LEARNING IN-PLACE RESIDUAL HOMOGENEITY FOR IMAGE DETAIL ENHANCEMENT

He Jiang[1]*, HuangKai Cai*[1]*, Jie Yang*[1]*

[1]Institute of Pattern Recognition and Image Processing, Shanghai Jiaotong University, China

ABSTRACT

In this paper, we put forward and demonstrate a novel method in image and video detail enhancement-- in-place residual homogeneity (IP). In-place residual homogeneity is a regular law we find in testing different blocks in database, that is, residual blocks with slight different resolutions hold homogenous structures. By learning this homogeneity, we guess that it might be a good description of image's detail layer. Then images are enhanced by designed framework and accelerated by proposed fast in-place search method. Unlike most algorithms that need to adjust parameters by manual to get best performance, our approach is adaptive. Besides, many algorithms will change images' intensity, but our IP can keep natural images from over enhancement. Moreover, IP is also robust to low bit rate H.265 encoder and decoder system and runs faster than most popular methods. The last but not least, it can be easily FPGA implemented as well. Numbers of experiments testify that our algorithm is robust with good performance both subjectively and objectively.

Index Terms— In-place residual homogeneity, detail enhancement, fast in-place search, FPGA implemented.

1. INTRODUCTION

With development of information technology, billions of digital images are created every day. But many image details are degraded by noise or resolution limitations, so detail enhancement algorithm is highly required. Many algorithms are proposed to improve this problem. Bilateral filter [1] is a classical filter, though it's effective in many situations, it may have unwanted gradient reversal artifacts [2], [3]. He *et al* proposed a guided filter [2] and [3], it has better behaviors near edges and enables applications like de-hazing and matting. Li *et al* [4] proposed an l_0-based filter that can improve halo artifacts. Fattal *et al* [5] used global optimization based filters, getting multi-scale detail exaction under weighted least squares optimization, which is called WLS filter. It reduces de-blurring artifacts of [1], [2] and [3], providing an excellent foundation. Kou *et al* [7] proposed a new l_0-based algorithm, preserving sharp edge better. Xu *et al* [8] recovered details by analyzing scale in-variance of fractal dimension, extending to true texture enhancement. In this paper, we present a novel detail enhancement method via learning in-place residual homogeneity, that is, residual images with slight different resolutions are structure homogenous. There exist two procedures in our algorithm, fast in-place search and block match.

Corresponding author*: jieyang@sjtu.edu.cn (Jie Yang). This research is partly supported by NSFC, China (No: 61572315) and Committee of ST, Shanghai, China (No: 17JC1403000).

2. IN-PLACE RESIDUAL HOMOGENEITY OF IMAGE

2.1 Definition of in-place residual homogeneity

Given the original image $\mathbf{L_0}$, we can get $\mathbf{L_1}=\beta\times\mathbf{L_0}$ and $\mathbf{L_2}=(1/\beta)\times\beta\times\mathbf{L_0}$ using bi-linear [1], where β is set as 1.25. Bi-linear in [1] is a loss interpolation method, so $\mathbf{L_0}$ and $\mathbf{L_2}$ are not exactly same, we define $\mathbf{H_0}=\mathbf{L_0}-\mathbf{L_2}$ as residual part of $\mathbf{L_0}$. As we can see in Fig. 3, $\mathbf{L_0}$ and $\mathbf{L_1}$ are with homogenous structure. So block $[x: x+m, y: y+n]$ in $\mathbf{L_0}$ have a strong probability to be homogenous with one of blocks $[\beta x +p: \beta x+p+m, \beta y+t: \beta y+t+n]$ in $\mathbf{L_1}$, where p and t are offset numbers ranging from -2 to 2. If $\mathbf{H_1}$ is generated the same way as $\mathbf{H_0}$, that is, $\mathbf{H_1}=\mathbf{L_1}-(1/\beta)\times\beta\times\mathbf{L_1}$, the structure between $\mathbf{H_1}$ and $\mathbf{H_0}$ should be homogenous, so $\mathbf{H_1}$ and $\mathbf{H_0}$ must be satisfied with $\mathbf{H_1}[\beta x+t: \beta x+t+m, \beta y+p: \beta y+p+n] \approx \mathbf{H_0}[x: x+m, y: y+n]$, which we call this in-place residual homogeneity, revealing facts that most homogenous residual blocks in $\mathbf{H_0}$ exist just in restricted areas of $\mathbf{H_1}$.

