

# Photographic Appearance Enhancement via Detail-based Dictionary Learning

Zhi-Feng Xie<sup>1,3</sup> (谢志峰), Shi Tang<sup>1</sup> (唐仕), Dong-Jin Huang<sup>1,3</sup> (黄东晋), You-Dong Ding<sup>1,3</sup> (丁友东), *Member, CCF*, and Li-Zhuang Ma<sup>2,3</sup> (马利庄), *Member, CCF*

<sup>1</sup> *Department of Film and Television Engineering, Shanghai University, Shanghai 200072, China*

<sup>2</sup> *Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

<sup>3</sup> *Shanghai Engineering Research Center of Motion Picture Special Effects, Shanghai 200072, China*

E-mail: zhifeng\_xie@shu.edu.cn; tangshimail@163.com; {djhuang, ydding}@shu.edu.cn; ma-lz@cs.sjtu.edu.cn

Received December 15, 2016

Revised January 25, 2017

**Abstract** A number of edge-aware filters can efficiently boost the appearance of an image by detail decomposition and enhancement. However, they often fail to produce photographic enhanced appearance due to some visible artifacts, especially noise, halos and unnatural contrast. The essential reason is that the guidance and constraint of high-quality appearance aren't enough sufficient in the process of enhancement. Thus our idea is to train a detail dictionary from a lot of high-quality patches in order to constrain and control the entire appearance enhancement. In this paper, we propose a novel learning-based enhancement method for photographic appearance, which includes two main stages: dictionary training and sparse reconstruction. In the training stage, we construct a training set of detail patches by extracting from some high-quality photos, and then train an overcomplete detail dictionary by iteratively minimizing an  $\ell_1$ -norm energy function. In the reconstruction stage, we employ the trained dictionary to reconstruct the boosted detail layer, and further formalize a gradient-guided optimization function to improve the local coherence between patches. Moreover, we propose two evaluation metrics to measure the performance of appearance enhancement. The final experimental results have demonstrated the effectiveness of our learning-based enhancement method.

**Keywords** image enhancement, dictionary learning, edge-aware filter

## 1 Introduction

Detail enhancement is a fundamental image processing technique which can enhance the appearance of an image by extracting and boosting its hidden detail layer. In the past

decade, a number of edge-aware filters have been proposed to decompose an image into base layer and detail layer. A high-quality enhanced appearance not only depends on the quality of detail decomposition but also often needs to man-

---

This work was supported by the National Natural Science Foundation of China (No. 61303093, 61402278, 61472245), the Innovation Program of the Science and Technology Commission of Shanghai Municipality (No. 16511101300), the Natural Science Foundation of Shanghai (No. 14ZR1415800), and the Gaofeng Film Discipline Grant of Shanghai Municipal Education Commission.

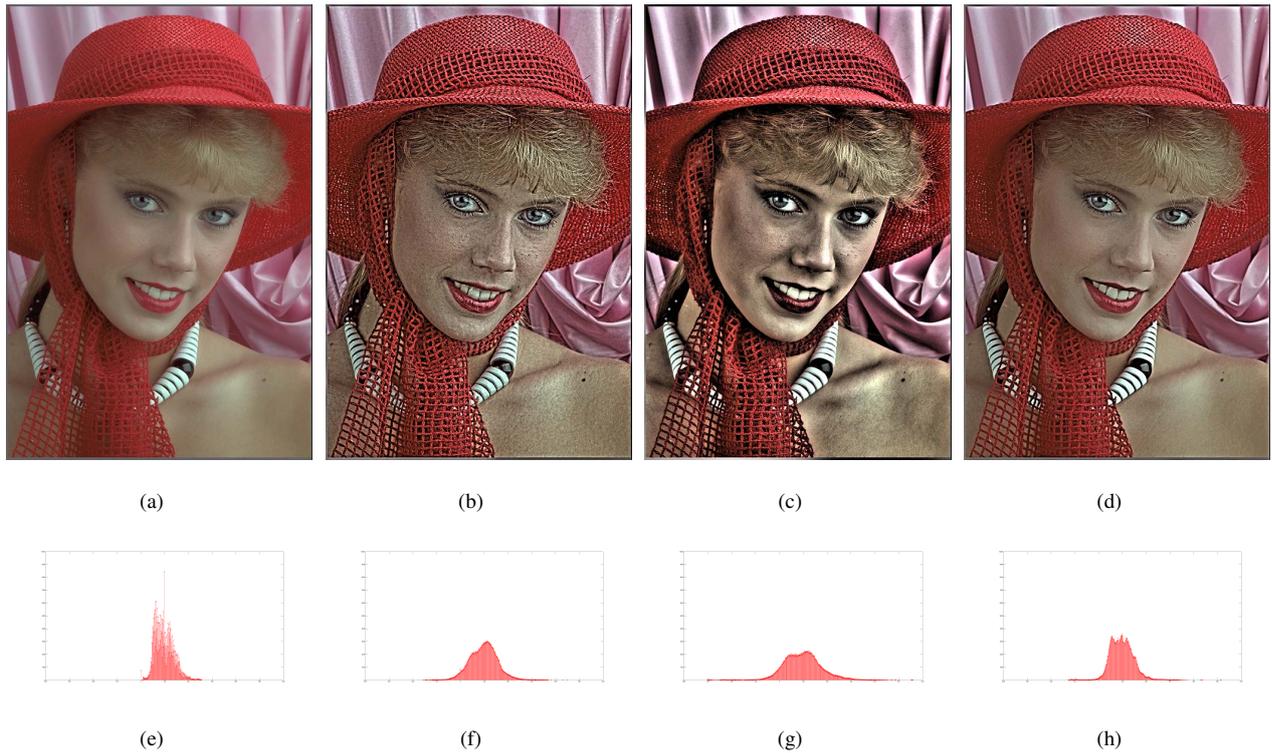


Fig. 1. Photographic appearance enhancement. (a) Source image. (b) The enhanced result by bilateral filter [1] ( $w = 5, \sigma = [16, 0.1]$ ). (c) The enhanced result by guided filter [2] ( $r = 16, \epsilon = 0.01$ ). (d) Our reconstructed result on the detail layer of (c). (e) Histogram of (a). (f) Histogram of (b). (g) Histogram of (c). (h) Histogram of (d). Note: The edge-aware filters employ a fixed scale factor ( $n = 5$ ) to boost their detail layers and perform a unified detail enhancement.

usually find out the optimal enhancement parameters. However, the manual adjustment is hard to yield a photographic appearance because of some visible appearance artifacts, including noise, halos and unnatural contrast. As shown in Fig.1(b), we employ a state-of-the-art bilateral filter [1] to extract the detail layers and boost it using a scale factor, but their scale-based enhanced results easily suffer from the severe issues about noise, edge and contrast. In fact, these issues are derived from a lack of reasonable detail constraint in the process of enhancement.

The recent guided filter [2] introduces a guidance image to constrain and control the detail decomposition, especially the gradient reversal of bilateral filter [1]. As shown in Fig.1(c), its filtering output is locally a linear transform of the guidance image, which can effectively reduce the noise, edge and contrast issues. But the guidance image is

often the filter input itself and its guided info does not derive from some real photos with high-quality appearance. Thus its guidance constraint is not completely appropriate to photographic enhancement and its scale-based enhanced results cannot still get rid of unnatural appearance. Unfortunately, similar with the above state-of-the-art filters, we observe that most of other popular filters [3–6] also fail to fully guarantee the appearance quality in detail enhancement.

