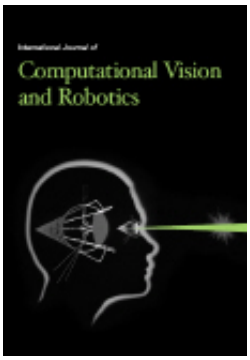


International Journal of Computational Vision and Robotics

ISSN online: 1752-914X - ISSN print: 1752-9131
https://www.inderscience.com/ijcvr

Image enhancement based on skin-colour segmentation and smoothness

Haitao Sang, Bo Chen, Shifeng Chen, Li Yan

DOI: 10.1504/IJCVR.2021.10036485

Article History:

Received:	02 November 2020
Accepted:	16 December 2020
Published online:	30 November 2022

Image enhancement based on skin-colour segmentation and smoothness

Haitao Sang, Bo Chen and Shifeng Chen*

College of Information Engineering,
Lingnan Normal University,
Zhanjiang 524048, China
Email: sanght@lingnan.edu.cn
Email: chenbo20040607@126.com
Email: chens@lingnan.edu.cn
*Corresponding author

Li Yan

College of Science,
Guangdong University of Petrochemical Technology,
Maoming 525000, China
Email: bynd_yanli@163.com

Abstract: The image restoration tasks represented by image denoising, super-resolution and image deblurring have a wide range of application background, and have become a research hotspot in academia and business circles. A novel image enhancement algorithm based on skin texture preservation is proposed in this paper. The mask has been obtained using the Gaussian fitting, which can have a box blur for many times and for skin feathering. The denoising smoothing image is fused with the original image mask to preserve the hair details of the original image and enhance the edge details of the contour, so as to provide more effective information for the extraction of edge features. Compared with different methods of image smoothing algorithms, this algorithm is more effective in smoothing the skin edge contour and achieving better detection of images. Experimental results show that the proposed algorithm has strong adaptive capacity and significant effect on most images detection. Specifically, it can moderately smooth the edges of the areas with many details, leaving no traces of an artificial process. The proposed algorithm with image enhancement has a wide range of practicality.

Keywords: image enhancement; image restoration; image generation and synthesis; texture preserving smoother; skin-colour model.

Reference to this paper should be made as follows: Sang, H., Chen, B., Chen, S. and Yan, L. (2023) 'Image enhancement based on skin-colour segmentation and smoothness', *Int. J. Computational Vision and Robotics*, Vol. 13, No. 1, pp.1–20.

Biographical notes: Haitao Sang received his PhD degree from Harbin University of Science And Technology, in 2017. He is currently an Associate Professor with Lingnan Normal University. His research interests are in the field of large-scale visual artificial intelligence, including visual analysis, recognition, search and mining.

Bo Chen received his PhD degree from Dalian University of Technology, in 2002. He is currently a Professor with Lingnan Normal University. His research interests include complex networks and bioinformatics.

Shifeng Chen received his MS degree in Computer Science and Technology from Dalian University, in 2012, and the PhD degree from Dalian Maritime University, in 2018. He is currently a Lecturer with the School of Information Engineering, Lingnan Normal University, Zhanjiang, China. His current research interests include combinatorial optimisation using heuristic and metaheuristic techniques.

Li Yan received her PhD degree from Nanjing Agricultural University, in 2016. She is currently an Associate Professor with Guangdong University of Petrochemical Technology. Her research interests include image processing and bioinformatics.

1 Introduction

With the continuous advancement of digital image imaging equipment technology, the number of images obtained through camera, mobile phones, video monitoring and other channels shows an exponential growth trend. Images have become an important means for people to perceive the world and exchange information with the outside world. However, in the process of image acquisition, transmission and storage, the complex imaging factors in reality (such as noise, low light, camera shake, object motion) will lead to the degradation of image quality (such as noise, blur, distortion), thus reducing the visual perception quality of image. In recent years, through the exploration of reasonable restoration model and efficient restoration method to restore high-quality clear image, it has become a research hotspot in academia and business circles.

In the process of digitisation and transmission of digital images, they are often affected by noise interference from imaging equipment and external environment, and different types of complex noise are introduced. The task of image denoising requires not only to remove the noise in the image as much as possible, but also to keep the original image edge, texture and other details. At the same time, digital images will exist in the form of lower resolution in the process of acquisition, transfer and preservation, which affects the completeness of local details and the richness of image information. Compared with the high cost of improvement of hardware equipment and the limitations of its development, image super-resolution algorithms are actually more valuable. For the common problem of image blurring, how to effectively estimate the blur process, deal with noise and estimate error, will be very important to restore high-quality and clear images.

At present, image restoration tasks represented by image denoising, super-resolution and image deblurring have a wide range of application backgrounds, such as the daily life applications on mobile phones, the face and license plate image restoration technology for criminal investigation and evidence collection, and the related applications in the fields of medical image, remote sensing and satellite imaging. In addition, the restoration and enhancement of low-quality images also lay the foundation for high-level semantic analysis tasks. Therefore, image enhancement and image synthesis have important

research significance and practical value, and have been widely concerned by scholars at home and abroad.

In this paper, we put forward an image enhancement based on skin-colour segmentation and smoothness method by computing the magnitude of target gradients. We first tested the performance of the mean-shift, EMD, and beeps filter algorithms to smooth or denoise images. The contrastive analysis demonstrated effectiveness of the mean-shift denoising method. We then used the Gaussian skin-colour model to perform skin-colour segmentation on the original image, and then blurred the skin area by a Gaussian function to realise feathering and the creation of mask. The image denoised by the mean-shift algorithm were then fused with the original image. Finally, the enhancement effectiveness was tested by an eight neighbourhood gradient edge detection operator.

The organisation of rest of the paper is as follows: Section 2 describes related research work in detail. Section 3 presents smooth denoising method. Proposed countermeasures along with details of segmentation and enhancement setting are employed in Section 4. Section 5 presents the experimental results and discussion. Last but not least, Section 6 concludes the paper.