Definition:
For image $\mathbf{X}$, block $\mathbf{X_1}=\mathbf{X}[x: x+m, y: y+n]$ in $\mathbf{X}$, block $\mathbf{X_2}=\boldsymbol{\beta}\mathbf{X}(t, p)=\boldsymbol{\beta}\mathbf{X}[\beta x+t: \beta x+t+m, \beta y+p: \beta y+p+n]$ in $\boldsymbol{\beta}\mathbf{X}$, if $\min\{\mathrm{SAD}(\mathbf{X_1}, \mathbf{X_2})\}<\mathit{threshold}$, $\mathbf{X_1}$ is in-place homogenous with $\mathbf{X_2}$, that is $\mathbf{X_1}\sim\mathbf{X_2}$, where *threshold* is set as $4mn$.

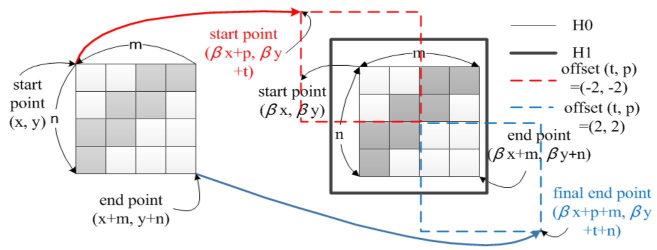


Fig. 1 Details of in-place residual homogeneity

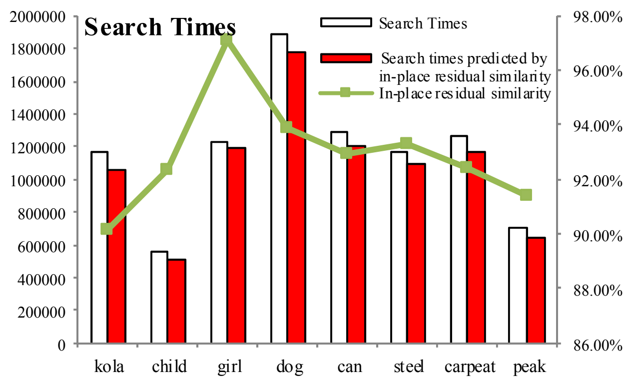


Fig. 2. Image in-place residual similarity statistics

To evaluate such IP, we divide $\mathbf{H_0}$ into 4×4 sub blocks, then search $\mathbf{H_0}$ to find their match blocks in database and count search times. Besides, we also apply IP to find $\mathbf{H_0}$'s match blocks in $\mathbf{H_1}$. For example, for all sub blocks in image kola, we search 1.2×10^6 times in database, but 10^6 search times among them are unnecessary. Match blocks just locate in their corresponding in-place homogenous areas in $\mathbf{H_1}$. Fig. 2 shows that average in-place residual similarity in test is 93%.