As a powerful learning technique, dictionary learning (sparse representation) [7, 8] has been widely applied to image restoration, super-resolution, denoising, and so on. The learning-based method assumes that each small patch can be represented sparsely in an overcomplete dictionary learned from a lot of training patches. Therefore, the appearance constraint of training patches can be effectively introduced into the reconstruction of new patches by dictionary learn-

ing. To improve the appearance fidelity, our idea is to perform the learning-based detail reconstruction instead of the traditional scale-based enhancement. More specifically, given a low-quality input image, we employ an overcomplete detail dictionary learned from a lot of high-quality image patches to reconstruct the detail layer of its unknown enhanced image. Obviously, with the guidance of high-quality appearance, the new detail reconstruction will be more beneficial to generating a high-fidelity enhanced result.

In this paper, we propose a novel appearance enhancement method based on the above dictionary learning pipeline, which includes two main stages: dictionary training and sparse reconstruction. In the training stage, we first employ some advanced cameras with HDR mode (e.g., Canon 5D mark III and Nikon D800) to collect enough high-quality photos. Their HDR mode can take the multi-exposure photos in a scene to compose a high dynamic range image. Then we randomly extract a lot of small patches from these images with photographic appearance and simply compute their local intensity differences as detail patches. Finally, we construct a training set of detail patches and train an overcomplete detail dictionary by iteratively minimizing an  $\ell_1$ -norm energy function. In brief, we hope that the trained dictionary can simulate the high-quality enhanced appearance of advanced cameras.

In the reconstruction stage, we first decompose a low-quality input image into base layer and detail layer by some edge-aware filters. Then we employ a scale factor to boost its detail layer and construct a lot of overlapped detail patches. Next, we compute the sparse coefficients of each detail patch by minimizing an  $\ell_1$ -norm reconstruction function with the trained dictionary and the initially boosted patch. To improve the local coherence between patches, we further formalize a gradient-guided function to optimize the reconstructed detail patches. Finally, we integrate all new detail patches into a complete detail layer and add it to the

input image for a final enhanced result. Compared with their histograms in Figs.1(e)~1(h), the detail constraint of high-quality photo is reasonably involved into appearance enhancement. The original artifacts of noise, edge and contrast are effectively eliminated in Fig.1(d).

Furthermore, we propose two evaluation metrics with no-reference and reference quantitative assessments to verify the performance of our method. The metrics can compute the EME (The Measure of Enhancement) and PSNR (Peak Signal to Noise Ratio) values to measure the quality of appearance enhancement. The experimental results demonstrate that our new method can successfully solve the original appearance issues in detail enhancement. In summary, the key contributions of this paper include:

- We introduce dictionary learning into detail enhancement, which employs dictionary training and sparse reconstruction to improve the appearance quality in the process of enhancement.
- We propose two objective evaluation metrics based on EME and PSNR, which measures the quality of appearance enhancement with no-reference and reference quantitative assessments.

The outline of the paper is as follows. Section 2 briefly reviews existing relevant work. Section 3 presents our new enhancement model. Section 4 and Section 5 introduce dictionary training and sparse reconstruction in our method. Section 6 and Section 7 evaluate the proposed method through two evaluation metrics and a number of experiments. Finally, Section 8 briefly concludes the work and discusses possible future works.

## 2 Related Work

In this paper, some popular techniques, including edge-aware filter, photo enhancement and dictionary learning, are

closely related to our learning-based appearance enhancement.

**Edge-aware Filter** Unlike traditional filters, edge-aware filters are a group of special image manipulation techniques due to their faithfulness to image structures. These filters can effectively preserve image edges when they decompose an image into base layer and detail layer. Perona and Malik [9] first proposed a new definition of scale-space and realized edge-preserving smoothing using a diffusion process. Tomasi and Manduchi [10] proposed a typical bilateral filter to smooth images while preserving edges, by means of a nonlinear combination of nearby image values. Later the useful bilateral filter has many improved and accelerated versions [1, 11–16].

Farbman *et al.* [3] proposed an alternative edge-preserving smoothing operator, based on the weighted least squares optimization framework, which is particularly well suited for progressive coarsening of images and for multi-scale detail extraction. Fattal [17] proposed an edge-avoiding wavelets to achieve nonlinear data-dependent multiscale edge-preserving image filtering and processing at linear computation times. Subr *et al.* [18] defined detail as oscillations between local minima and maxima, and then proposed a new model for edge-preserving multi-scale image decomposition. Paris *et al.* [4] constructed Laplacian pyramid to achieve edge-preserving smoothing by manipulating gradients at each scale of the pyramid. Gastal and Oliveira [19] proposed an efficient domain-transform-based method to perform high-quality edge-preserving filtering of images and videos in real time. Xu *et al.* [5] proposed a new smoothing operator based on  $\ell_0$ -norm gradient minimization by controlling the number of non-zero gradients in an image.

He *et al.* [2] proposed a fast and non-approximate linear-time filter to generate the high-quality filtering output by considering the content of a guidance image. Li *et*

*al.* [20] proposed an explicit mixed-domain edge-aware filter for image manipulation in real space and DCT transform space. Zhang *et al.* [6] proposed a new framework to filter images with the complete control of detail smoothing under a scale measure, which is based on a rolling guidance implemented in an iterative manner. Wang *et al.* [21] proposed least-squares images as a basis for edge-preserving image smoothing, and Shao *et al.* [22] also proposed a new edge-preserving image decomposition method by joint weighted least squares. Xu *et al.* [23] recently introduced deep convolutional neural network to train and approximate various filters without knowing the original models, which can achieve up to 200x acceleration.

As a fundamental technique, the edge-aware filters have been widely applied to image processing and computer vision, such as image smoothing, detail enhancement, edge extraction, non-photorealistic rendering, and so on. In this paper, we employ several state-of-the-art filters [1–6] to extract and boost the detail layer of an image for appearance enhancement. However, although the original edge structure can well be preserved, their enhanced results are still hard to yield a high-quality photographic appearance due to noise, unnatural contrast, color distortion, halos, and so on. Therefore, we will introduce reasonable detail constraints to better guide and control their enhancement process, and then eliminate visible artifacts to improve the final appearance quality.

**Photo Enhancement** Besides edge-aware filters, there are many other techniques and tools for enhancing photographs. Several popular commercial softwares (Adobe’s Photoshop, Google’s Picasa and Microsoft’s Office Picture Manager) can be used to well achieve a variety of photo enhancement effects. As a typical technique, histogram equalization [24–28] uses the cumulative distribution function of image histogram to achieve image contrast enhancement. Content-aware enhancement methods [29–34] exploit the context of

images to correct and improve the quality of enhancement. Some other methods employ the intelligent editing and processing techniques [35–39] of visual media to enhance the low-quality images and videos. Also worth mentioning are techniques that reproduce artworks [40–44] by simulating artistic or photographic effects from images.

In summary, these tools and techniques can yield various enhancement effects by manipulating image features, such as color, detail, histogram, context, exposure and so on. However, they often need to manually adjust their optimal parameters for high-quality enhancement, and are still difficult to completely eliminate the aforementioned enhancement issues. Thus similar to data-driven content-aware methods [30, 33], we should develop a new idea for detail enhancement to reduce complex human intervention and meanwhile improve enhancement quality.