2 Literature review

The application of neural network in image denoising and restoration can be traced back to at least 2002. In Suzuki et al. (2002), proposed that image denoising or edge enhancement can be achieved by learning multi-layer neural network. In 2009, the authors proposed a completely convoluted convolutional neural network (CNN) for denoising, and found that CNN can get the same or even better performance as wavelet and Markov random field methods (Jain and Seung, 2009). Then, in 2012, the authors combined sparse coding and deep network training with automatic encoder for image denoising (Xie et al., 2012). With the rapid growth of training data and computing power, multi-layer perceptron (MLP) has gradually achieved the same or higher performance as the block matching 3D (BM3D) algorithm in denoising and other issues (Burger et al., 2012). In 2014, the authors proposed a discriminative learning model based on successive iterative learning for image restoration: cascade of shrinkage fields (CSF) based on semi quadratic splitting algorithm (Schmidt et al., 2014), this method learned the phase model parameters from the training data by expanding the prediction process into iterative learning algorithm. In 2015, the authors proposed a nonlinear reaction diffusion (TNRD) model from the perspective of reaction diffusion equation, learned the filter and response function of each iteration, and explained the model from the perspective of recurrent neural network (Chen and Pock, 2015). Inspired by CSF and TNRD, in Zhang et al. (2017), designed a deep denoising network based on convolutional neural network (DnCNN), the model uses convolution neural network to separate noise from noise image from function regression through end-to-end residual learning, and achieves significantly better denoising results than other methods. At the same time, a series of network structure-based improvements have been proposed. Residual encoder-decoder network (REDNet) adopts the deep convolution coding decoding framework based on the symmetric jump link, so that in the reverse process, information can be directly transferred from the top layer to the bottom layer (Mao et al., 2016). Memory network (MemNet) further proposes a long-term memory model for image denoising (Tai et al.,

2017); multi-level wavelet CNN (MWCNN) proposes a MWCNN framework by reducing the discrete, the combination of wave transform and convolution network is beneficial to the restoration of image details (Liu et al., 2018).

Rich edge texture information is contained in images, and the grey-scale mutation along the vertical direction of the texture makes the image have non-stationary signal characteristics, which can be processed by non-stationary signal analysis. The empirical mode decomposition method is an adaptive processing method that performs well on non-stationary signal processing in recent years, and can avoid the disadvantages of traditional non-stationary signal processing methods. In view of the achievements of EMD in one-dimensional non-stationary signal processing, researchers have tried to extend EMD into the field of images (Hao et al., 2017). And then bi-dimensional empirical mode decomposition is developed, achieving good results in image processing such as image fusion and feature extraction.

The above methods usually need to train models for different noise levels, which is not only lack of flexibility, but also can not be applied to the real noise image with more complex degradation process. In Lee et al. (2009), use the Gaussian mixture model (GMM) and Bayesian segmentation algorithm to segment the skin area, and finally use the Gaussian smooth skin area to achieve automatic skin beautification. But the effect is general, in the area of rich texture, it will cause unstable GMM parameters, resulting in inaccurate segmentation of the skin area, and eventually cause image blur. In Chen et al. (2010), combined with downsampling, skin colour detection algorithm and feature classifier, segmented skin colour region, smoothed skin with bilateral edge preserving filter, and finally fused image with Poisson's clone algorithm. This algorithm can filter out facial defects, but Poisson's clone calculation is complex and cannot be processed in real-time. In Liang et al. (2014), use edge preserving filter to calculate a quadratic energy minimum functional equation to decompose the layer, which is divided into smooth layer, brightness layer and colour layer. Then use the prior information of face features, Canny edge detection operator and adaptive editing propagation technology to generate adaptive perceptual mask. The image enhancement effect is good, but it has a large time complexity and is difficult to process video. In Qiu and Dai (2016), use chroma channel histogram to segment skin colour, combine double index edge to keep filtering smooth face, Gauss feather fusion image, and use log curve to whiten face, the effect of beautification is better; however, while removing facial defects, the texture information of skin will be lost. In Zhang et al. (2018), put forward a fast and flexible denoising model fast and flexible denoising network (FFDNet) by taking noise map as network input, the model can deal with different noise levels and spatial correlation noise at the same time. The work of kernel prediction networks (KPN) also shows that if the noise level of the input image is taken as the network input, the resulting network will have better robustness and generalisation ability for a wider range of noise levels (Mildenhall et al., 2018).

In recent years, the convolution network image denoising model based on discrimination learning has also made great progress. However, the existing models usually need network training based on pairs of clear and noise images. In real applications, such as CT, MRI, microscope images, it is usually difficult to obtain a large number of clear images, which greatly limits the practicability of convolution denoising network. In 2018, the authors put forward the Noise2Noise model, and realised image denoising by using noise image pairs that obey the same distribution without clear image (Lehtinen et al., 2018). In Krull et al. (2019), further proposed that noise free image pairs

are not needed, and based on the input noise image itself, combined with blind spot network, self-monitoring training can achieve the same denoising effect as the traditional method. The degradation process of real image often involves not only noise, but also image blur and resolution degradation due to camera shake, object motion and other reasons. Convolution neural network has been successfully applied in image denoising tasks, and has also achieved considerable development in image super-resolution problems, such as super resolution convolution neural network (SRCNN) (Dong et al., 2015), very deep network for super resolution (VDSR) (Kim et al., 2016), super resolution residual network (super resolution residual network, SRResNet) (Ledig et al., 2017). However, the existing CNN model usually needs to set up training network for the specific fuzzy kernel and noise level in the degradation model, which is not only lack of flexibility, but also unable to cope with the more general image restoration task. Therefore, how to use model flexibly in image restoration has more important research value.

3 Smooth denoising

Image smoothing refers to the methods that highlight the low frequency, primary components of an image; or the methods that suppress interferences (noises and the high-frequency components of the signal). This kind of image processing aims for the reduction of image noises, rendering a smooth gradual change of image brightness, reduction of abruptness of the gradient, and enhancing quality of the image. In the following, a brief introduction is given to the mean-shift, EMD, and beeps filter algorithms, which are all popular smooth denoising methods at present.

