

# ALRe: Outlier Detection for Guided Refinement

Mingzhu Zhu<sup>1</sup>, Zhang Gao<sup>2</sup>, Junzhi Yu<sup>1</sup>, Bingwei He<sup>3</sup>, and Jiantao Liu<sup>3</sup>

<sup>1</sup> BIC-ESAT, College of Engineering, Peking University, Beijing, PR China.  
yujunzhi@pku.edu.cn

<sup>2</sup> Institute of Automation, Chinese Academy of Sciences, Beijing, PR China.

<sup>3</sup> the Department of Mechanical Engineering, Fuzhou University, Fuzhou, PR China.

**Abstract.** Guided refinement is a popular procedure of various image post-processing applications. It produces *output image* based on *input* and *guided images*. *Input images* are usually flawed estimates containing kinds of noises and outliers, which undermine the edge consistency between *input* and *guided images*. As improvements, they are refined into *output images* with similar intensities of *input images* and consistent edges of *guided images*. However, outliers are usually untraceable and simply treated as zero-mean noises, limiting the quality of such refinement. In this paper, we propose a general outlier detection method for guided refinement. We assume local linear relationship between *output* and *guided images* to express the expected edge consistency, based on which, the outlier likelihoods of *input* pixels are measured. The metric is termed as ALRe (anchored linear residual) since it is essentially the residual of local linear regression with an equality constraint exerted on the measured pixel. Valuable features of the ALRe are discussed. Its effectiveness is proven by applications and experiment.

**Keywords:** Anchored linear residual; Outlier detection; Guided refinement; Local linear assumption; Linear regression.

## 1 Introduction

Many computer vision and image processing applications require to calculate *output images* fusing the intensities of *input images* and partial edges of *guided images*, such as matting [1], guided smoothing [2], depth map restoration [3], transmission and disparity refinements [4], [5]. The edges of *input images* are expected to be associated with *guided images*. However, it is usually not the case due to theoretical flaws, missing information or noises. As improvements, *output images* are produced based on global optimizations [3], [6], [7] or local filters [8], [9], [10] to ensure both intensity similarity and edge consistency.

Edge consistency is commonly described as “*Output image has an edge only if guided image has an edge*” and controlled in two ways: 1) punishing large edge strength ratios between *output* and *guided images* [2], [11]; 2) assuming local linear relationship between *output* and *guided images* [12].

Despite a clear purpose, outliers are rarely handled. Except for some applications whose outliers could be traced based on models [13] or hardwares [5],

intensity similarities are mostly measured by L2-norms [11], [10]. It implies that all the estimates of *input images* are inliers, and the edge consistency is undermined by zero-mean noises only. This obviously wrong simplification limits the quality of guided refinement.

In this paper, we propose a general outlier detection method for guided refinement referred as ALRe (anchored linear residual). It is shown that the outlier likelihoods provided by ALRe could effectively improve the quality of refinement. It is even comparable to model-based and hardware-based detections. ALRe has three advantages: 1) it does not require any prior knowledge of the application; 2) it has the feature of asymmetry, which exactly expresses the concept of edge consistency; 3) its complexity is  $O(N)$  where  $N$  is the pixel number.

ALRe is based on the local linear assumption between *output* and *guided images*, which has been proven as an effective and general assumption for guided refinement [1], [12]. The more the measured pixel against the assumption, the higher its outlier likelihood. The metric is referred as ALRe because it is based on the residual of local linear regression with an equality constraint exerted on the measured pixel. Compared to linear residual, edge strength ratio and SSIM [14], ALRe has important advantages which are especially suitable for its task.

The remaining contents are arranged as follows. Section 2 surveys guided refinement algorithms, applications and existing solutions for outlier detection. Section 3 proposes the method named ALRe. Section 4 analyzes the relations and differences between ALRe and other applicable methods. Section 5 conducts experiments on various applications and Section 6 provides quantitative results. Section 7 gives the conclusion.

## 2 Related Works

Related works are introduced with respect to four guided refinement algorithms, including WLS (weighted least squares) [11], JBF (joint bilateral filter) [2], GF (guided filter) [12] and WMF (weighted median filter) [15].

WLS (weighted least squares) [11] has a straightforward definition following the concept of guided refinement closely. It is defined as

$$E(q) = \sum_i \underset{\Delta}{w_i} (q_i - p_i)^2 + \lambda \sum_{(i,j) \in J} a_{ij} (q_i - q_j)^2, \quad (1)$$

where  $p$  is *input image*,  $q$  is *output image*,  $i$  is pixel index,  $(i, j) \in J$  means pixels  $i$  and  $j$  are adjacent to each other,  $\lambda$  balances the two terms. There are two kinds of weights in Eq. 1. Smooth weight  $a_{ij}$  is negatively correlated with the distance between  $\mathbf{I}_i$  and  $\mathbf{I}_j$ , where  $\mathbf{I}$  is *guided image*. Data weights  $w$  can be calculated based on the inlier fidelities of  $p$ . The weights is not defined in the original [11], but can be easily inserted without increasing complexity.

