^2$, and $\lambda > 0$.

Cai et al. [40] proved that the NNM problem of (10) can be easily solved by the singular value thresholding (SVT) method; that is,

$$X_{k+1} = U \mathcal{S}_\lambda(\Sigma) V^T, \qquad (11)$$

where X_{k+1} is the solution to (10), $Y^* = U\Sigma V^T$ is the singular value decomposition (SVD) of Y^*, $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ is the diagonal matrix with singular values, and $\mathcal{S}_\lambda(\Sigma)$ is the soft-thresholding function on Σ as

$$\mathcal{S}_\lambda(\Sigma)_i = \max(\sigma_i - \lambda, 0). \qquad (12)$$

Recently, Gu et al. [34] suggested that, as the soft-thresholding operator $\mathcal{S}_\lambda(\Sigma)$ shrinks each singular value with the same λ in order to pursue the convex property, it ignores the prior knowledge about the singular values. Compared with the small singular values, the larger ones are generally associated with the major projection orientation of the matrix in the lower subspace and they should be shrunk less to preserve the major data components. Thus, they extended the standard nuclear norm to the weighted nuclear norm:

$$\|X\|_{w,*} = \sum_i |w_i \sigma_i(X)|_1, \qquad (13)$$

where $\mathbf{w} = [w_1, \dots, w_n]$ and $w_i \geq 0$ is a nonnegative and assigned to the singular value $\sigma_i(X)$ and the weighted nuclear norm minimization (WNNM) problem could be formulated as

$$\widehat{X} = \arg \min_X \left\{ \|Y^* - X\|_F^2 + \lambda \|X\|_{w,*} \right\}. \qquad (14)$$

Generally, the model in (14) is not convex, but Xie et al. [41] further proved that if the weights, assigned to the singular vales, are arranged in ascending orders, the globally optimal solution could be obtained with the following theorem.

Theorem 1 (see [41]). *If the weights $\mathbf{w}$ satisfy $0 \leq w_1 \leq w_2 \leq \cdots \leq w_n$, then the WNNM problem (16) has a globally optimal solution:*

$$\widehat{X} = U \mathcal{S}_w(\Sigma) V^T, \qquad (15)$$

where $Y^* = U\Sigma V^T$ denotes the SVD of matrix Y^*, Σ is the diagonal matrix with singular values arranged in descending

order; that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, and $\mathcal{S}_w(\Sigma)$ is the soft-thresholding operator:

$$\mathcal{S}_w(\Sigma)_i = \max(\sigma_i - w_i, 0). \tag{16}$$

WNNM has been applied to image denoising and achieved better results than state of the arts, such as LSSC [42] and BM3D [43], by PSNR and SSIM values. Moreover, WNNM is effective in preserving the local structures of images and generating less visual artifacts.

3.3. Tongue Image Inpainting by Using WNNM.

In this section, we proposed a WNNM-based nonlocal inpainting method for repairing tongue image specular reflection areas. First, the WNNM-based inpainting model was introduced to utilize the nonlocal information of tongue image. Then, we described the proposed optimization algorithm by iteratively solving a series of standard WNNM problems. Finally, we analyzed the convergence of our algorithm.

Given a tongue image B with specular reflection $\overline{\Omega}$, the proposed WNNM-based tongue inpainting model is formulated as

$$\min_{X \in \mathfrak{R}^{m \times n}} F(X) = \left\{\frac{1}{2} \|P_\Omega(X) - B\|_F^2 + \lambda \|X\|_{w,*}\right\}, \tag{17}$$

where X is the inpainted image, Ω denotes the nonreflection area, and P_Ω is a linear projection operator defined in (7). The model in (17) can be obtained by substituting the standard nuclear norm with weighted nuclear norm in (9). Because of the introduction of the projection operator P_Ω, the model in (17) cannot be directly solved by WNNM algorithm in [34]. Thus, we adopt the following iterative shrinkage algorithm. In each iteration, we consider the following surrogate function [44] of $F(X)$ at a given point X_k:

$$Q(X, X_k) := f(X_k) + \langle \nabla f(X_k), X - X_k \rangle$$
$$+ \frac{L_f}{2} \|X - X_k\|_F^2 + \lambda \|X\|_{w,*}, \tag{18}$$

where $f(X_k) = (1/2)\|P_\Omega(X_k) - B\|_F^2$ and L_f is the Lipschitz constant of $\nabla f(X_k)$. $Q(X, X_k)$ satisfies that (i) $Q(X, X_k) \geq F(X)$ for any X and (ii) $Q(X_k, X_k) = F(X_k)$. By ignoring the constant term, the minimization of (18) can be formulated as

$$\arg\min_X Q(X, X_k) \doteq \left\{\frac{L_f}{2} \|X - Y_k\|_F^2 + \lambda \|X\|_{w,*}\right\}, \tag{19}$$

where $Y_k = X_k - (1/L_f)P_\Omega^T(P_\Omega(X_k) - B)$. Equation (19) has the same form of (14), which could be solved by using WNNM. Finally, we can get the minimizer of (17) by iteratively solving (19) with proximal gradient as described in Algorithm 1.

There are many repeated patterns in the tongue image which are useful for tongue image inpainting. In order to take advantage of this nonlocal information, we split the tongue image into many 5×5 pixels patches. For a single patch p_j, many nonlocal similar patches could be found across the whole image. Stretching each similar patch to a vector and stacking them into a matrix Z_p, it is intuitively that Z_p should be of low rank. By using (19), we could get the estimation of Z_p as

$$\widehat{Z}_p = \arg\min_Z \left\{\frac{L_f}{2} \|Z - Z_p\|_F^2 + \lambda \|Z\|_{w,*}\right\}, \tag{20}$$

which could be solved by the generalized soft-thresholding method in Theorem 1. Specifically, the weight w_i assigned to the ith singular value, $\sigma_i(Z_p)$, of Z_p as $w_i = c\sqrt{n}/(\sigma_i(Z_p) + \varepsilon)$, where $c > 0$ is a constant, n is the number of similar patches in Z_p, and $\varepsilon = 10^{-6}$ is to avoid dividing by zero. In this paper c is set to 10^{-3}, λ is set to 0.1, L_f and N are set to 2 and 5, respectively. After that, the patch p_j can be inpainted by finding the most similar one to it in $\widehat{Z}_p$. Finally, by aggregating all the patches $\widehat{p}_j$ together, the whole image X can be updated.

3.4. Convergence of the WNNM-Based Nonlocal Inpainting Method.

In this section, we proved the convergence of WNNM-based inpainting method which could be summarized as in the following theorem.

Theorem 2. *To solve $\min_{X \in \mathfrak{R}^{m \times n}} \{F(X) = (1/2)\|P_\Omega(X) - B\|_F^2 + \lambda\|X\|_{w,*}\}$, one iteratively performs the two steps*

$$Y_k = X_k - P_\Omega^T(P_\Omega X_k - B), \tag{21}$$

$$X_{k+1} = \arg\min_X \|X - Y_k\|_F^2 + \lambda \|X\|_{w,*} \tag{22}$$

and has

$$F(X_{k+1}) \leq F(X_k). \tag{23}$$

Proof. For image inpainting, we have $L_f = 1$. Define the surrogate function $Q(X, X_k)$ as

$$Q(X, X_k) = \frac{1}{2} \|P_\Omega X_k - B\|_F^2$$
$$+ \frac{1}{2} \langle (I - P_\Omega^T P_\Omega)(X_k - X), X_k - X \rangle \tag{24}$$
$$+ \lambda \|X\|_{w,*}.$$