3.1 The mean-shift algorithm

In 2007, the authors pointed out the errors made by Comaniciu in proving the convergence of the mean-shift algorithm; they re-proved the local convergence of the algorithm and laid a ground work for the in-depth development of the mean-shift algorithm (Wen and Cai, 2007). In the field of imaging processing, it is often required to analyse the data in a multi-dimensional space, which implies that multivariate kernel functions are need to perform non-parametric density estimation. Let X be a d-dimensional Euclidean space and x be a point in the space, which can be represented by a vector. The norm of x is $||x||^2 = x^T x$. Let R denote the real domain. When performing non-parametric density estimation, there are two ways to generate the multivariate kernel function:

$$K^P(x) = \prod_{i=1}^{d} k_1(x) \tag{1}$$

$$K^S(x) = a_{k,d} k_1(\|x\|) \tag{2}$$

where the multivariate kernels $K^P(x)$ and $K^S(x)$ are both functions from X to R; $k_1(x)$ is a univariate kernel function. The function $K^P(x)$ is a production of univariate kernel functions. The function $K^S(x)$ originates from a rotation of $k_1(x)$ in the space R^d; it is thus radial symmetric. The constant $a_{k,d}$ is used to ensure that the integration of $K^S(x)$ is 1.

Radial symmetric functions, which must satisfy the condition are more suitable to be used in computer vision.

$$K(x) = c_{k,d} k(\|x\|^2) \tag{3}$$

where $K(x)$ is a radial symmetric kernel function; $k(x)$, with $x \geq 0$, is called the contour function of $K(x)$; The normalising constant $C_{k,d}$, which makes the integration of $K(x)$ be 1, is strictly positive. Therefore, $k(x)$ is actually a section function of $K(x)$.

As mentioned in the above, x is a point in the d-dimensional Euclidian space X. Now let $k_h(x)$ be a kernel function of the space. The estimated probability density at the point x in the R^d-space is:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} k_h (x - x_i) \tag{4}$$

The formula used by the mean-shift algorithm to computing the eigen probability density is:

$$\hat{f}_{h,k}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\left\| \frac{x - x_i}{h} \right\| \right) \tag{5}$$

Figure 1 Mean shift vectors

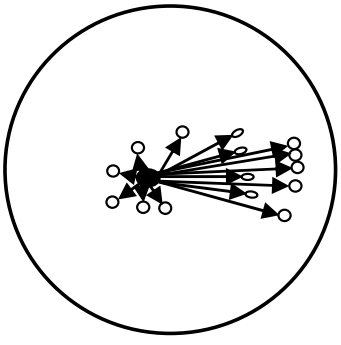


Figure 1 illustrates the mean shift vectors. The centre x of the kernel function is represented by the solid dot; the sampling points x_i are represented by the open dots; the arrows indicate the deviations of the sampling points from the centre x of the kernel function. The physical meaning of the arrows is that the average deviation would point to the densest sampling point, which is the direction of density gradient.

The mean-shift smoothing algorithm adopts a differentiable kernel function, using the gradient of the density estimation operator to estimate the density gradient,

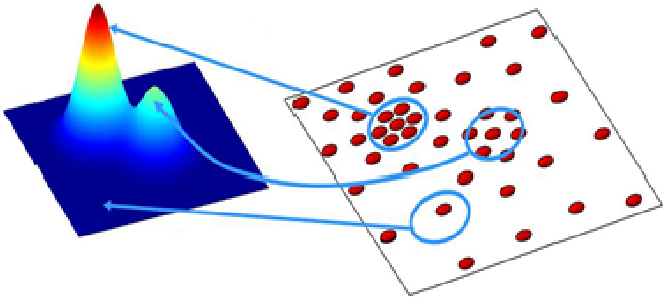
$$\nabla f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} \nabla k\left(\frac{x - x_i}{h} \right) \tag{6}$$

Here the Epanechnikov kernel function is used,

$$K_E(x) = \begin{cases} (d+2)(1-x^T x)/(2c_d); & \text{if} \quad x^T x < 1 \\ 0; & \text{otherwise} \end{cases} \tag{7}$$

where c_d represents the volume of the unit hypersphere in the d-dimensional space. Figure 2 gives the density estimation function of a two-dimensional random dataset.

Figure 2 Density estimation function of random datasets (see online version for colours)



The Epanechnikov kernel function is isotropic, having the same weight coefficient for every parameter of the eigenvector, which allows us to re-write equation (6) into:

$$\begin{aligned} \nabla f(x) &= \frac{1}{n(h^d c_d)} \frac{d+2}{h^2} \sum_{x_i \in s_h(x)} (x_i - x) \\ &= \frac{n_x}{n(h^d c_d)} \frac{d+2}{h^2} \left(\frac{1}{n_x} \sum_{x_i \in s_h(x)} (x_i - x) \right) \end{aligned} \tag{8}$$

where $S_h(x)$ represents the hypersphere centred at x, with radius h and volume $h^d c_d$ in the d-dimensional eigenspace. The hypersphere includes n_x data points. We define:

$$M_h(x) = \frac{1}{n_x} \sum_{x_i \in s_h(x)} (x_i - x) \tag{9}$$

where $(x_i - x)$, known as the mean shift vector, can be understood as the deviation of the sampling point x_i from the centre x. By the function $M_h(x)$, all the deviations (of the k sampling points enclosed by the area S_h) from the centre x are summed up and then averaged. If the sampling points x_i follow the probability distribution function $f(x)$, then the sampling points within the area S_h tend to fall along the direction of the gradient of the probability density. The term $\frac{n_x}{n(h^d c_d)}$ in equation (8) is the kernel density estimation obtained from the hypersphere $S_h(x)$. From,

$$\nabla f(x) = f(x) \frac{d+2}{h^2} M_h(x) \tag{10}$$

One obtains:

$$M_h(x) = \frac{h^2}{d+2} \frac{\nabla f(x)}{f(x)} \tag{11}$$

Because the density gradient of the Epanechnikov kernel function at point x has the same direction as the mean-shift vector, one can obtain the direction of the gradient by calculating the mean-shift vector, which points to the direction of maximal density.