In the field of haze removal, Fattal [16] proposed color-lines model for transmission estimation, whose outlier likelihoods are related to the variances of fitting lines. Berman *et al.* [13] proposed haze-lines model. Its outlier likelihoods are

related to the effective lengths of haze-lines. During the refinement, these likelihoods are used to form the data weights  $w$  of Eq. 1, which brings out obvious robustness since pixels not following their models affect the results little. Unfortunately, these outlier detection methods are only applicable to these models, thus can not be generalized to other theories or applications.

JBF [2] is an intuitive and easy-to-implement algorithm for guided refinement. It produces *output images* by smoothing *input images*, thus the intensity similarity is guaranteed. Filter kernels are calculated based on *guided images*, as

$$\begin{cases} q_i = \sum_{j \in \Omega_i} K_{ij} p_j \\ K_{ij} = \frac{1}{Z_i} w_i s_{ij} c_{ij} \\ s_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_s^2}\right) \\ c_{ij} = \exp\left(-\frac{\|\mathbf{I}_i - \mathbf{I}_j\|^2}{\sigma_c^2}\right), \end{cases} \quad (2)$$

where  $\mathbf{x}$  is pixel coordinate,  $Z_i$  is normalizing parameter,  $\Omega_i$  is the local region centered at pixel  $i$ . There are three kinds of weights including data weights  $w$ , distance weight  $s$  and color weight  $c$ . The parameters  $\sigma_s$  and  $\sigma_c$  adjust the sensitivities of the spatial and color similarities respectively. Data weights  $w$  can be calculated based on the inlier fidelities of  $p$ . The weights is not defined in the original [2], but can be easily inserted without increasing complexity.

In the field of disparity estimation, outliers can be detected by cross check [5], which only accepts estimates bidirectionally supported by stereo matching. Pixels have different estimates between left-to-right and right-to-left matchings are considered as outliers, and their  $w$  in Eq. 2 are set as zeros. Unfortunately, this method is also not generalizable because specific hardwares are required.

GF [12] is an efficient algorithm for guided refinement. It assumes local linear relationship between *output* and *guided images*, and conducts linear regression to approach *input images*. Therefore, both intensity similarity and edge consistency are considered. It is defined as

$$\begin{cases} q_i = \frac{1}{|\Omega|} \sum_{k \in \Omega_i} w_{ik}^a (\mathbf{a}_k^T \mathbf{I}_i + b_k) \\ (\mathbf{a}_k, b_k) = \arg \min_{\mathbf{a}_k, b_k} \sum_{j \in \Omega_k} \left( R_{jk}(\mathbf{a}_k, b_k) + \epsilon w_{jk}^w \mathbf{a}_k^T \mathbf{a}_k \right) \\ R_{jk}(\mathbf{a}_k, b_k) = w_j w_{jk}^t (\mathbf{a}_k^T \mathbf{I}_j + b_k - p_j)^2, \end{cases} \quad (3)$$

where  $\mathbf{a}$  and  $b$  are linear parameters,  $\epsilon$  suppresses large  $\mathbf{a}$  for smoothness.

GF are improved in many researches. Anisotropic guided filter [10] introduces the weights  $w_{ik}^a$  and weighted guided filter [8] introduces the weights  $w_k^w$ . Dai *et al.* [9] relaxed local support region  $\Omega$  to the entire image domain, and introduced the weights  $w_{jk}^t$  based on minimum spanning tree. Additionally, some

researches about bidirectional guided filter can be found in [17], [18]. These methods introduce various benefits, such as stronger edge-preserving behavior and less halo-effect. However, the inlier fidelities of each  $p_j$  is not considered, which should be controlled by the weight  $w_j$ .

WMF [15] produces  $q$  by picking values from  $p$ . It is robust to outliers because unpicked pixels have no impact on the result. WMF is defined as

$$\left\{ \begin{array}{l} h(i, v) = \sum_{j \in \Omega_i} w_j w_{i,j} \delta(p_j - v) \\ q_i = v' \\ \sum_{v=l}^{v'} h(i, v) \leq \frac{1}{2} \sum_{v=l}^u h(i, v), \quad \sum_{v=l}^{v'+1} h(i, v) > \frac{1}{2} \sum_{v=l}^u h(i, v), \end{array} \right. \quad (4)$$

where  $\delta(x)$  is 1 if  $x$  equals 0, and is 0 otherwise. The weight  $w_{i,j}$  depends on the input  $\mathbf{I}_i$  and  $\mathbf{I}_j$ . It can be calculated based on the kernel of any edge-ware filter. Despite the robustness, WMF might fail when filter size is large or some outliers happen to be the medians. This problem can be improved if the fidelity of each single pixel is available. It requires the weight denoted as  $w_j$ . It is not originally included in [15], but can be easily realized without increasing complexity.