**Dictionary Learning** Dictionary learning [7, 8] has been proven to be a very powerful learning technique in computer vision and image processing, and it has been successfully applied to image restoration [45, 46], super-resolution [47, 48], denoising [49, 50], and so on. It uses a dictionary of primitive elements learned from the signal and decomposes the signal into these primitive elements. The process basically involves two steps: training dictionary and computing sparse coefficients for representing the signal using the dictionary elements.

In this paper, we first introduce dictionary learning into the detail enhancement based on edge-aware filters. It assumes that each small detail patch can be sparsely represented in an overcomplete detail dictionary learned from a lot of training detail patches. Thus the appearance constraint of training detail patches can be mapped into the boosted detail patches by dictionary learning, which can effectively solve many appearance issues in the process of detail enhancement. When there are a training set of high-quality photos, we will employ the trained dictionary to reconstruct

a new boosted detail layer and produce a final enhanced result with photographic appearance.

### 3 Our Enhancement Model

Generally, most of state-of-the-art methods extract image detail and then boost it to yield an enhanced appearance. A photographic appearance effect depends on how appropriate to obtain and enhance the detail layer, like the production of an advanced camera. Here, a typical enhancement model can be simply defined as a following function:

$$L = I + d^* = I + n \times d, \quad (1)$$

where  $L$  is an enhanced result;  $I$  is a source image;  $d^*$  is a boosted detail layer;  $d$  is an initial detail layer;  $n$  is a scale factor, which is fixed to 5 in this paper. In (1), the initial detail layer can be produced by many edge-aware filters, which can decompose an image into base and detail components. The traditional model employs the scale factor to boost the detail layer, but it is easy to bring some visible appearance artifacts, especially strong noise, uncomfortable edges and unnatural contrast.

To eliminate the artifacts, we tend to reconstruct a photographic appearance with a detail dictionary trained from a lot of high-quality enhanced photos. Thus we assume that the boosted detail patch can be sparsely represented in an appropriately chosen overcomplete dictionary. Based on the assumption, we further propose a new learning-based enhancement model as follows:

$$d_x^* = n d_x \approx \mathbf{D} \alpha_x, \quad (2)$$

where  $x$  is a small patch of image and  $d_x^* \in \mathbb{R}^N$ ,  $N$  is the number of pixels in each small patch  $x$ ;  $\mathbf{D} \in \mathbb{R}^{N \times K}$  is an overcomplete dictionary trained from detail patches sampled from high-quality photos,  $K$  is the number of atoms in the dictionary  $\mathbf{D}$ ;  $\alpha$  is a coefficient vector of sparse representation, for some  $\alpha \in \mathbb{R}^K$  with  $\|\alpha\|_0 \ll K$ . (2) indicates

that the boosted detail patch  $d_x^*$  is represented as a sparse linear combination with respect to the detail dictionary  $D$  with  $K$  atoms ( $K > N$ ), and the unknown coefficients  $\alpha_x$  is a vector with very few ( $\ll N$ ) nonzero entries.

Obviously, the new model focuses on how to solve the sparse coefficients. It is a NP-hard optimization problem because the desired coefficients should be sufficiently sparse. With regard to its difficulty, given a trained dictionary  $D$ , we define an  $\ell_1$ -norm minimization function to constrain the approximation to the boosted detail layer  $nd_x$  and balance the sparsity of the solution:

$$\alpha_x^* = \arg \min_{\alpha_x} \|D\alpha_x - nd_x\|_2^2 + \lambda \|\alpha_x\|_1, \quad (3)$$

where  $\alpha_x^*$  is the optimal sparse coefficients;  $\lambda$  is a balance parameter between sparsity and fidelity, which is fixed to 0.1 in this paper. After the optimal solution of coefficients is performed, the new boosted detail layer can be efficiently reconstructed as  $d_x^* = D\alpha_x^*$ . In (3), the optimization and reconstruction must have an overcomplete dictionary pre-trained from a lot of sampled detail patches. Thus we will explain the training process of the key dictionary in Section 4.

#### 4 Dictionary Training

According to (2), the prerequisite of dictionary training is a lot of high-quality enhanced images. A conventional method is to manually boost low-quality images and then yield their enhanced results. However, the manual enhancement is very difficult to preserve a unified quality, and thus their enhanced results are artificial and unsatisfactory to construct an ideal training set. Fortunately, there are some advanced cameras (e.g., Canon 5D Mark III and Nikon D800) with HDR mode, which can simultaneously take several photos with different exposure levels and then integrate them into a higher-quality HDR photo. In this way, we can collect a lot of high-quality training images with photo-

graphic appearance instead of the enhanced results of manual adjustment. After that, a training set can be effectively constructed by sampling from these high-quality photos.

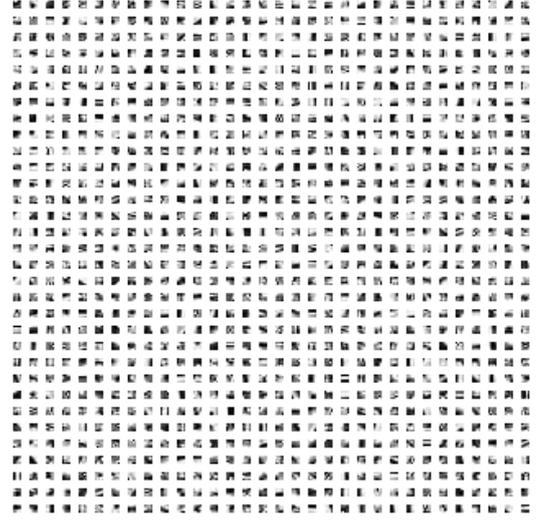


Fig. 2. Dictionary training. The dictionary with 1024 atoms can be learned with 100,000 detail patches ( $5 \times 5$ ) sampled from high-quality images with photographic appearance. Note: The detail dictionary is normalized for the purpose of better illustration.

To simulate the photographic enhancement effect, we first randomly extract a lot of small patches from the training set  $P = \{L^1, L^2, \dots, L^m\}$ , and then simply compute their local intensity differences to obtain a training set of detail patches  $Y = \{y_1, y_2, \dots, y_n | y_x = L_x^t - \text{mean}(L_x^t), t \in [1, m]\}$ . Here, the dictionary learning must fully guarantee the sparsity of coefficients recovered from  $\ell_1$ -norm minimization in (3). Thus we define a new  $\ell_1$ -norm minimization function to solve the detail dictionary and meanwhile enforce the sparsity of coefficients:

$$D = \arg \min_{D, \alpha} \|Y - D\alpha\|_2^2 + \lambda \|\alpha\|_1, \quad (4)$$

where each column of  $D$  should remove the scaling ambiguity by an  $\ell_2$ -norm constraint  $\|D_i\|_2^2 \leq 1, i \in [1, K]$ ; the  $\ell_1$ -norm constraint of  $\alpha$  is to enforce sparsity;  $\lambda$  is a balance parameter between them, which is fixed to 0.1. In (4), the optimization usually performs in an alternative manner over