Once the mean-shift vector $M_h(x)$ is obtained, one can perform a parallel shift of the current window according to the size and direction of the mean-shift vector, and then do the subsequent computations until convergence. The iterative mean-shift algorithm can be expressed by:

$$y_{k+1} = \frac{1}{n_k} \sum_{x_i \in S_h(y_k)} x_i \tag{12}$$

After the $(k + 1)^{\text{th}}$ iteration, only those data points change their positions that are enclosed by the circle centred at y_k with a radius h. For a given image $I(i, j)$, a dataset can be generated by including every pixel's coordinates and its normalised intensity, namely $(i, j, I(i, j) * c)$. The smoothing algorithm is as follows:

1 Set the normalisation constant

$$c = \frac{height + width}{2} \times \frac{1}{255} \tag{13}$$

2 Initialise the dataset

$$I(i, j) \to (i, j, I(i, j) * c) \quad j = 1, 2, \cdots, n \tag{14}$$

3 Let $k = 1$ and let $y_k = x_i$.

4 Calculate y_{k+1} by the mean-shift process until convergence.

5 Assignment operation

$$I_{smoothed}\left(x_j(1), x_j(2) = y_k(3)\right) \tag{15}$$

3.2 The EMD algorithm

Empirical mode decomposition (EMD) was put forward by Huang et al. (1998) and Huang and Shen (2005), on the basis of their in-depth study of 'instantaneous frequency'. By EMD one needs not to set the basis functions. Using the local characteristics and time scale information, one can decompose the observed signal into multiple IMFs, each of which carries the local characteristics of the signal. By analysing the local characteristics one can better understand and interpret the characteristic information of the signal, which is apparently very advantageous in processing nonlinear signals and can greatly enhance signal-to-noise ratio. Therefore, EMD has been widely used in smooth denoising of images.

The denoising methods are primarily based on wavelet analysis. The image is first converted into a one-dimensional signal to be input into EMD, which yields multiple IMF components. The signal decomposition can be considered as a data screening process. It goes as follows. One first converts the original signal $x(t)$ [Figure 3(a)] into a new signal $h_1(t)$ [Figure 3(c)]. The original signal $x(t)$ is bounded and thus the upper and lower envelopes can be traced out [Figure 3(b)]. Let the mean of the upper and lower envelopes be $m_1(t)$. Then $h_1(t)$ is obtained by subtracting $m_1(t)$ from $x(t)$:

$$x(t) - m_1(t) = h_1(t) \tag{16}$$

If this first result is unsatisfactory, then the screening process can be repeated by letting $h_1(t)$ be the original signal. The average of the upper and lower envelops of $h_1(t)$ is denoted by $m_{11}(t)$. The second result ($h_{11}(t)$) is obtained by:

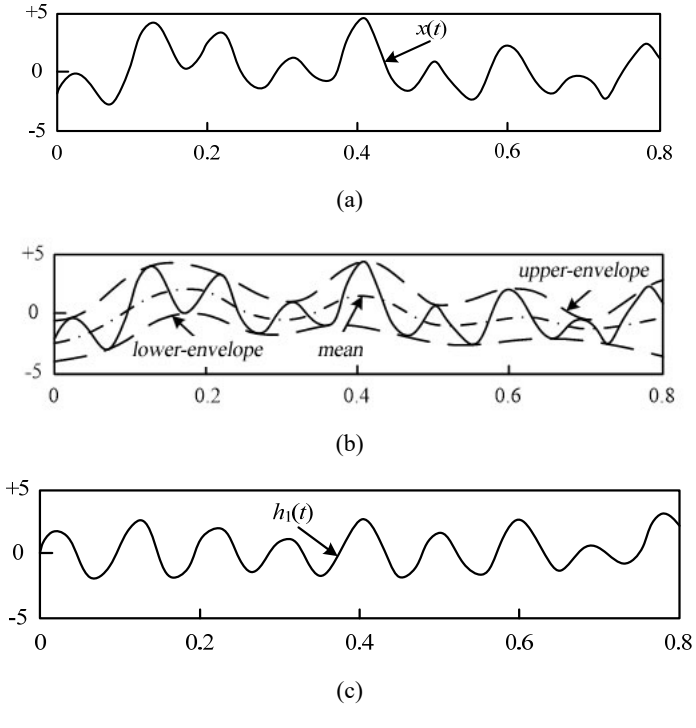
$$h_1(t) - m_{11}(t) = h_{11}(t) \tag{17}$$

By iterating such screenings many times, one finds that the k^{th} result $h_{1k}(t)$:

$$h_{1(k-1)}(t) - m_{1k}(t) = h_{1k}(t) \tag{18}$$

is satisfactory. This $h_{1k}(t)$ is used as the first order IMF. Let $c_1(t)$ denote the first order IMF of $x(t)$, namely $c_1(t) = h_{1k}(t)$.

Figure 3 Decomposition in EMD, (a) the original signal (b) obtaining the upper envelop, the lower envelop, and their average $m_1(t)$ (c) the new signal $h_1(t)$ is obtained by subtracting $m_1(t)$ from $x(t)$



3.3 Beeps filter

Beeps is the abbreviation for bi-exponential edge-preserving smoother, which is the smoothing filter put forward by Thevenaz et al. (2012). It can smooth images and is especially effective in removing freckles, wrinkles, and pimples in the image. The filter first performs a horizontal iterative computation for the image data $I_{ori}(x)$, and then performs a vertical iterative computation for the data, which finally obtains the processed image $I_{hv}(x)$:

$$I_{hv}(x) = B_{hv}\left(I_{ori}(x)\right) \tag{19}$$

Likewise, one can first performs a vertical iterative computation for the image data $I_{ori}(x)$, and then performs a horizontal computation for the data, which obtains the processed image $I_{vh}(x)$:

$$I_{vh}(x) = B_{vh}\left(I_{ori}(x)\right) \tag{20}$$

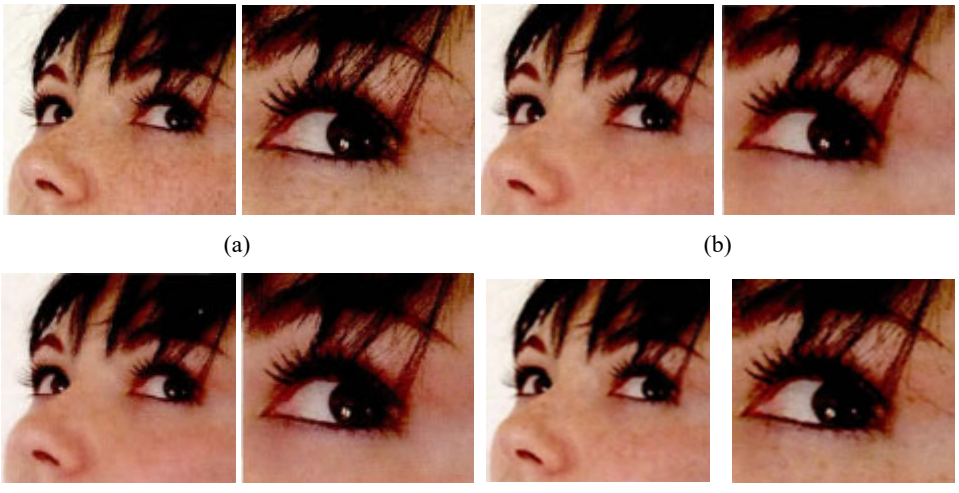
By a weighted combination of equations (19) and (20), the two images merge into the final image.