2.2 In-place homogeneity between $\mathbf{L_1}$ and $\mathbf{L_2}$

Suppose $\boldsymbol{P}\mathbf{L_0}=\mathbf{L_1}$, $\boldsymbol{Q}\mathbf{L_1}=\mathbf{L_2}$, $\mathbf{H_0}=\mathbf{L_0}-\mathbf{L_2}$, $\mathbf{L_1}-\boldsymbol{P'Q'}\mathbf{L_1}=\mathbf{H_1}$, where $\boldsymbol{P}$ and $\boldsymbol{Q}$, $\boldsymbol{P'}$ and $\boldsymbol{Q'}$ are up-scale and down-scale matrix with different rows and columns. According to our experiment, $\mathbf{H_0}$ and $\mathbf{H_1}$ are in-place residual homogenous, that is to say, $\mathbf{H_0}\,[x{:}\ x+m,\ y{:}\ y+n] \sim \beta\mathbf{H_0} = \mathbf{H_1}\,[\beta x+t{:}\ \beta x+t+m,\ \beta y+p{:}\ \beta y+p+n]$. With these conditions, conclusion can be drawn that in-place homogeneity exists between $\mathbf{L_1}$ and $\mathbf{L_2}$. The proof is as follows:

$\mathbf{H_0} \sim \mathbf{H_1} \Rightarrow \mathbf{H_0}\,[x{:}\ x+m,\ y{:}\ y+n] = \mathbf{H_1}\,[\beta x+t{:}\ \beta x+t+m,\ \beta y+p{:}\ \beta y+p+n]$, $\beta\mathbf{H_0} \subset \mathbf{H_1} \Rightarrow \beta\mathbf{H_0} \sim \mathbf{H_1} \Rightarrow \mathbf{H_0} \sim \beta\mathbf{H_0} \sim \mathbf{H_1} \Rightarrow (\boldsymbol{I}-\boldsymbol{PQ})^{-1}\,\mathbf{H_0} \sim (\boldsymbol{I}-\boldsymbol{PQ})^{-1}\,\beta\mathbf{H_0} \sim (\boldsymbol{I}-\boldsymbol{P'Q'})^{-1}\,\mathbf{H_1}$.

$\mathbf{L_2} = \mathbf{L_0}-\mathbf{H_0} = ((\boldsymbol{I}-\boldsymbol{PQ})^{-1}-\boldsymbol{I})\,\mathbf{H_0} \approx (\boldsymbol{I}+\boldsymbol{PQ}+\boldsymbol{PQ}^2-\boldsymbol{I})\,\mathbf{H_0} = (\boldsymbol{PQ}+\boldsymbol{PQ}^2)\,\mathbf{H_0} \sim (\boldsymbol{I}+\boldsymbol{PQ}+\boldsymbol{PQ}^2)\,\mathbf{H_0} \approx (\boldsymbol{I}-\boldsymbol{PQ})^{-1}\,\mathbf{H_0} \sim (\boldsymbol{I}-\boldsymbol{P^*Q^*})^{-1}\,\mathbf{H_1} = \mathbf{L_1} \Rightarrow \mathbf{L_2} \sim \mathbf{L_1} \Rightarrow \mathbf{L_2}\,[x{:}\ x+m,\ y{:}\ y+n] = \mathbf{L_1}\,[\beta x+t : \beta x+t+m,\ \beta y+p{:}\ \beta y+p+n]$, where in demonstration above, $\boldsymbol{I}$ is identity matrix, and eigenvalues of $\boldsymbol{PQ}$ are small, $(\boldsymbol{I}-\boldsymbol{PQ})^{-1} \approx \boldsymbol{I}+\boldsymbol{PQ}+\boldsymbol{PQ}^2$.

3. FAST IN-PLACE SEARCH

In order to avoid unnecessary search, accurate search is full of great importance. In this paper, we put forward a fast in-place search method between $\mathbf{L_1}$ and $\mathbf{L_2}$. $\mathbf{L_0}$ is up-scaled to get $\mathbf{L_1}$ then down-scaled to acquire $\mathbf{L_2}$, due to demonstration above, $\mathbf{L_2}$ is seen as an in-place homogenous part of $\mathbf{L_1}$. So $\mathbf{L_1}\,[\beta x+t : \beta x+t+m,\ \beta y+p{:}\ \beta y+p+n] \approx \mathbf{L_2}\,[x{:}\ x+m,\ y{:}\ y+n]$. As Fig. 3 shows, only four nearest neighbor pixels of 5×5 match block in $\mathbf{L_1}$ are used For instance, if a pixel locates at coordinate [2, 2] in $\mathbf{L_2}$, β=1.25, 2×1.25=2.5, the most in-place homogenous pixel has a very large probability locating between [2, 2] and [3, 3].