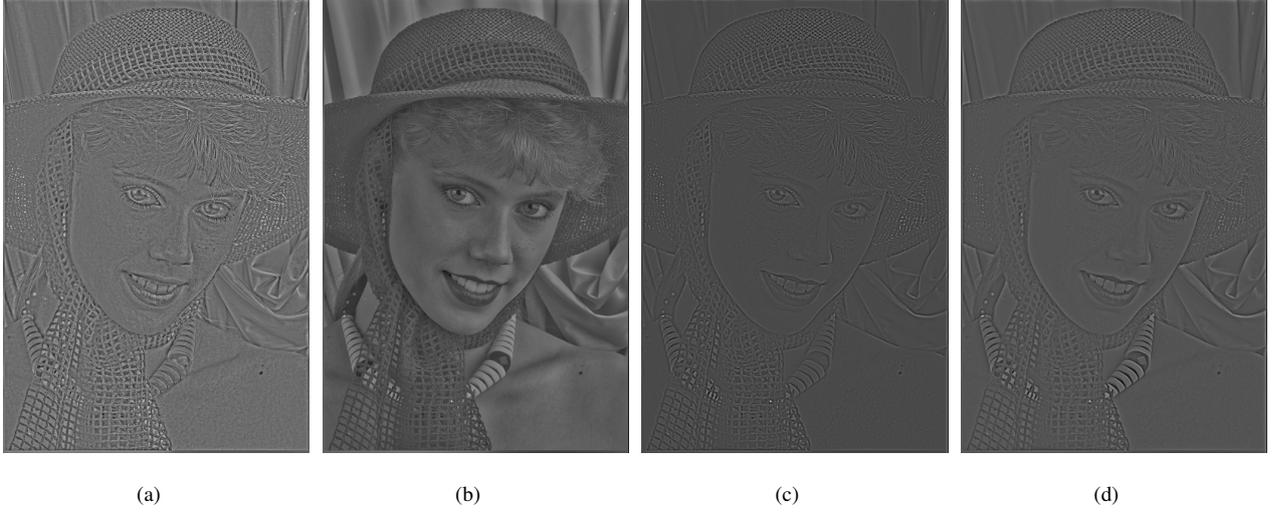


Fig. 3. Detail reconstruction and optimization. (a) The boosted detail layer of bilateral filter [1]. (b) The boosted detail layer of guided filter [2]. (c) Our reconstructed detail layer on (b). (d) The final optimized detail layer on (c). Note: All of detail layers are normalized for the purpose of better illustration.

two unknowns  $D$  and  $\alpha$ .

First of all, we employ a Gaussian random matrix to initialize a detail dictionary  $D$ , each column of which is normalized. Then we fix the detail dictionary  $D$  and update the sparse coefficients  $\alpha$  by the following minimization function:

$$\alpha = \arg \min_{\alpha} \|Y - D\alpha\|_2^2 + \lambda \|\alpha\|_1, \quad (5)$$

which can be solved efficiently through linear programming. Third, we turn to fix the sparse coefficients  $\alpha$  and update the detail dictionary  $D$  by the  $\ell_2$ -norm minimization function:

$$D = \arg \min_D \|Y - D\alpha\|_2^2, \quad (6)$$

which is regarded as a typical problem about quadratically constrained quadratic programming. Finally, we iterate between (5) and (6) until the optimization problem (4) converges. So far we have finished the entire training process and obtained a satisfactory detail dictionary. As shown in Fig.2, the dictionary with 1024 atoms is successfully trained using a set of 100,000 detail patches ( $5 \times 5$ ). We will employ the dictionary to reconstruct a high-quality appearance for defective enhanced image in Section 5.



Fig. 4. Appearance comparison with and without detail optimization. (a) The enhanced appearance without detail optimization. (b) The enhanced appearance with detail optimization. Detail optimization can effectively improve the local coherence of detail reconstruction and produce a higher-quality appearance.

## 5 Sparse Reconstruction

The sparse reconstruction aims to improve the appearance of enhanced result by dictionary learning. Besides the trained dictionary, (3) demonstrates that the initial boosted

detail layer is also necessary to reconstruction. Here, we employ a fast and non-approximate linear-time filter [2] to extract the detail layer, and then simply multiply a scale factor to boost it. Certainly, the other edge-aware filters can also be chosen to produce the detail layers if you want to optimize their enhanced effects. After that, all preparations for the sparse reconstruction are complete, including the detail dictionary and the initial boosted detail layer.

Formally, given a detail dictionary  $D$  and an initial boosted detail  $nd_x$  in a small patch  $x$ , we first optimize (3) to yield its sparse coefficients  $\alpha_x^*$ . Secondly, we collect all reconstructed detail patches  $d_x^* = D\alpha_x^*$  and integrate them into a complete detail layer  $d^*$ . Thirdly, we construct a gradient-guided optimization function to further improve the local coherence of enhancement:

$$\hat{d}^* = \arg \min_{\hat{d}^*} \|\hat{d}^* - d^*\|_2^2 + \lambda_1 \|\nabla \hat{d}^* - \nabla g\|_2^2, \quad (7)$$

where  $\hat{d}^*$  is the optimized detail layer;  $\nabla g$  is the guided gradient, which is often related to source image;  $\lambda_1$  is a regularization parameter, which is fixed to 0.05 in this paper. We hardly find an ideal guided gradient because local coherence is not guaranteed after traditional detail enhancement. Compared with  $nd$ , we observe that the source image has better local coherence. Thus we choose its gradient as suboptimal guided gradient and set a small value  $\lambda_1$  to balance local coherence and enhanced effect. Finally, the optimized detail layer  $\hat{d}^*$  is merged into a final enhanced image  $L^* = I + \hat{d}^*$ . The entire reconstruction process is summarized as Algorithm 1. Fig.3(a) and Fig.3(b) are the boosted detail layers of bilateral filter and guided filter in Fig.1. Their appearance issues can be solved by detail reconstruction and optimization in Fig.3(c) and Fig.3(d).

Although sparse reconstruction without (7) can also get the enhanced appearance back on track by suppressing noise, eliminating halos and correcting contrast, some incoherent enhanced effects may be existed between patches because the reconstruction for each patch is local and separate.

As shown in Fig.4, the enhanced results suffer from some visible artifacts that reduce the local coherence of enhancement. In contrast, the optimized detail by (7) can effectively improve the local coherence and produce a higher-quality appearance.

## 6 Evaluation Metric

To evaluate the performance of our new model, we define a EME value [26–28] (The Measure of Enhancement) as no-reference quantitative assessment by dividing image into nonoverlapping blocks, finding a measure based on minimum and maximum intensity values in each block, and averaging them. First, the enhanced result  $L^*$  is divided into  $k_1 k_2$  nonoverlapping subblocks  $L_{i,j}^*$  of size  $w_1 \times w_2$ . Then we employ these subblocks to compute the EME of  $L^*$  by the following function:

$$EME(L^*) = \frac{1}{k_1 k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} 20 \cdot \ln \frac{\max(L_{i,j}^*)}{\min(L_{i,j}^*)}, \quad (8)$$

where  $\max(L_{i,j}^*)$  and  $\min(L_{i,j}^*)$  are the maximum and minimum intensity levels in block  $L_{i,j}^*$ . (8) indicates that the higher EME is, the stronger the enhancement degree is. Finally, it is worth to note that EME is highly sensitive to noise and contrast. For example, we compute the EME values for four images in Fig.1: (a) 10.39 (b) 54.87 (c) 56.55 (d) 26.46, and observe that the EME value is increased by our enhancement method but the higher EME value does not mean the better appearance. In practice, both quantitative and subjective assessments should combine together to evaluate the enhancement quality.