Since $\|P_\Omega\|_2 \leq 1$, $(\mathbf{I} - P_\Omega^T P_\Omega)$ is positive semidefinite, where $\mathbf{I}$ is the identity matrix. Thus we can obtain that

$$Q(X, X_k) \geq Q(X, X) = f(X). \tag{25}$$

Input: The tongue image B, iteration number N, and L_f
(1)   Initialize: $X_0 = B$, $Y_0 = X_0 - (1/L_f)P_\Omega^T(P_\Omega(X_0) - B)$
(2)   **for** $k = 1 : N$ **do**
(3)   **for** each patch p_j in Y_{k-1} **do**
(4)   Find similar patches matrix Z_j
(5)   Estimate weight vector $\mathbf{w}$
(6)   Singular value decomposition $[U, \Sigma, V] = \text{SVD}(Z_j)$
(7)   Estimate patch $\widehat{Z}_j = U \mathcal{S}_w(\Sigma)V^T$ and $\widehat{p}_j$
(8)   **end for**
(9)   Aggregate all the single patches $\widehat{p}_j$ to form the image X_k
(10) $Y_k = X_k - (1/L_f)P_\Omega^T\left(P_\Omega\left(X_k\right) - B\right)$
(11)   **end for**
Output: $X = X_N$.

ALGORITHM 1: WNNM-based nonlocal inpainting method.

Given X_k, with the surrogate function in (24), we have

$$\min_X Q\left(X, X_k\right)$$

$$= \min_X \frac{1}{2}\left\|P_\Omega X_k - B\right\|_F^2$$

$$+ \frac{1}{2}\left\langle\left(I - P_\Omega^T P_\Omega\right)\left(X - X_k\right), X - X_k\right\rangle$$

$$+ \lambda\left\|X\right\|_{w,*}$$

$$= \min_X \frac{1}{2}\left\|X - \left(I - P_\Omega^T P_\Omega\right)X_k - P_\Omega^T B\right\|_F^2 + \lambda\left\|X\right\|_{w,*} \qquad (26)$$

$$+ C$$

$$= \min_X \frac{1}{2}\left\|X - \left(X_k - P_\Omega^T\left(P_\Omega X_k - B\right)\right)\right\|_F^2$$

$$+ \lambda\left\|X\right\|_{w,*} + C,$$

where C is the constant term w.r.t. to X. Equation (26) is actually the iterative process of updating Y_k using gradient descent as

$$Y_k = X_k - P_\Omega^T\left(P_\Omega X_k - B\right). \qquad (27)$$

With WNNM, we can obtain the optimal solution of (26); $X_{k+1} = \arg\min_X Q(X, X_k)$. Then one can get

$$Q\left(X_{k+1}, X_k\right) \le Q\left(X_k, X_k\right) = F\left(X_k\right). \qquad (28)$$

From (25), one can derive that

$$Q\left(X_{k+1}, X_k\right) \ge Q\left(X_{k+1}, X_{k+1}\right) = F\left(X_{k+1}\right). \qquad (29)$$

Combining (28) and (29), we have

$$F\left(X_{k+1}\right) = Q\left(X_{k+1}, X_{k+1}\right) \le Q\left(X_{k+1}, X_k\right)$$

$$\le Q\left(X_k, X_k\right) = F\left(X_k\right). \qquad (30)$$

From (30), we get $F(X_{k+1}) \le F(X_k)$.   □

From Theorem 2, the proposed algorithm can guarantee to decrease the loss function along with iterations until convergence to a fixed point.

4. Experimental Results

In this section, we validated the effectiveness of the proposed method for tongue image inpainting. For better evaluation, both synthetic images and real images were used in the experiment. The inpainting results of the TV-based method [21] were used for comparison.

4.1. Experimental Results on Synthetic Images. In this section, we quantitatively compared the performance of the competing methods on synthetic images. The synthetic images were constructed as follows: we first detect the specular reflections of one real tongue image and then randomly put these reflections onto nonreflection tongue images to synthesize new images for inpainting. The benefit was that, on one hand, we could ensure that the topological structure of reflections was as natural as real ones. On the other hand, we could conveniently calculate the peak signal to noise ratio (PSNR) since we had the ground truth.
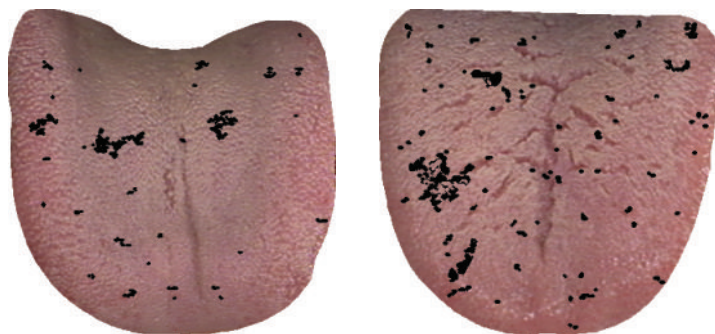
We used two ways to detect the highlight reflections of tongue image. One was the morphological dilation based method in [20] and the other was our superpixel based method. Synthetic images with both kinds of reflections were used in the experiments of WNNM-based inpainting and TV-based inpainting, respectively.