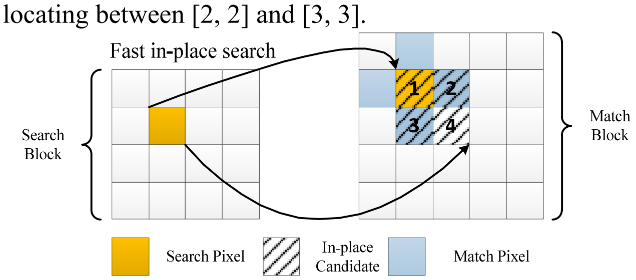


Fig. 3. Fast in-place search

To get accurate position of best pixel in four in-place candidates, function (1) must be solved.

$$\mathbf{X} = \arg\min \sum_{a,b,c,d} \sum_{\Delta x=0}^{1} \sum_{\Delta y=0}^{1} L_2(a-1:a+1, b-1:b+1) - L_1(c+\Delta x-1:c+\Delta x+1, d+\Delta y-1:d+\Delta y+1) \qquad (1)$$

,where $\mathbf{X}$ is the best location of four candidates [Δx, Δy]. Let [a, b] and [c, d] be start search point in $\mathbf{L_2}$ and $\mathbf{L_1}$. Block in $\mathbf{H_0}$ can be transferred from [a-1: a+1, b-1: b+1] to [$c+\Delta x$-1: $c+\Delta x$+1, $d+\Delta y$-1: $d+\Delta y$+1], generating a new $\mathbf{H_1}$, thus in-place residual homogeneity between $\mathbf{H_0}$ and $\mathbf{H_1}$ is preserved in this way. Popular methods based on operators will lead to image blurring, but the detail of $\mathbf{H_1}$ based on IP is well protected. To evaluate performance of our fast in-place search, we compare it to full search and ways in [6]. Table 1 shows average running time and reduced PSNR value of the first 300 frames in selected 720p videos. It can be seen that our method is much faster with tiny quality loss.

Table 1: Video running time and reduced PSNR

Selected videos	[6]'s time	Our time	Reduced time	Reduced PSNR
Foreman	9.54s	**0.64s**	79.28%	-0.074dB
Football	9.77s	**0.68s**	76.97%	-0.065dB
Mobile	9.65s	**0.64s**	79.35%	-0.078dB
Tennis	10.21s	**0.59s**	80.31%	-0.083dB

4 RESULTS AND DISCUSSION

4.1 System schematic diagram

The system schematic diagram is shown in Fig. 4. First, in-place residual homogeneity between $\mathbf{H_0}$ and $\mathbf{H_1}$ is found in our research. Second, fact is testified that in-place homogeneity also exists in $\mathbf{L_1}$ and $\mathbf{L_2}$. Third, we protect homogenous residual blocks in $\mathbf{H_0}$ and $\mathbf{H_1}$, and design this framework, and image $\mathbf{L_{enhance}} = \mathbf{L_0} + \alpha$ **Detail** is stand for final enhanced result, where $\alpha = 2$ in this paper.