Moreover, besides the above no-reference evaluation, we collect or produce high-quality and low-quality image pairs as source images and their corresponding ground truth images for reference evaluation. Here, the ground truth images can be obtained by two ways: (1) we take some photos with high-quality appearance as ground truth images by advanced cameras; (2) we manually adjust photo appear-

**Algorithm 1** Sparse Reconstruction**Input:** low-quality image  $I$  and trained detail dictionary  $D$ .**Output:** high-quality enhanced image  $L^*$ .

- 1: Extract the detail layer  $d$  from the low-quality image  $I$  by edge-aware filter.
- 2: Compute the boosted detail  $n \times d$ ,  $n$  is a default scale factor.
- 3: **for** each  $3 \times 3$  or  $5 \times 5$  patch  $x$  of the low-quality image  $I$ , taken starting from the upper-left corner with 1 pixel overlap in each direction, **do**
- 4: Obtain the boosted detail patch  $nd_x$  of the patch  $x$ .
- 5: Compute the sparse coefficients  $\alpha_x^*$  by minimizing the  $\ell_1$ -norm optimization function  $\|D\alpha_x - nd_x\|_2^2 + \lambda\|\alpha_x\|_1$  with the detail dictionary  $D$  and the boosted detail patch  $nd_x$ ,  $\lambda$  is a default balance parameter.
- 6: Generate the reconstructed detail patch  $d_x^* = D\alpha_x^*$ .
- 7: Integrate the new detail patch  $d_x^*$  into the complete detail  $d^*$ .
- 8: **end for**
- 9: Compute the average of  $d_i^*$  in each pixel  $i$  because the patches are overlapped.
- 10: Optimize the reconstructed detail  $d^*$  by the gradient-guided function  $\|\hat{d}^* - d^*\|_2^2 + \lambda_1\|\nabla\hat{d}^* - \nabla g\|_2^2$  with the guided gradient  $\nabla g$  and the regularization parameter  $\lambda_1$ .
- 11: **return** a final enhanced result  $L^* = I + \hat{d}^*$ .

ances to generate ground truth images by image editing tools. Then the appearance effects of low-quality images are improved by our new enhancement method. Finally, we measure the difference between the enhance images  $L^*$  and the corresponding high-quality ground truth images  $T$  by computing the following PSNR value (Peak Signal to Noise Ratio):

$$PSNR(L^*, T) = 20 \cdot \log_{10}\left(\frac{\text{MAX}}{\sqrt{\sum \|L_i^* - T_i\|^2/m}}\right), \quad (9)$$

where MAX is the maximum level of intensity;  $m$  is the number of pixels. (9) indicates that the larger PSNR is, the better the enhancement fidelity is. The increment of PSNR represents the improvement in appearance enhancement. Generally, the enhanced result with photographic appearance should have a large PSNR value, but the visible artifacts will severely reduce the PSNR value.

## 7 Experimental Results

The experimental environment involves a computer with a 3GHz CPU of Intel Core i5-2400 and 4GB memory, 64-bit Windows operating system, and Matlab version 7.11. In this paper, we train a overcomplete dictionary by sampling 100,000 patches and performing an iterative optimization, so it has to take about 30 minutes for dictionary training. On the other hand, sparse reconstruction for all small patches is also time-consuming, and it takes about 100 - 300s in the different resolutions ( $400 \times 300$ ,  $600 \times 480$ , and  $800 \times 600$ ). Here, all of traditional filters choose their optimal or default parameters for visual comparison.

Generally, we employ guided filter [2] to produce the initial boosted detail layers and then introduce them into our sparse reconstruction. However, the various edge-aware filters have their respective appearance issues, and is our model still effective to other edge-aware filters? Thus we first produce five enhanced results in Figs.5(d)~9(d) by



Fig. 5. Appearance enhancement with bilateral filter [1]. (a) Source image (PSNR = 20.86dB). (b) The enhanced result of bilateral filter ( $w = 5, \sigma = [16, 0.1], n = 5$ , PSNR = 19.51dB). (c) Our reconstructed result on the detail layer of (b) (PSNR = 30.87dB). (d) Ground truth image with high-quality appearance.



Fig. 6. Appearance enhancement with WLS filter [3]. (a) Source image (PSNR = 25.82dB). (b) The enhanced result of WLS filter ( $\lambda = 2, \alpha = 2, n = 5$ , PSNR = 16.10dB). (c) Our reconstructed result on the detail layer of (b) (PSNR = 32.69dB). (d) Ground truth image with high-quality appearance.



Fig. 7. Appearance enhancement with L0 filter [5]. (a) Source image (PSNR = 28.27dB). (b) The enhanced result of L0 filter ( $\lambda = 2e - 2, \kappa = 2, n = 5$ , PSNR = 14.00dB). (c) Our reconstructed result on the detail layer of (b) (PSNR = 35.36dB). (d) Ground truth image with high-quality appearance.

our new method and indicate them as ground truth images with high-quality appearance. Then we yield the boosted detail layers of other five state-of-the-art filters [1, 3–6] in Figs.5(b)~9(b), and reconstruct these detail layers to respectively obtain their enhanced results in Figs.5(c)~9(c). At last, we compute their PSNR values by (9) and observe their appearance quality to verify the effectiveness of our new method. In a word, our model can train a overcomplete de-

tail dictionary by sampling a lot of patches from high-quality photos and employ it to effectively constrain and guide the entire detail enhancement. Thus the learning-based reconstruction is very robust to different appearance issues and the quality of its production is also reliable and guaranteed.

To further verify the advantage of our new model, we adjust their individual enhancement parameters to obtain a same EME value, and then compare their enhanced appear-



Fig. 8. Appearance enhancement with laplacian filter [4]. (a) Source image (PSNR = 22.73dB). (b) The enhanced result of laplacian filter ( $\sigma_r = 0.4, \alpha = 0.01, \beta = 1$ , PSNR = 19.70dB). (c) Our reconstructed result on the detail layer of (b) (PSNR = 31.94dB). (d) Ground truth image with high-quality appearance.

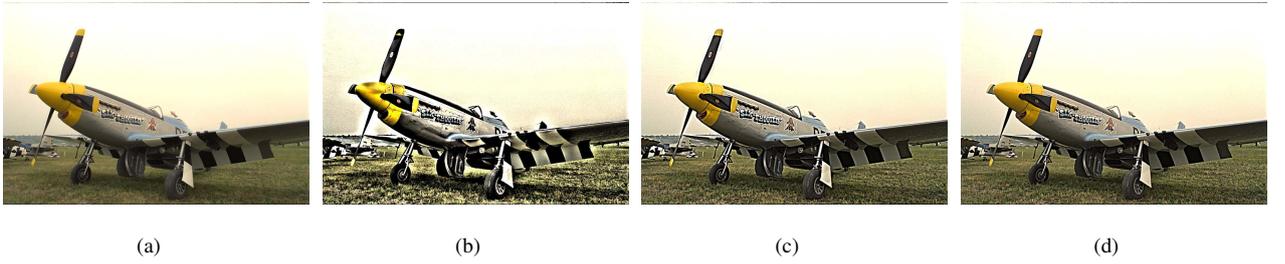


Fig. 9. Appearance enhancement with rolling filter [6]. (a) Source image (PSNR = 26.37dB). (b) The enhanced result of rolling filter ( $\sigma_s = 16, \sigma_r = 0.1, n = 5$ , PSNR = 17.39dB). (c) Our reconstructed result on the detail layer of (b) (PSNR = 31.27dB). (d) Ground truth image with high-quality appearance.

ances in Fig.10 and Fig.11. When the EME value is increased to a high level, the quality of traditional detail enhancement is hardly preserved because of severe artifacts, such as noise, color distortion, unnatural contrast, halos, and so on. Certainly, users can manually adjust the parameters of edge-aware filter to yield a better enhanced result, but the manual adjustment may consume too much time and energy and not still completely guarantee the enhanced quality. Fortunately, for a high-level detail enhancement, our new method can constrain the over-enhanced detail layer and improve its enhanced appearance by dictionary learning. Compared with the other six filters [1–6], a same EME value has different enhanced appearances and our two enhanced results without visible artifacts are obviously superior to their results.