Figure 7 shows the comparison experimental results of competing methods on synthetic tongue images with reflections detected by morphological dilation based method, while Figure 8 shows the results on the synthetic tongue images with reflections detected by the proposed superpixel based method.
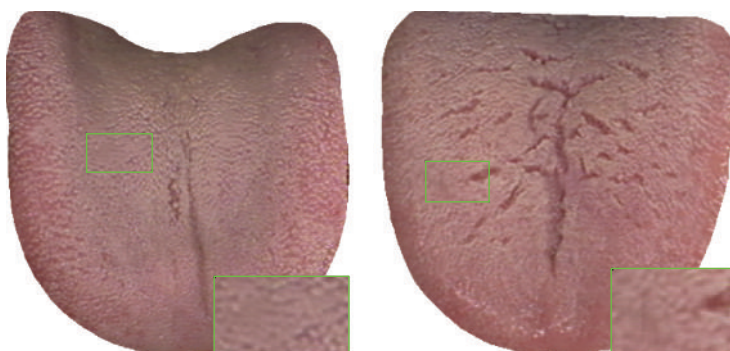
From Figures 7, 8, and 9, we could find that for the same synthetic tongue image, in terms of PSNR, the proposed inpainting method obtained better result than the TV-based inpainting method. Furthermore, the TV-based inpainting method usually finds the pixels via minimizing total variation value, while the proposed method usually finds the similar pixels to the area around the reflections. Thus, compared with the TV-based method, the proposed method is much better in handling the case of tongue image with large reflection areas by involving more texture information, as shown in Figure 9.
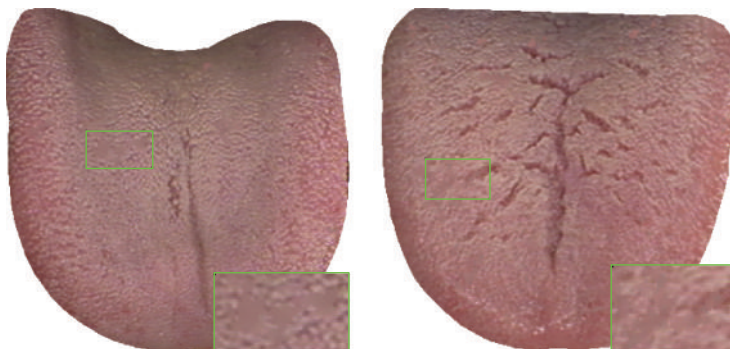
(a) Ground truth

(b) Synthetic tongue images

(c) Results of WNNM-based method (PSNR: left: 30.37, right: 31.56)

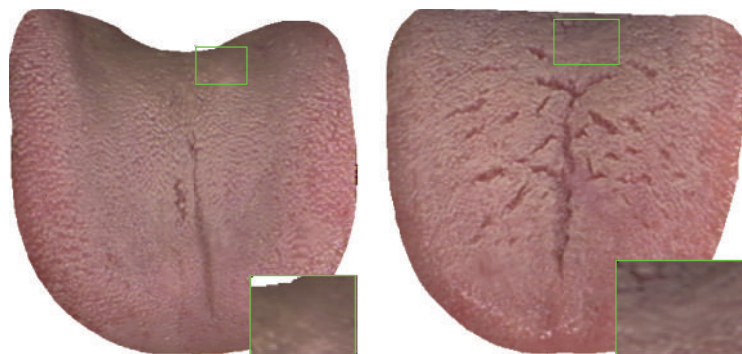(d) Results of TV-based method (PSNR: left: 29.32, right: 30.06)