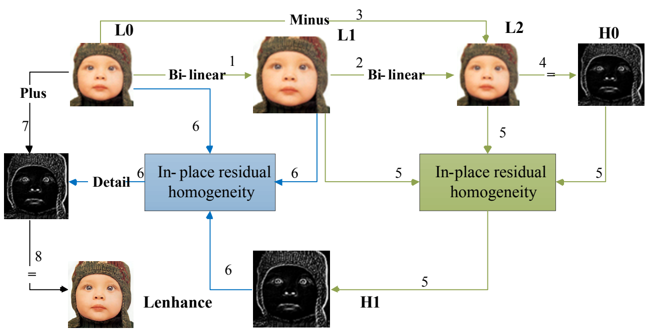


Fig. 4. System of in-place residual homogeneity

4.2 Subjective experimental processing result

To demonstrate performances of our IP, methods in the experiment are implemented in MATLAB 2013a and they are run on an Intel Core(TM) i7 CPU 3.4GHz machine. Fig. 5 illustrates five examples based on [2], [3], [5] and our IP. Parameter r and eps in [2], [3] are set as 10 and 0.01, and parameters in [5] are set as follows. (val0=2, val1=1, val2=1, exposure=1, saturation= 1.1, gamma=1). Comparing to mainstream algorithms including guide image filter [2], [3], weighted least square filter [5], l_0-based method [6] and [7], the proposed approach has three advantages. First, image signal parameters in [3], [4] and [5] are set globally, thus they are not suitable for different image contents. Parameters must be changed to get best performance by manual. However, our algorithm utilizes in-place homogeneity in image itself and it is robust to most image styles. Second, our algorithm keeps images from being over enhanced, which are serious in results of [2], [3], [5]. They always change image intensity. However, intensity in our IP is not changed and results are more visible. Third, our method is low complexity yet powerful. It can also be used in FPGA applications. [4], [5], [6] and [7] all contain complex operations or optimization regulation terms, so all of them cannot be easily hardware implemented. But our method is suitable for practical applications. Fig. 6 show two group images and their corresponding objective intensity curves. Each group has four images including original image, result of GF [2], [3], WLS [5] and our IP. They are marked with four different color lines respectively, whose width is one pixel, and values of these lines are regarded as y-axis of coordinate. From four intensity curves, we can see that curves of [2], [3] and [5] sometimes deviate from original image signals, where happened in over enhanced regions, but our curves follows closely to curve, It's proved our IP is a texture true enhancement method.

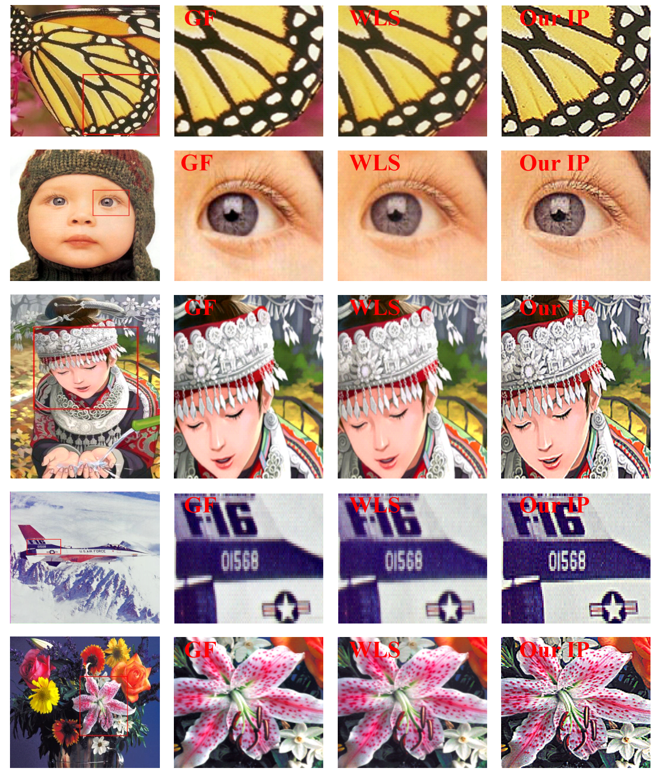
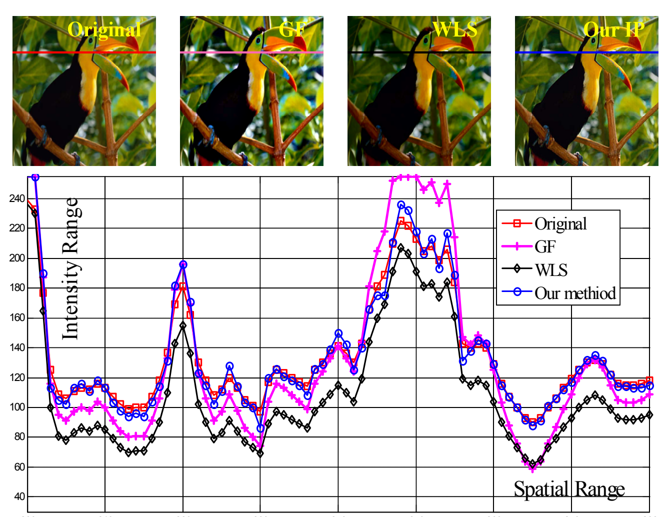