$$I(x) = \left(I_{hv}(x) + I_{vh}(x)\right)/2 \tag{21}$$

3.4 Analysis of the experimental results

To compare smoothing performance of the mean-shift, the EMD, and the beeps filter algorithms, we performed simulations with an ordinary personal computer (Intel Core i7 CPU, 3.60 GHz, 8 G RAM) with MATLAB 2017. The test images were from the internet. The results are presented in Figure 4.

Figure 4 Smoothing by the three algorithms, (a) the input image (b) the image smoothed by the EMD algorithm (c) the image smoothed by the beeps filter (d) the image smoothed by the mean-shift algorithm (see online version for colours)



(a) (b)

(c) (d)

From Figure 4 one sees that the mean-shift algorithm can moderately remove noises in the face, while preserving details such as hair and the background of eyes. The performances of the other two methods are not ideal. The EMD algorithm can smooth the face, but the detailed hair information is lost. The non-ideal performance may be due to the wavelet-based denoising, while simple and easy to implement, has certain limitations: if the threshold value is too small, the algorithm cannot effectively remove the noises;

if the threshold value is too large, the details of the original image are also removed. Therefore, the EMD algorithm has the shortcoming of overly relying on the range of the threshold [Figure 4(b)]. The beeps filter smoothing is a little overkill, because the iteration and weighted processing are performed twice, one along the horizontal direction and the other along the vertical direction, which would lead to the loss of important local features of the face, e.g., the freckles. It is thus not suitable for face detection [Figure 4(c)]. Smoothing by the mean-shift algorithm, on the other hand, can greatly enhance the contrast of the image. And it is a real-time process [Figure 4(d)].

4 Skin-colour segmentation and enhancement

4.1 Skin-colour detection and segmentation

In a YC_bC_r colour space, if the influence from the brightness distribution can be ignored, then the two chroma components, C_b and C_r, tend to have consistent distributions; and they together manifest a two-dimensional Gaussian distribution. Through computing the similarity of every pixel of the colour image, one can determine for a pixel the probability of it belonging to a specific skin-colour area.

$$m = E(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{22}$$

The two-dimensional skin-colour Gaussian model is denoted by $G = (m, C)$, where is the mean; represents the covariance matrix; $x_i = (C_b, C_r)^T$ is the value of skin-colour pixel i in the training sample; n is the number of pixels in the sample.

$$C = E\left[(X-m)(X-m)^T\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i-m)(x_i-m)^T \tag{23}$$

This Gaussian skin-colour model allows for the calculation of the probability of a pixel belonging to a certain skin-colour and the generation of the skin-colour likelihood image.

$$p(C_b, C_r) = \exp\left[-0.5(x-m)^T C^{-1}(x-m)\right] \tag{24}$$

The skin area of image is relatively bright in the C_r channel and is relatively dark in the C_b channel. To avoid noise interference, one can perform morphology closed and open operations respectively in the C_r and C_b channels, which obtains the normalised histograms $H(x_{Cr})$ and $H(x_{Cb})$ and the corresponding cumulative distribution functions $C(x_{Cr})$ and $C(x_{Cb})$. Based on experience, the cumulative distribution of C_r is [0.26, 0.96] and that of C_b is [0.04, 0.74].

To validate this method, we tested both single and multiple images (Figure 5). One sees that the Gaussian skin-colour model, when combined with morphology, can realise skin-colour segmentation while avoiding the existence of isolated points.

Figure 5 Skin-colour segmentation on single or multiple images (see online version for colours)

4.2 Image enhancement

Smooth denoising of face image has the inherent problem of overly smoothing that often destroys the face edge details adjacent to the hairs. To protect such detailed information, it is necessary to fuse the smoothed image (by the mean-shift algorithm) and the mask (created from the original image by skin region segmentation). To do it in real-time, one can use the method of Gaussian fitting-based feather blending. The mask is first created by Gaussian fitting-based feathering of the skin region, which needs to repeat multiple times to obtain the mean. Then Gaussian blurring is performed three times for the mean blurred image. These processes create the mask through feathering the skin area P_{sk}.

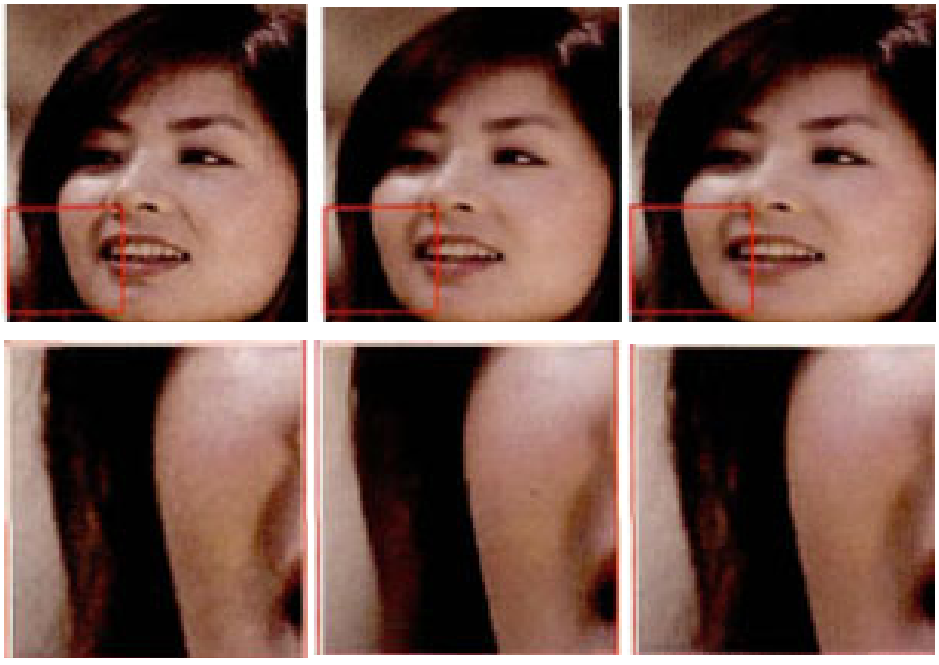
$$M_{sk} = 3a_{Blur}\left(P_{sk}\right) \tag{25}$$