We also want to quantitatively compare the influence of different dictionary sizes for appearance enhancement. Therefore, we first train two detail dictionaries

with 512 atoms and 1024 atoms respectively. Then we collect 200 source images with high-quality appearance (<http://pan.baidu.com/s/1qX9s5ha>) and degrade their detail layers to produce their corresponding low-quality images. Next we employ two different dictionaries to reconstruct these low-quality images and yield their enhanced results. Finally, we compute their PSNR values by (9) and compare the enhancement performance of two dictionaries. As shown in Fig.12, the PSNR mean value of dictionary with 1024 atoms is 0.1578dB bigger than that of dictionary with 512 atoms, but its reconstruction is also more time-consuming.



Fig. 10. Appearance comparison with same EME value. (a) Source image (EME = 12.06). (b) bilateral filter [1] ( $w = 1, \sigma = [16, 0.1], n = 8, \text{EME} = 43.79$ ). (c) WLS filter [3] ( $\lambda = 2, \alpha = 2, n = 2.1, \text{EME} = 43.41$ ). (d) laplacian filter [4] ( $\sigma_r = 0.4, \alpha = 0.01, \beta = 1, \text{EME} = 43.52$ ). (e) L0 filter [5] ( $\lambda = 2e - 2, \kappa = 2, n = 3.7, \text{EME} = 43.59$ ). (f) rolling filter [6] ( $\sigma_s = 16, \sigma_r = 0.1, n = 2, \text{EME} = 43.41$ ). (g) guided filter [2] ( $r = 16, \epsilon = 0.01, n = 4, \text{EME} = 43.58$ ). (h) Our method (EME = 43.82). Note: A boosted detail layer of guided filter ( $r = 16, \epsilon = 0.01, n = 15$ ) is introduced into our sparse reconstruction for same EME value. The partial enlarged details in (a)~(h) are shown in the third row.

**Table 1.** The correct choice number of our enhanced results in 10 selective images.

Tester	Num	Tester	Num	Tester	Num
User #1	1	User #6	0	User #11	1
User #2	1	User #7	2	User #12	2
User #3	2	User #8	1	User #13	1
User #4	1	User #9	1	User #14	0
User #5	0	User #10	0	User #15	1

Furthermore, we construct a testing set of 90 real high-quality photos and 10 enhanced images by our new model. Here, the boosted detail layers of guided filter [2]

( $r = 16, \epsilon = 0.01, n = 5$ ) are introduced into our sparse reconstruction. Then we organize fifteen testers to compare their appearances and find the 10 enhanced images from the testing set. As shown in Table 1, most of users cannot easily pick out our results according to noise, contrast, color, edge, and so on. The subjective evaluation demonstrates the effectiveness of our learning-based enhancement, which can produce high-quality photographic appearance.

## 8 Conclusion and Discussion

In this paper, we present a new learning-based method for photographic appearance enhancement. The method assumes that each small detail patch can be sparsely repre-

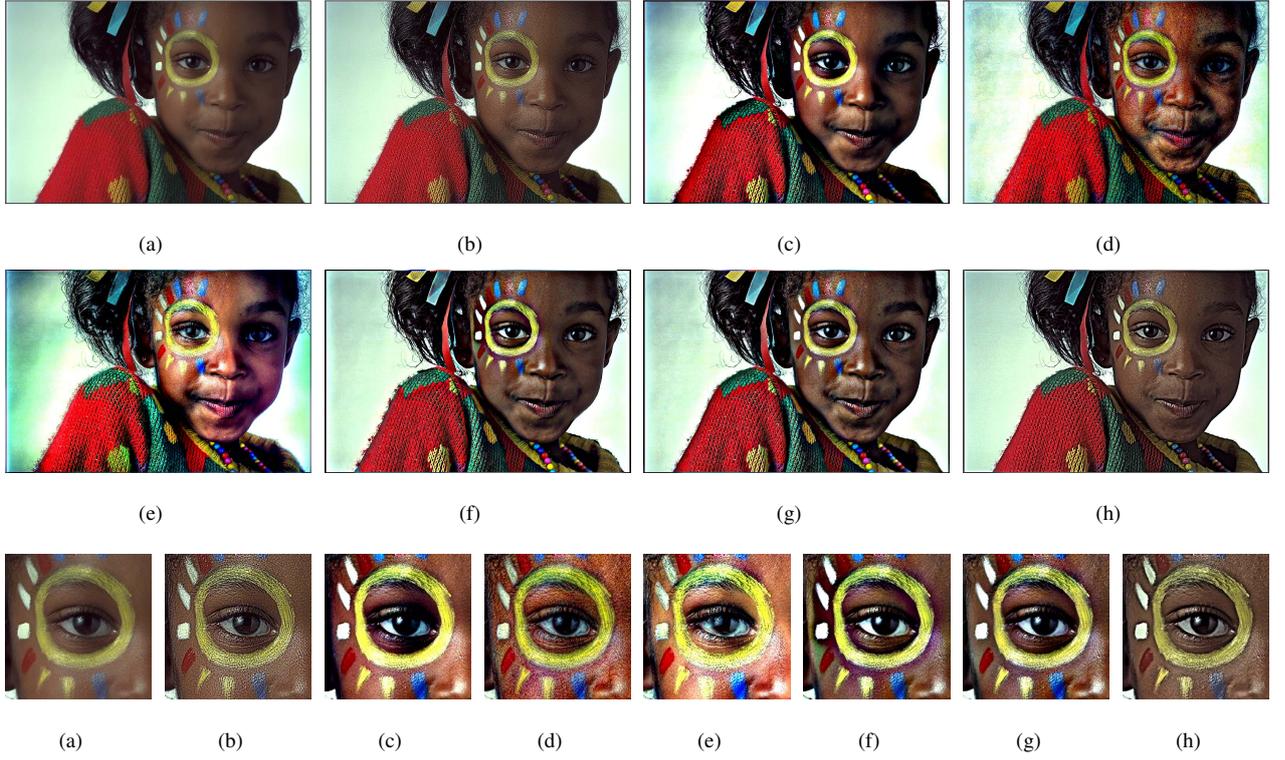


Fig. 11. Appearance comparison with same EME value. (a) Source image (EME = 13.03). (b) bilateral filter [1] ( $w = 1, \sigma = [16, 0.1], n = 6, \text{EME} = 45.03$ ). (c) WLS filter [3] ( $\lambda = 2, \alpha = 2, n = 3, \text{EME} = 45.62$ ). (d) laplacian filter [4] ( $\sigma_r = 0.4, \alpha = 0.01, \beta = 1, \text{EME} = 45.43$ ). (e) L0 filter [5] ( $\lambda = 2e - 2, \kappa = 2, n = 5, \text{EME} = 45.18$ ). (f) rolling filter [6] ( $\sigma_s = 16, \sigma_r = 0.1, n = 2.8, \text{EME} = 45.52$ ). (g) guided filter [2] ( $r = 16, \epsilon = 0.01, n = 5, \text{EME} = 45.98$ ). (h) Our method (EME = 45.74). Note: A boosted detail layer of guided filter ( $r = 16, \epsilon = 0.01, n = 13.5$ ) is introduced into our sparse reconstruction for same EME value. The partial enlarged details in (a)~(h) are shown in the third row.

sented in an overcomplete detail dictionary. On the one hand, we sample a lot of detail patches from high-quality photos and construct a training set to learn an overcomplete detail dictionary by an iterative  $\ell_1$ -norm minimization. On the other hand, we use the detail dictionary to reconstruct each boosted detail patch in source image and improve local coherence between patches by a gradient-guided optimization. Moreover, we propose a no-reference quantitative assessment metric and a reference evaluation metric to evaluate the performance of appearance enhancement. The final experimental results can verify our learning-based method's effectiveness for high-quality appearance enhancement.