FIGURE 7: Comparision of inpainting results of synthetic tongue images with reflection detected by morphological dilation based method.
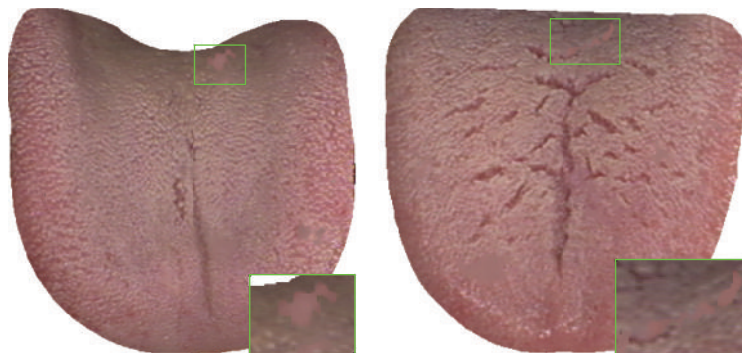
(a) Ground truth

(b) Synthetic tongue images

(c) Results of our WNNM-based method (PSNR: left: 32.29, right: 31.34)

(d) Results of the TV-based method (PSNR: left: 31.39, right: 30.47)

FIGURE 8: Comparison of inpainting results of synthetic tongue images with reflection detected by superpixel based method.

(a) Ground truth

(b) Tongue image with missing block

(c) Inpainting result of our WNNM-based method (PSNR = 25.39)

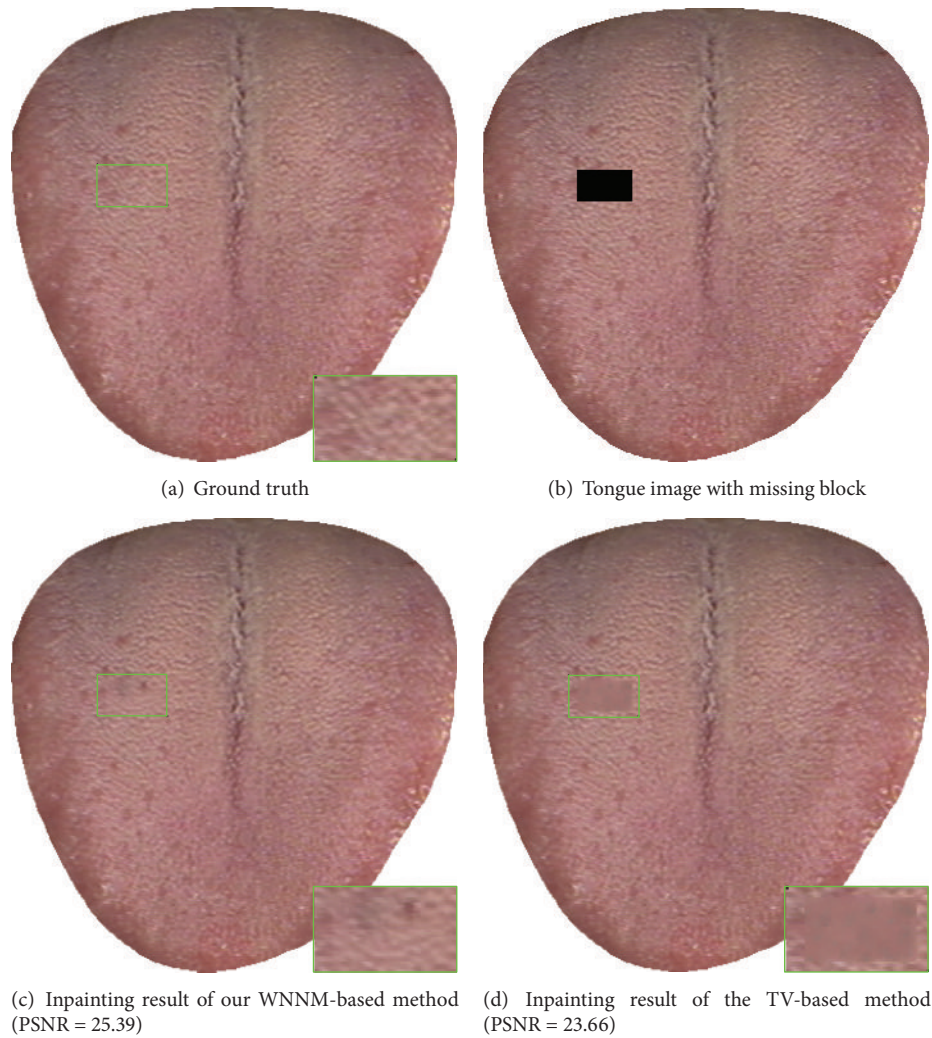(d) Inpainting result of the TV-based method (PSNR = 23.66)