where a_{Blur} is the fuzzy function after triple averaging. The original image and the smoothed image are then fused:

$$I' = \left(1 - M_{sk}\right) I_{ori} + M_{sk} I \tag{26}$$

where I_{ori} is the original image; I is the smoothed image; and I' is the fusion image (Figure 6). By comparing Figures 6(a), 6(b) and 6(c), one sees that the fusion image appears to be natural (without artificial processing) but at the same time keeps important details of the non-skin region. The algorithm is very effective.

Figure 6 Comparison among the original, smoothed, and enhanced image, (a) the original image (b) the smoothed image (c) the enhanced image (see online version for colours)



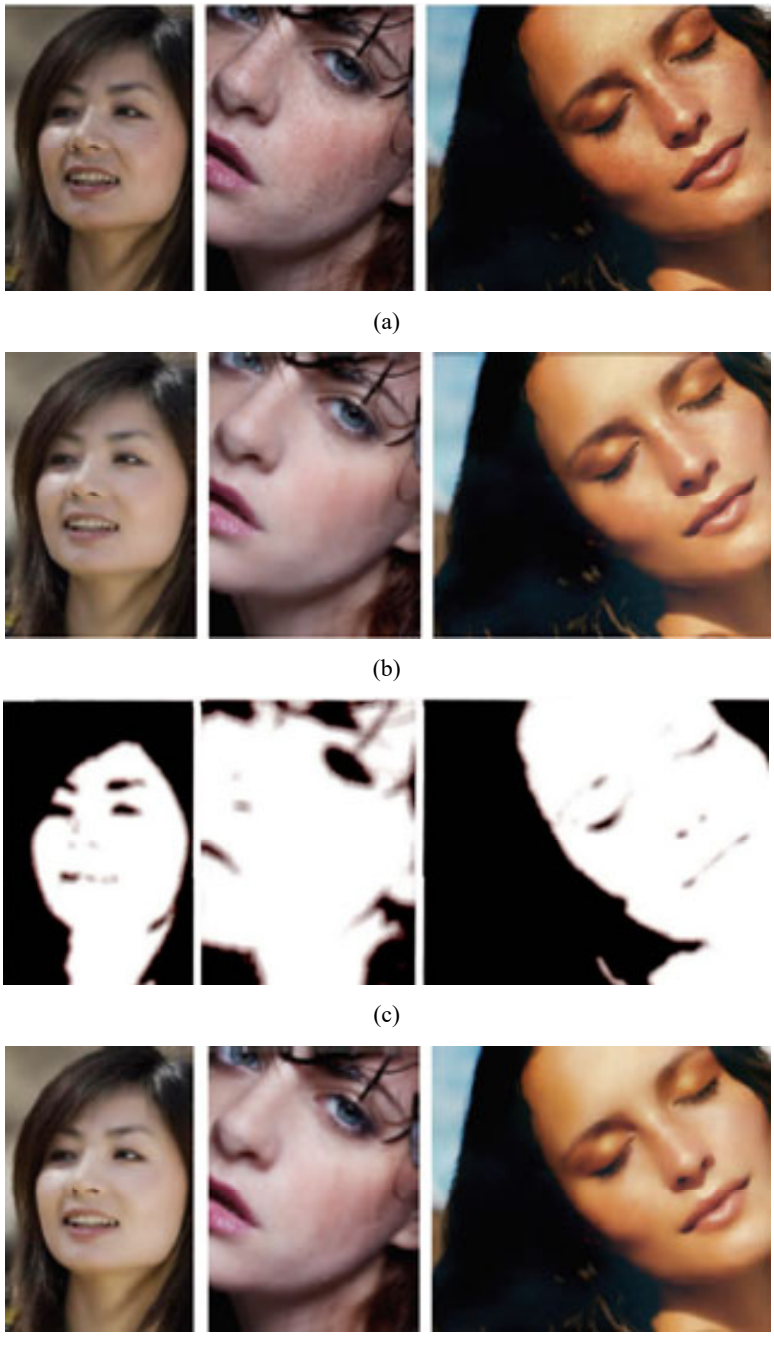
(a) (b) (c)

4.3 Enhancement experiments and their analysis

To test effectiveness of the fusion algorithm, we processed three test images obtained from the internet with an ordinary personal computer (Intel Core i7 CPU, 3.60 GHz, 8 G RAM) with MATLAB 2017. The results are presented in Figure 7.

The main reasons why the present algorithm can achieve moderate smoothing are as follows. First, the mean-shift algorithm can effectively realise smooth denoising, which preserves detailed information such as the face edge and improves facial contrast. Second, the morphological processing by Gaussian skin-colour segmentation can avoid the loss of details of the non-skin regions such as the background hairs, which makes the enhanced images more realistic, leaving no trace of artificial processing.