Fig. 5. Experimental results of GF [2], [3], WLS [5], and Our IP

4.3 Objective intensity curve



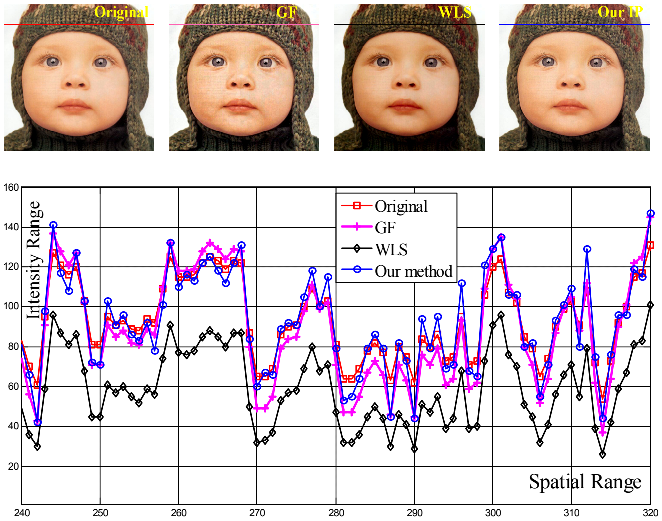


Fig. 6. Objective intensity curves, The red, pink, black, blue line refer to original image, GF[2], [3], WLS[5], and our IP(in-place method) respectively

4.4 Robust to low bit rate H.265 encoder and decoder

H.265 video standard is widely used in video coding and transmission. Low bit rate H.265 encoder might ignore image detail to meet the demand of transmission bandwidth, that is to say, if an enhanced video flows in a low bit rate H.265 encoder and flow out from a H.265 decoder, whose detail will suffer a great loss. But our IP is adaptive to that. PSNR loss of our method is almost negligible. Moreover, image detail enhancement is a kind of traditional image processing algorithm. If it can be well embedded in H.265 system, it must be full of great value. We cut out first 300 frames in video *Life of PI.yuv* (720p) to test IP algorithm. The left and right side of black line are input video and enhanced video after 2M/s H.265 systems respectively. Furthermore, average PSNR loss of these frames is **0.087dB**, it can be almost negligible.

4.5 Another faster version

To cater to fast application, we simplified our model to a faster version, which we call it IP2, as Fig. 7 shows. Unlike proposed IP algorithm, the module based on in-place homogeneity is used only once, resulting in a nearly twice faster algorithm. Having test the intensity curve of IP2 system, it does not perform well enough as GF [2], [3] and WLS [5]. However, for images or videos with great demand in fast speed, IP2 algorithm might be a good choice. Table 2 shows running time of each algorithm. Method WLS sometimes fails in high resolution formats like 2K due to its high computation, our IP and IP2 is robust at popular existing video formats. The last but not least, both IP and IP2 are faster than algorithm [2], [3], and [5].