FIGURE 9: Comparison of inpainting result on tongue image with large reflection area.

In Figure 9(b), we manually add a block to original tongue image to simulate the large reflection area. It is easy to find that the inpainting result of the proposed method is more similar to the ground truth, which is much better than the TV-based method by involving more texture information of tongue body.

4.2. Experimental Results on Real Tongue Images. For better evaluation of the proposed inpainting method, we used two real tongue images with serious specular reflection areas to validate its performance. The experimental inpainting results of three real tongue images are shown in Figure 10. Moreover, on the lower-right of each image of Figures 10(b) and 10(c), we further showed the enlarged subimages of the inpainting images. From Figure 10, one could observe that the proposed method would achieve satisfactory visual results with much texture information of tongue image while part of horizontal or vertical discontinuity can be observed from the results obtained by the TV-based method.
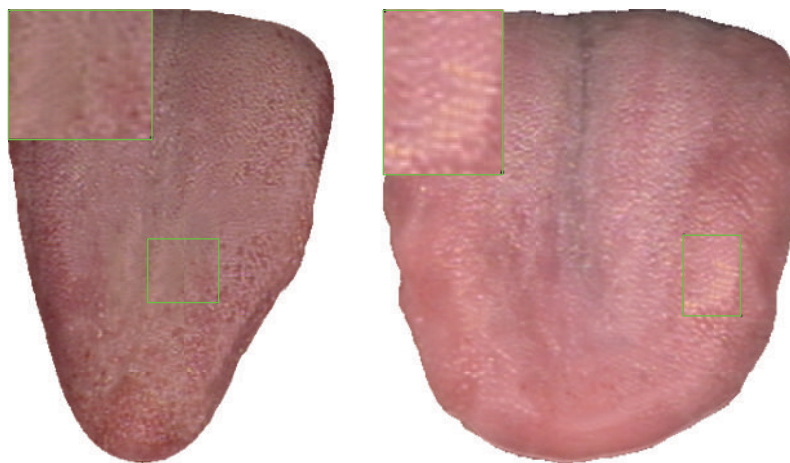
Based on the results on synthetic and real tongue images, we showed that the proposed method was better than TV-based method, either in terms of PSNR or in the visual results. The proposed method filled the reflection area with more texture information of tongue image, which was useful in the tongue texture analysis applications. Compared with TV-based inpainting method, the proposed method was more suitable for the specular reflection removal of tongue images.
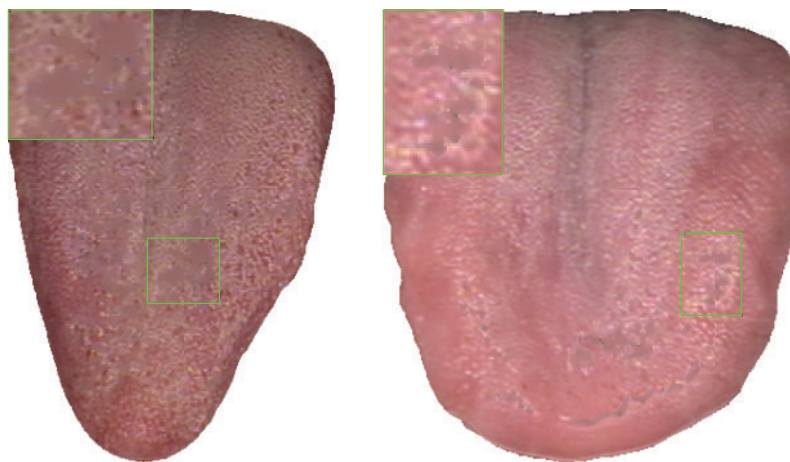
5. Conclusion