As introduced, outlier detections for guided refinement are either absent or not generalizable. The weights marked by triangles in Eq. 1, Eq. 2, Eq. 3 and Eq. 4 are only available for specific models and hardwares. We argue that a general outlier detection method should exist since guided refinements have the same purpose whatever the applications. In this paper, we propose ALRe, which to our best knowledge is the first attempt to this problem.

### 3 Anchored Linear Residual

We assume local linear relationship between *ouput images*  $q$  and *guided images*  $\mathbf{I}$ , which means

$$q_i = \mathbf{a}_k^T \mathbf{I}_i + b_k, \quad i \in \Omega_k, \quad (5)$$

where  $(\mathbf{a}_k, b_k)$  are linear parameters,  $\Omega_k$  is the local region centered at pixel  $k$ . If *input images*  $p$  are simply  $q$  affected by zero-mean noise  $n$ , as

$$p = q + n, \quad (6)$$

the optimal  $q$  can be solved by linear regression. However, the residual might be large due to outliers, which significantly against the local linear assumption.

The degree of a given pixel  $k$  against the local linear assumption is measured on  $\Omega_k$  by three steps: 1) assume  $k$  is an inlier with no noise; 2) find the optimal  $q_i$  based on the local linear assumption, where  $i \in \Omega_k$ ; 3) check how well the inlier assumption and local linear assumption are supported. The first step assumes  $k$  is an anchored pixel, which means

$$p_k = q_k = \mathbf{a}_k^T \mathbf{I}_k + b_k. \quad (7)$$

The second step means to find the  $(\mathbf{a}_k, b_k)$  minimizing

$$e_k = \frac{1}{\sum_{i \in \Omega_k} w_i} \sum_{i \in \Omega_k} w_i (q_i - p_i)^2 = \frac{1}{\sum_{i \in \Omega_k} w_i} \sum_{i \in \Omega_k} w_i (\mathbf{a}_k^T \mathbf{I}_i + b_k - p_i)^2, \quad (8)$$

where  $w_i$  is the fidelity of pixel  $i$  being an inlier. The last step is implemented based on the residual  $e_k$ , as

$$w_k = \frac{\frac{1}{\max(\text{LB}, \min(\text{UB}, \sqrt{e_k}))} - \frac{1}{\text{UB}} + \epsilon}{\frac{1}{\text{LB}} - \frac{1}{\text{UB}} + \epsilon}, \quad (9)$$

where (LB, UB) are the bounds of  $\sqrt{e}$ . When  $e_k$  is out of the bounds, pixel  $k$  is considered as pure inlier and outlier respectively.  $\epsilon$  is a small number for numerical stability. In this paper, (LB, UB,  $\epsilon$ ) equals (0.01, 0.3, 0.001).

With some algebraic manipulations on Eq. 7 and Eq. 8, it yields

$$\begin{cases} \mathbf{C}_k = \sum_{i \in \Omega_k} w_i (\mathbf{I}_i - \mathbf{I}_k) (\mathbf{I}_i - \mathbf{I}_k)^T \\ \mathbf{a}_k = \mathbf{C}_k^{-1} \sum_{i \in \Omega_k} w_i (p_i - p_k) (\mathbf{I}_i - \mathbf{I}_k) \cdot \\ \mathbf{b}_k = p_k - \mathbf{a}_k^T \mathbf{I}_k \end{cases} \quad (10)$$

In programming, it is

$$\begin{cases} \mathbf{C}_k = (\overline{w\mathbf{I}\mathbf{I}^T})_k + (\overline{w})_k \mathbf{I}_k \mathbf{I}_k^T - (\overline{w\mathbf{I}})_k \mathbf{I}_k^T - \mathbf{I}_k (\overline{w\mathbf{I}^T})_k + \boldsymbol{\epsilon} \\ \mathbf{a}_k = \mathbf{C}_k^{-1} \left( (\overline{wp\mathbf{I}})_k - p_k (\overline{w\mathbf{I}})_k - (\overline{wp})_k \mathbf{I}_k + (\overline{w})_k p_k \mathbf{I}_k \right) \\ b_k = p_k - \mathbf{a}_k^T \mathbf{I}_k \\ e_k = \left( \mathbf{a}_k^T (\overline{w\mathbf{I}\mathbf{I}^T})_k \mathbf{a}_k + (\overline{w})_k b_k^2 + (\overline{wp^2})_k + 2b_k \mathbf{a}_k^T (\overline{w\mathbf{I}})_k \right. \\ \left. - 2\mathbf{a}_k^T (\overline{wp\mathbf{I}})_k - 2b_k (\overline{wp})_k \right) / \left( (\overline{w})_k + \epsilon \right) \end{cases}, \quad (11)$$