Figure 7 Enhanced images for face, (a) the original image (b) the images smoothed by the mean-shift algorithm (c) the masks (d) the enhanced images (see online version for colours)



(a)

(b)

(c)

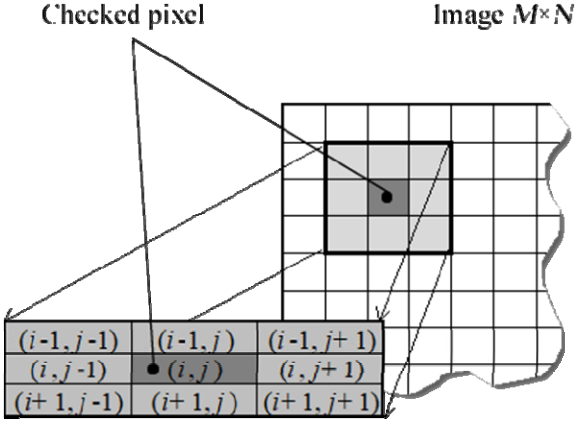
(d)

5 Edge detection of images

5.1 Edge detection based on the eight neighbourhood gradient operators

The main shortcoming of the traditional edge detection methods such as Roberts, Prewitt, Sobel, and Canny operators is the sensitivity to noises and the low precision of edge location. The Canny operator has relatively better performance of edge detection – it can detect weak edges – but problems can easily emerge such as the false edges and the loss of details of true edges, because the operator compute the magnitude and direction of the gradient by finite difference averaging only in the 2 × 2 neighbourhood.

Figure 8 The eight neighbourhood



Therefore, we propose to perform edge detection by computing the gradient in an eight neighbourhood. In an image, an arbitrary point always has eight neighbouring points. If the present point $G(i, j)$ is an edge point, then the next edge point must be with the eight neighbourhood of $G(i, j)$ (Figure 8). Using the eight neighbourhood, the noises can be further reduced and the edge of the image can be accurately localised.

In an eight neighbourhood, one first calculate the first order partial derivative of the point $I(i, j)$ along the x and y directions, respectively:

$$G_x(i, j) = \frac{\begin{array}{c} I(i, j+1) - I(i, j-1) + I(i-1, j+1) - I(i-1, j-1) \\ +I(i+1, j+1) - I(i+1, j-1) \end{array}}{2} \tag{27}$$

$$G_y(i, j) = \frac{\begin{array}{c} I(i+1, j) - I(i-1, j) + I(i+1, j-1) - I(i-1, j-1) \\ +I(i+1, j+1) - I(i-1, j+1) \end{array}}{2} \tag{28}$$

which determines the direction of the gradient vector.

$$\theta(i, j) = \arctan\frac{G_y}{G_x} \tag{29}$$

along which the magnitude of the gradient can be determined.

$$|\nabla| M(i, j) = \sqrt{G_x^2 + G_y^2} \tag{30}$$

The steps of edge detection by the eight neighbourhood gradient operator are as follows:

1 Sweep the image from top down and from left to right for every pixel in the image.

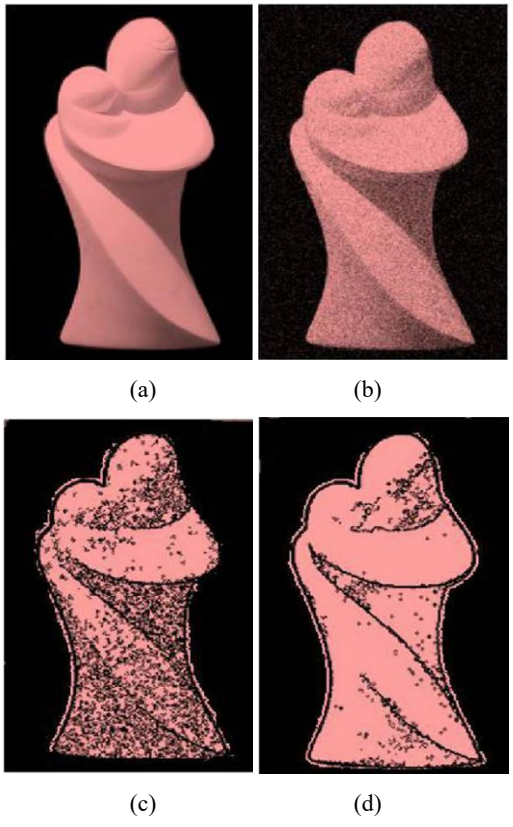
2 If the grey value of the pixel is zero, as well as all the eight neighbouring points, then the pixel is an interior one. Delete the pixel (i.e., set the grey value to 255). If all the eight neighbouring points are not zero, then the pixel is isolated noise point. Delete the pixel. Otherwise the pixel is an edge point. Keep the pixel.

3 Keep all the background points.

5.2 Edge detection experiments and their analysis

To highlight the denoising effects of the eight neighbourhood operator, we processed a simple image of sculpture (Figure 9) by two different algorithms (the Canny operator and the eight neighbourhood operator) and with the software MATLAB 2017. We found that the eight neighbourhood operator is better than the Canny operator in terms of both image details preservation and denoising.

Figure 9 Smooth denoising for a sculpture image by different algorithms, (a) original sculpture (b) noisy image (c) Canny operator (d) eight neighbourhood operator (see online version for colours)



(a) (b)

(c) (d)

Figure 10 Edge detection on a pre-processed standard test image, (a) noisy image (b) smoothed image (c) enhanced image (see online version for colours)

(a) (b) (c)

To prove that the fusion images contain more effective edge information, we used the eight neighbourhood operator to detect the edges of two sets of images. The first set is about a standard human face test image (Figure 10). The upper left one is the original image with Gaussian noise (0.001) added. The upper middle and upper right images are the smoothed and fusion images, respectively. We then used the eight neighbourhood operator to detect the edges for the three upper images. The results are presented in the corresponding lower positions. The second set is about an image obtained from the internet (Figure 11). Following the same procedures as the first set, the image was pre-processed (noise addition, smoothing, fusing), followed by edge detection.

We found that the quality of edge detection depends on the pre-processing of the image. For the noisy image, the edge detection obtains bad contrast and fuzzy details [Figures 10(a) and 11(a)]. For the smoothed image, the edge detection obtains better contrast but the edge is still fuzzy [Figure 10(b) and 11(b)]. For the enhanced image, the edge detection obtains good contrast and sharpness along the border between the face and the hairs; it provides more detailed and effective information about the face edge [Figure 10(c) and 11(c)].