In this paper, we proposed a WNNM-based nonlocal inpainting method for specular reflections removal of tongue image. In the proposed method, superpixel segmentation was adopted to handle the specular reflections detection task. Then, based on the nonlocal self-similarity, we transformed the problem of specular reflections inpainting of tongue image to a matrix completion problem and further proved the convergence of the proposed WNNM-based tongue image inpainting method.

(a) Original tongue images with specular reflections

(b) Inpainting results of our WNNM-based method

(c) Inpainting results of the TV-based method

Figure 10: Comparison of inpainting results of WNNM-based and TV-based method.

We evaluated the performance of the proposed approach by comparing with the TV-based tongue image inpainting method. The comparison experimental results on both synthetic tongue images and real tongue image showed that the proposed method could achieve not only higher PSNR, but also more satisfactory visual effects, which could involve more texture information of tongue body, and was more suitable for the specular reflections removal of tongue images.

Conflict of Interests

Acknowledgment

References

[1] S. E. Strome, K. McClatchey, P. R. Kileny, and C. F. Koopmann Jr., "Neonatal choristoma of the tongue containing glial tissue: diagnosis and surgical considerations," *International Journal of Pediatric Otorhinolaryngology*, vol. 33, no. 3, pp. 265–273, 1995.

[2] J. K. Anastasi, L. M. Currie, and G. H. Kim, "Understanding diagnostic reasoning in TCM practice: tongue diagnosis," *Alternative Therapies in Health and Medicine*, vol. 15, no. 3, pp. 18–28, 2009.

[3] B. Kirschbaum, *Atlas of Chinese Tongue Diagnosis*, Eastland, Seattle, Wash, USA, 2000.

[4] Q. Li, Y. Wang, H. Liu, Y. Guan, and L. Xu, "Sublingual vein extraction algorithm based on hyperspectral tongue imaging technology," *Computerized Medical Imaging and Graphics*, vol. 35, no. 3, pp. 179–185, 2011.

[5] Z. Liu, H. Wang, H. Xu, and S. Song, "3D Tongue reconstruction based on multi-view images and finite element," *Advances in Information Sciences and Service Sciences*, vol. 3, no. 11, 2011.

[6] Y. Cai, "A novel imaging system for tongue inspection," in *Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference*, pp. 159–163, May 2002.

[7] B. Pang, D. Zhang, N. Li, and K. Wang, "Computerized tongue diagnosis based on Bayesian networks," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 10, pp. 1803–1810, 2004.

[8] L. Jiang, W. Xu, and J. Chen, "Digital imaging system for physiological analysis by tongue colour inspection," in *Proceedings of the 3rd IEEE Conference on Industrial Electronics and Applications (ICIEA '08)*, pp. 1833–1836, June 2008.

[9] Y. J. Jeon, K. Kim, J. Do, H. Ryu, and J. Kim, "The development of a tongue diagnosis system and the evaluation of reproducibility," *Korean Journal of Oriental Medicine*, vol. 14, no. 3, pp. 97–102, 2008.

[10] X. Wang and D. Zhang, "A high quality color imaging system for computerized tongue image analysis," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5854–5866, 2013.

[11] B. Huang, J. Wu, D. Zhang, and N. Li, "Tongue shape classification by geometric features," *Information Sciences*, vol. 180, no. 2, pp. 312–324, 2010.

[12] Q. Li and Z. Liu, "Tongue color analysis and discrimination based on hyperspectral images," *Computerized Medical Imaging and Graphics*, vol. 33, no. 3, pp. 217–221, 2009.