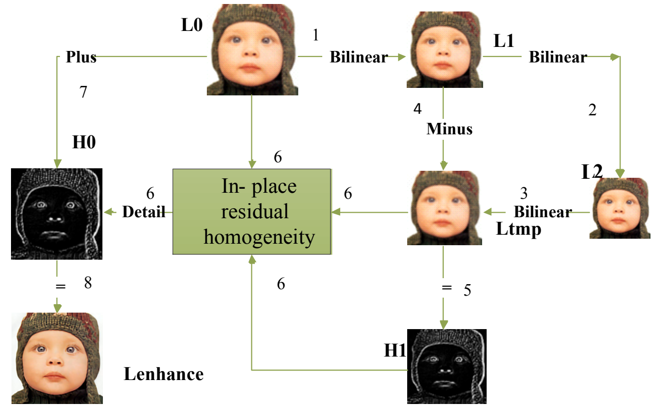


Fig. 7 Faster version algorithm, IP2

Table 2: Different algorithms' running time

Images	Format	GF/s	WLS/s	Our IP/s	Our IP2/s
Mickey	CIF	0.12	1.04	**0.09**	**0.05**
Child	CIF	0.14	1.07	**0.10**	**0.06**
Lena	512×512	0.56	3.02	**0.32**	**0.18**
Pepper	512×512	0.58	3.12	**0.35**	**0.19**
Plane	720p	2.73	23.12	**1.76**	**0.86**
Tulips	720p	2.24	17.43	**1.56**	**0.78**
Kangaroo	1080p	4.08	35.56	**2.68**	**1.45**
Chips	1080p	4.13	36.33	**2.84**	**1.57**
Harbor	2K	8.55	×	**5.91**	**2.97**
Building	2K	8.74	×	**6.05**	**3.44**

5 CONCLUSIONS

In this paper, we propose and demonstrate a novel method in image detail enhancement. It is based on in-place residual homogeneity (IP). We get an idea in our research statistics. Images are enhanced and accelerated by IP framework and fast in-place search. Fig. 5 and Fig. 6 indicate final output images are more natural looking with minimum average distance in intensity curve comparing to original images' edge regions, meaning that IP is a texture true enhancement approach with best edge preserving ability. Moreover, IP is robust to low bit rate H.265 encoder and decoder, with tiny PSNR loss in sample tests. Furthermore, we offer a faster version of IP called IP2. Table 2 tells us that both IP and IP2 are faster than traditional methods. The last but not least, IP itself is easily hardware implemented with practical value, resulting in both subjective and objective improvement.

REFERENCES

[1] Tomasi C, Manduchi R. Bilateral filtering for gray and color images", ICCV [J].1998:839 - 846.

[2] He K, Sun J, Tang X. Guided Image Filtering [M] Computer Vision, ECCV 2010Springer Berlin Heidelberg, 2010:1-14.
[3] He K, Sun J, Tang X. Guided Image Filtering [J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2013, 35(6):1397-1409.
[4] Zhengguo Li, Jinghong Zheng, Zijian Zhu. Content adaptive guided image filtering[C] IEEE International Conference on Multimedia and Expo. 2014:1-6.
[5] Farbman Z, Fattal R, Lischinski D, et al. Edge-preserving decompositions for multi-scale tone and detail manipulation.[J]. ACM Transactions on Graphics, 2008, 27(3):15-19.
[6] Freedman G, Fattal R. Image and Video Upscaling From Local Self-Examples [J]. ACM Transactions on Graphics, 2011, 30(2): 474-484.
[7] Kou F, Chen W, Li Z, et al. Content Adaptive Image Detail Enhancement [J]. IEEE Signal Processing Letters, 2015, 22(2):211-215.
[8] Xu H, Zhai G, Yang X. Single Image Super-resolution With Detail Enhancement Based on Local Fractal Analysis of Gradient [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2013, 23(10):1740-1754.