Although our method can yield a high-quality appearance enhancement result, it must take more computing re-

sources in time and memory for training an overcomplete detail dictionary and reconstructing each small detail patch. Therefore, compared with some edge-aware filters, a main limitation is that our enhancement method is inefficient in dictionary training and sparse reconstruction, which causes that it is not well adapted to some fast enhancement cases. We plan to refine the entire process of training and reconstruction and introduce some acceleration techniques for the proposed pipeline. Another limitation is that our model cannot control the amount of detail enhancement, and thus the initial scale factor might still have effects on final enhancement result. In the future, we hope to exclude the initial input of scaled-based enhancement and refine the entire sparse optimization to directly produce the final boosted re-

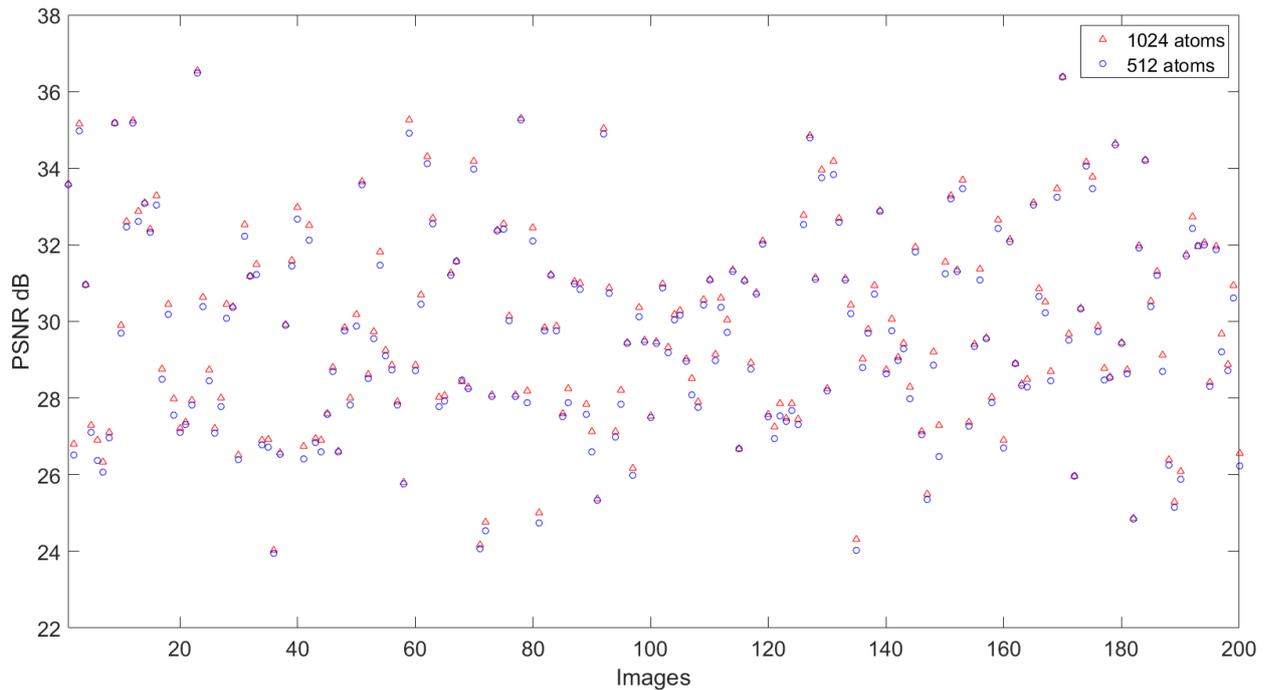


Fig. 12. PSNR comparison of 200 photos using dictionaries of different sizes (512 atoms - blue circle and 1024 atoms - red triangle). The PSNR value in vertical axis is from 22dB to 38dB; the photo number in horizontal axis is from 1 to 200. The PSNR mean value of dictionary with 1024 atoms is 0.1578dB bigger than that of dictionary with 512 atoms.

sult. Moreover, based on space-time consistency, we would also extend our appearance enhancement model to handle the video streams.

## References

- [1] Paris S, Durand F. A fast approximation of the bilateral filter using a signal processing approach. In *Proc. ECCV*, May 2006, pp.568–580.
- [2] He K M, Sun J, Tang X O. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(6): 1397–1409.
- [3] Farbmán Z, Fattal R, Lischinski D, Szeliski R. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics*, 2008, 27(3): 67(1–10).
- [4] Paris S, Hasinoff S, Kautz J. Local laplacian filters: edge-aware image processing with a laplacian pyramid. *ACM Transactions on Graphics*, 2011, 30(4): 68(1–12).
- [5] Xu L, Lu C W, Xu Y, Jia J Y. Image smoothing via L0 gradient minimization. *ACM Transactions on Graphics*, 2011, 30(6): 174(1–12).
- [6] Zhang Q, Shen X Y, Xu L, Jia J Y. Rolling guidance filter. In *Proc. ECCV*, June 2014, pp.815–830.
- [7] Wright J, Ma Y, Mairal J, Sapiro G, Huang T S, Yan S. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 2010, 98(6): 1031–1044.
- [8] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse

- representation. *IEEE Transactions on Signal Processing*, 2006, 54(11): 4311–4322.
- [9] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12: 629–639.
- [10] Tomasi C, Manduchi R. Bilateral filtering for gray and color images. In *Proc. ICCV*, Jan. 1998, pp.839–846.
- [11] Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 2002, 21(3): 257–266.
- [12] Bae S, Paris S, Durand F. Two-scale tone management for photographic look. *ACM Transactions on Graphics*, 2006, 25(3): 637–645.
- [13] Weiss B. Fast median and bilateral filtering. *ACM Transactions on Graphics*, 2006, 25(3): 519–526.
- [14] Chen J W, Paris S, Durand F. Real-time edge-aware image processing with the bilateral grid. *ACM Transactions on Graphics*, 2007, 26(3): 103(1–9).
- [15] Porikli F. Constant time  $O(1)$  bilateral filtering. In *Proc. CVPR*, June 2008, pp.1–8.
- [16] Yang Q X, Tan K H, Ahuja N. Real-time  $O(1)$  bilateral filtering. In *Proc. CVPR*, June 2009, pp.557–564.
- [17] Fattal R. Edge-avoiding wavelets and their applications. *ACM Transactions on Graphics*, 2009, 28(3): 22(1–10).
- [18] Subr K, Soler C, Durand F. Edge-preserving multiscale image decomposition based on local extrema. *ACM Transactions on Graphics*, 2009, 28(5): 147(1–9).
- [19] Gastal E, Oliveira M. Domain transform for edge-aware image and video processing. *ACM Transactions on Graphics*, 2011, 30(4): 69(1–12).
- [20] Li X Y, Gu Y, Hu S M, Martin R. Mixed-domain edge-aware image manipulation. *IEEE Transactions on Image Processing*, 2013, 22(5): 1915–1925.
- [21] Wang H, Cao J J, Liu X P, Wang J M, Fan T R, Hu J P. Least-squares images for edge-preserving smoothing. *Computational Visual Media*, 2015, 1(1): 27–35.
- [22] Shao P, Ding S H, Ma L Z, Wu Y S, Wu Y J. Edge-preserving image decomposition via joint weighted least squares. *Computational Visual Media*, 2015, 1(1): 37–47.
- [23] Xu L, Ren J, Yan Q, Liao R J, Jia J Y. Deep edge-aware filters. In *Proc. ICML*, July 2015, pp.1669–1678.
- [24] Gonzalez R, Woods R. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.
- [25] Stark J A. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on Image Processing*, 2000, 9(5): 889–896.
- [26] Arici T, Dikbas S, Altunbasak Y. A histogram modification framework and its application for image contrast enhancement. *IEEE Transactions on Image Processing*, 2009, 18(9): 1921–1935.
- [27] Celik T, Tjahjadi T. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 2011, 20(12): 3431–3441.
- [28] Celik T. Two-dimensional histogram equalization and contrast enhancement. *Pattern Recognition*, 2012, 45(10): 3810–3824.
- [29] Fattal R, Agrawala M, Rusinkiewicz S. Multiscale shape and detail enhancement from multi-light image collections. *ACM Transactions on Graphics*, 2007, 26(3): 51(1–10).

- [30] Joshi N, Matusik W, Adelson E. and Kriegman, David J. Personal photo enhancement using example images. *ACM Transactions on Graphics*, 2010, 29(2): 12(1–15).
- [31] Kang S B, Kapoor A, Lischinski D. Personalization of image enhancement. In *Proc. CVPR*, June 2010, pp.1799–1806.
- [32] Caicedo J C, Kapoor A, Kang S B. Collaborative personalization of image enhancement. In *Proc. CVPR*, June 2011, pp.249–256.
- [33] Bychkovsky V, Paris S, Chan E, Durand F. Learning photographic global tonal adjustment with a database of input/output image pairs. In *Proc. CVPR*, June 2011, pp.97–104.
- [34] Kaufman L, Lischinski D, Werman M. Content-aware automatic photo enhancement. *Computer Graphics Forum*, 2012, 31(8): 2528–2540.
- [35] Lischinski D, Farbman Z, Uyttendaele M, Szeliski R. Interactive local adjustment of tonal values. *ACM Transactions on Graphics*, 2006, 25(3): 646–653.
- [36] Zhang F, Zhang X, Qin X Y, Zhang C M. Enlarging image by constrained least square approach with shape preserving. *Journal of Computer Science and Technology*, 2015, 30(3): 489–498.
- [37] Liu Q, Chen M Y, Zhou D H. Single image haze removal via depth-based contrast stretching transform. *Science China: Information Sciences*, 2015, 58(1): 1–17.
- [38] Zhang Q, Nie Y W, Zhang L, Xiao C X. Underexposed video enhancement via perception-driven progressive fusion. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(6): 1773–1785.
- [39] Cheng M M, Hou Q B, Zhang S H, Rosin P. Intelligent visual media processing: when graphics meets vision. *Journal of Computer Science and Technology*, 2017 Vol. 32(1): 110–121.
- [40] DeCarlo D, Santella A. Stylization and abstraction of photographs. *ACM Transactions on Graphics*, 2002, 21(3): 769–776.
- [41] Winnemoller H, Olsen S, Gooch B. Real-time video abstraction. *ACM Transactions on Graphics*, 2006, 25(3): 1221–1226.
- [42] Echevarria J, Wilensky G, Krishnaswamy A, Kim B, Gutierrez D. Computational simulation of alternative photographic processes. *Computer Graphics Forum*, 2013, 32(4): 7–16.
- [43] Son M J, Lee Y J, Kang H, Lee S Y. Art-photographic detail enhancement. *Computer Graphics Forum*, 2014, 33(2): 391–400.
- [44] Aydin T O, Smolic A, Gross M. Automated aesthetic analysis of photographic images. *IEEE Transactions on Visualization and Computer Graphics*, 2015, 21(1): 31–42.
- [45] Mairal J, Sapiro G, Elad M. Learning multiscale sparse representations for image and video restoration. *Multiscale Modeling & Simulation*, 2008, 7(1): 214–241.
- [46] Dong W, Zhang L, Shi G, Li X. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 2013, 22(4): 1620–1630.
- [47] Yang J, Wright J, Huang T S, Ma Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 2010, 19(11): 2861–2873.
- [48] Zhang Y, Liu J, Yang W, Guo Z. Image super-resolution based on structure-modulated sparse rep-

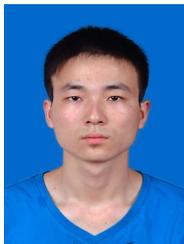
resentation. *IEEE Transactions on Image Processing*, 2015, 24(9): 2797–2810.

- [49] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 2006, 15(12): 3736–3745.
- [50] Zeng X, Bian W, Liu W, Shen J, Tao D. Dictionary pair learning on grassmann manifolds for image denoising. *IEEE Transactions on Image Processing*, 2015, 24(11): 4556–4569.



**Zhi-Feng Xie** received his Ph.D. degree in Computer Application Technology from Shanghai Jiao Tong University, China, in 2013. He was a Research Assistant at Department of Computer Science, City University of Hong Kong, Hong Kong, in 2011. He is now an Assistant Professor with Shanghai University, Shanghai, China. His research

interests include image/video editing, computer graphics, and digital media technology.



**Shi Tang** received his B.Sc. degree in Digital Media Technology from Shanghai University, China, in 2015. Now he is a M.S. candidate in Department of Film and Television Engineering, Shanghai University, Shanghai, China. His research interests include image processing, computer graphics, and digital media technology.



**Dong-Jin Huang** received his Ph.D. degree in Computer Applica-

tion Technology from Shanghai University, China, in 2011. He was a Post-doctor with University of Bradford, UK, from 2012 to 2013. He is now an Assistant Professor with Shanghai University, Shanghai, China. His research interests are augmented reality, computer vision and computer graphics.



**You-Dong Ding** received his Ph.D. degree in Mathematics from University of Science and Technology of China, in 1997. He was a Post-doctor at the Department of Mathematics of Fudan University from 1997 to 1999. He is now a Professor with Shanghai University, Shanghai, China. His research interests are computer graphics, image processing, and digital media technology.



**Li-Zhuang Ma** received his Ph.D. degree in Mathematics from Zhejiang University, China, in 1991. He was a Post-doctor at the Department of Computer Science of Zhejiang University from 1991 to 1993. He is now a Full Professor, Ph.D. Tutor, and the Head of Digital Media Technology and Data Reconstruction Lab at the Department of Computer

Science and Engineering, Shanghai Jiao Tong University, China since 2002. He is also the Chairman of the Center of Information Science and Technology for Traditional Chinese Medicine at Shanghai Traditional Chinese Medicine University. His research interests include computer-aided geometric design, computer graphics, and image processing.