[13] C. H. Li and P. C. Yuen, "Tongue image matching using color content," *Pattern Recognition*, vol. 35, no. 2, pp. 407–419, 2002.

[14] K. H. Kim, J.-H. Do, H. Ryu, and J.-Y. Kim, "Tongue diagnosis method for extraction of effective region and classification of tongue coating," in *Proceedings of the 1st International Workshops on Image Processing Theory, Tools and Applications (IPTA '08)*, pp. 1–7, IEEE, Sousse, Tunisia, November 2008.

[15] W. Huang, Z. Yan, J. Xu, and L. Zhang, "Analysis of the tongue fur and tongue features by naive Bayesian classifier," in *Proceedings of the International Conference on Computer Application and System Modeling (ICCASM '10)*, pp. V4-304–V4-308, October 2010.

[16] Z. Yan, K. Wang, and N. Li, "Computerized feature quantification of sublingual veins from color sublingual images," *Computer Methods and Programs in Biomedicine*, vol. 93, no. 2, pp. 192–205, 2009.

[17] Z. Gao, L. Po, W. Jiang, X. Zhao, and H. Dong, "A novel computerized method based on support vector machine for tongue diagnosis," in *Proceedings of the 3rd IEEE International Conference on Signal Image Technologies and Internet Based Systems (SITIS '07)*, pp. 849–854, December 2007.

[18] C. H. Siu, Y. He, and D. T. C. Thach, "Machine learning for tongue diagnosis," in *Proceedings of the 6th International Conference on Information, Communications and Signal Processing (ICICS '07)*, pp. 1–5, IEEE, Singapore, December 2007.

[19] Z. Yan, K. Wang, and N. Li, "An algorithm for extraction of near infrared sublingual veins," in *Proceedings of the International Symposium on Computational Intelligence and Design (ISCID '08)*, pp. 60–63, October 2008.

[20] B. Huang, *Tongue feature analysis and symptom diagnosis classification [Ph.D. dissertation]*, Harbin Institue of Technology, 2009.

[21] D. Jia, N. Li, C. Li, S. Li, and W. Zuo, "Reflection removal of tongue image via total variation-based image inpainting," in *Proceedings of the International Conference on E-Product E-Service and E-Entertainment (ICEEE 2010)*, pp. 1–4, Henan, China, November 2010.

[22] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[23] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Efficient hierarchical graph-based video segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2141–2148, June 2010.

[24] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1033–1038, IEEE, September 1999.

[25] R. Bornard, E. Lecan, L. Laborelli, and J.-H. Chenot, "Missing data correction in still images and image sequences," in *Proceedings of the ACM International Conference on Multimedia*, pp. 355–361, Juan les Pins, France, December 2002.

[26] I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-based image completion," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 303–312, 2003.

[27] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[28] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplar-based inpainting," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II-721–II-728, IEEE, June 2003.

[29] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1153–1165, 2010.

[30] C. Guillemot and O. Le Meur, "Image inpainting: overview and recent advances," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 127–144, 2014.

[31] P. Buyssens, M. Daisy, D. Tschumperlé, and O. Lézoray, "Exemplar-based inpainting: technical review and new heuristics for better geometric reconstructions," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1809–1824, 2015.

[32] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 60–65, June 2005.

[33] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.

[34] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 2862–2869, June 2014.

[35] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[36] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, "Practical low-rank matrix approximation under robust L1-norm," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 1410–1417, June 2012.

[37] N. Srebro and R. R. Salakhutdinov, "Collaborative filtering in a non-uniform world: learning with the weighted trace norm," in *Proceedings of the 24th Annual Conference on Neural Information Processing Systems (NIPS '10)*, pp. 2056–2064, December 2010.

[38] A. Nemirovski, *Efficient Methods in Convex Programming*, TECHNION, 2005.

[39] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 3, pp. 615–640, 2010.

[40] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[41] Q. Xie, D. Meng, S. Gu et al., "On the optimal solution of weighted nuclear norm minimization," http://arxiv.org/abs/1405.6012.

[42] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 2272–2279, IEEE, Kyoto, Japan, October 2009.

[43] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[44] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 1–20, 2000.