Figure 11 Edge detection on a pre-processed image from internet, (a) noisy image (b) smoothed image (c) enhanced image (see online version for colours)

(a) (b) (c)

6 Conclusions

In this paper, an image enhancement based on skin-colour segmentation and smoothness method is presented by computing the magnitude of target gradients. By comparing the performance of the mean-shift, EMD, and beeps algorithms on smooth denoising, we conclude that the mean-shift algorithm is better for removing noises in images. By using the Gaussian blurring to render skin-region feathering, the mask of the original image can be obtained. A enhanced image, with the details about the hairs preserved and with the face edge details enhanced, can then be obtained by merging the smoothed image and the mask. Finally, we perform edge detection on the enhanced image by using the octet neighbourhood operator, which proves that the enhanced image can also provide more effective information for the detection of edge characteristics. In the future, the depth model is not a black box, it is closely related to the traditional computer vision system, the various layers of neural network through joint learning, overall optimisation, so that the performance has been greatly improved. Applications related to image recognition are also driving the rapid development of deep learning in all aspects of network structure, layer design and training methods. It can be expected that in the next few years, deep learning will enter a period of rapid development in theory, algorithms and applications.

Acknowledgements

We thank the anonymous referees for their constructive comments, which helped to improve this paper.

This work is supported by Zhanjiang Science and Technology Project (Grant No. 2019B01076), Lingnan Normal University Nature Science Research Project (No. ZL2004), Yanling Excellent Young Teacher Program of Lingnan Normal University, Science and Technology Program of Maoming (No. 2019397) and Scientific Research Foundation for Talents Introduction, Guangdong University of Petrochemical Technology (No. 519025).

References

Burger, H.C., Schuler, C.J. and Harmeling, S. (2012) 'Image denoising: can plain neural networks compete with BM3D', in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2392–2399.

Chen, C.W., Huang, D.Y. and Fuh, C.S. (2010) 'Automatic skin color beautification', *Lecture Notes in Computer Sciences*, Vol. 30, pp.157–164, Springer, Heidelberg.

Chen, Y. and Pock, T. (2015) 'Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration', *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 39, No. 6, pp.1256–1272.

Dong, C., Loy, C.C., He, K. and Tang, X. (2015) 'Image super-resolution using deep convolutional networks', *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 38, No. 2, pp.295–307.

Hao, H., Wang, H.L. and Rehman, N.U. (2017) 'A joint framework for multivariate signal denoising using multivariate empirical mode decomposition', *Signal Processing*, Vol. 135, No. 6, pp.263–273.

Huang, N.E. and Shen, S.S.P. (2005) 'Hilbert-Huang transform and its applications', *World Scientific*, pp.1–209.

Huang, N.E., Shen, Z. and Long, S.R. (1998) 'The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis', *Proceedings of the Royal Society of London Series A – Mathematical Physical and Engineering Sciences*, Vol. 454, pp.903–995.

Jain, V. and Seung, S. (2009) 'Natural image denosing with convolutional networks', *Advanced in Neural Information Processing Systems (NeurIPS)*, pp.769–776.

Kim, J., Lee, J.K. and Lee, K.M. (2016) 'Accurate image super-resolution using very deep convolutional networks', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1646–1654.

Krull, A., Buchholz, T-O. and Jug, F. (2019) 'Noise2Void-learning denoising from single noisy images', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2129–2137.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A. et al. (2017) 'Photo-realistic single image super-resolution using a generative adversarial network', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4681–4690.

Lee, C., Schramm, M.T., Boutin, M. et al. (2009) 'An algorithm for automatic skin smoothing in digital portraits', *Proceedings of the 16th International Conference on Image Processing*, IEEE Computer Society Press, Los Alamitos, pp.3113–3116.

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M. and Aila, T. (2018) 'Noise2Noise: learning image restoration without clean data', *Proceedings of the 35th International Conference on Machine Learning (ICML)*.

Liang, L.Y., Jin, L.W. and Li, X.L. (2014) 'Facial skin beautification using adaptive region-aware masks', *IEEE Transactions on Cybernetics*, Vol. 44, No. 12, pp.2600–2612.

Liu, P., Zhang, H., Zhang, K. et al. (2018) 'Multi-level wavelet-CNN for image restoration', DOI: 10.1109/CVPRW.2018.00121.

Mao, X., Shen, C. and Yang, Y-B. (2016) 'Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections', *Advances in Neural Information Processing Systems (NeurIPS)*, pp.2802–2810.

Mildenhall, B., Barron, J.T., Chen, J., Sharlet, D., Ng, R. and Carroll, R. (2018) 'Burst denoising with kernel prediction networks', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2502–2510.

Qiu, J. and Dai, S. (2016) 'Fast facial beautification algorithm based on skin-color segmentation and smoothness', *Journal of Image and Graphics*, Vol. 21, No. 7, pp.865–874.

Schmidt, U., Jancsary, J., Nowozin, S. et al. (2014) 'Cascades of regression tree fields for image restoration', *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 38, No. 4, pp.677–689.

Suzuki, K., Horiba, I., Sugie, N. et al. (2002) 'Efficient approximation of neural filters for removing quantum noise from images', *IEEE Transactions on Signal Processing*, Vol. 50, No. 7, pp.1787–1799.

Tai, Y., Yang, J., Liu, X. and Xu, C. (2017) 'MemNet: a persistent memory network for image restoration', in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp.4539–4547.

Thevenaz, P., Sage, D. and Unser, M. (2012) 'Bi-exponential edge-preserving smoother', *IEEE Transactions on Image Processing*, Vol. 21, No. 9, pp.3924–3936.

Wen, Z-Q. and Cai, Z-X. (2007) 'Convergence analysis of mean shift algorithm', *Journal of Software*, Vol. 18, No. 2, pp.205–212, China.

Xie, J., Xu, L. and Chen, E. (2012) 'Image denoising and inpainting with deep neural networks', in *Advances in Neural Information Processing Systems (NeurIPS)*, pp.341–349.

Zhang, K., Zuo, W. and Zhang, L. (2018) 'FFDNet: toward a fast and flexible solution for CNN-based image denoising', *IEEE Transactions on Image Processing (TIP)*, Vol. 27, No. 9, pp.4608–4622.

Zhang, K., Zuo, W., Chen, Y., Meng, D. and Zhang, L. (2017) 'Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising', *IEEE Transactions on Image Processing (TIP)*, Vol. 26, No. 7, pp.3142–3155.