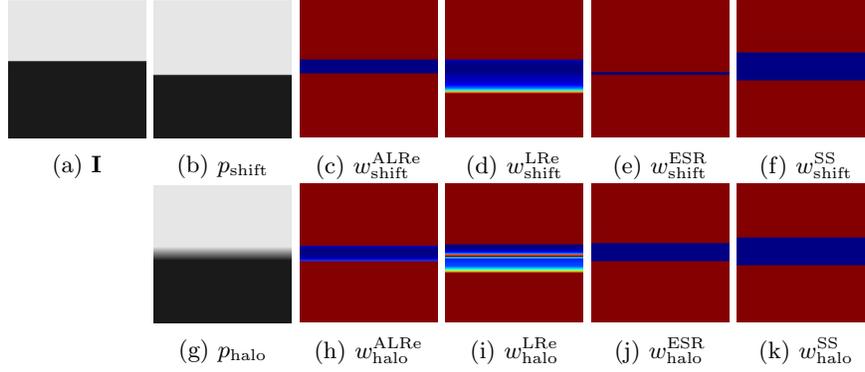
where  $\boldsymbol{\epsilon}$  is a diagonal matrix whose elements all equal  $\epsilon$ ,  $(\overline{p})_k$  is the mean value of  $p$  in  $\Omega_k$ , so do the others. The deduction is similar to GF [12], we recommend it to readers who need more details.

Since  $e$  and  $w$  are interdependent, we use an iteration strategy as  $w^t \rightarrow e^{t+1} \rightarrow w^{t+1} \rightarrow \dots$ , where  $w^0 = 1$ . The terminal condition is

$$\Delta e^{t+1} = \sum_k |e_k^{t+1} - e_k^t| < \epsilon. \quad (12)$$

In practice, it takes 5~10 iterations. The final ALRe of pixel  $k$ , that is  $e_k$ , represents its outlier likelihood. In most algorithms, the inlier fidelity  $w_k$  is preferred.

Note that, the mean values in Eq. 11 for all the  $k$  can be calculated by boxfilter with  $O(N)$  complexity, where  $N$  is the pixel number. The number of iterations is independent of  $N$ . Therefore, the algorithm is  $O(N)$  overall.



**Fig. 1. Comparison of outlier detection methods.** Weight maps are displayed in color. Warmer the color, larger the value. The edge of  $p_{\text{shift}}$  is misaligned with  $\mathbf{I}$ , and  $p_{\text{halo}}$  has halo-effect.

## 4 Analysis

### 4.1 Invariance and Asymmetry

ALRe is invariant to linear transforms on *guided images* and shifts on *input images* because

$$e(\alpha_p p + \beta_p, \alpha_{\mathbf{I}} \mathbf{I} + \beta_{\mathbf{I}}) \equiv \alpha_p^2 e(p, \mathbf{I}), \quad (13)$$

where  $(\alpha_p, \alpha_{\mathbf{I}}, \beta_p, \beta_{\mathbf{I}})$  are scalars satisfying  $\alpha_{\mathbf{I}}^2 \neq 0$ . It can be proven as

$$\begin{cases} \tilde{\mathbf{I}} = \alpha_{\mathbf{I}}^T \mathbf{I} + \beta_{\mathbf{I}} \\ \tilde{p} = \alpha_p p + \beta_p \end{cases} \Rightarrow \tilde{\mathbf{C}} = \alpha_{\mathbf{I}}^2 \mathbf{C} \Rightarrow \begin{cases} \tilde{a} = \frac{\alpha_p}{\alpha_{\mathbf{I}}} a \\ \tilde{b} = \alpha_p b \end{cases} \Rightarrow \tilde{e} = \alpha_p^2 e. \quad (14)$$

Eq. 13 also reveals the asymmetry of ALRe, which fulfills the concept of edge consistency. Given a pair of *input* and *guided images*, the sharpness can be tuned by  $\alpha_p$  and  $\alpha_{\mathbf{I}}$ . When  $\alpha_p$  is large, *input image* has sharp edges, ALRe is small only if *input* and *guided images* closely follow the local linear assumption due to the large  $\alpha_p^2$ . This fits the description “*Output image has an edge only if guided image has an edge*”. When  $\alpha_p$  is small, *input image* is smooth, ALRe is small because of the small  $\alpha_p^2$ . The sharpness of *guided image* is unessential. This fits another fold of the description, as “*Output image can be smooth whether guided image has an edge or not*”. In most applications, *guided image* has more edges than *input image*, but it will not lead to large ALRe because of this asymmetry.

### 4.2 vs. Linear Residual

Now contemplate the necessity of the iteration framework and equality constraint, without which, the algorithm degenerates to a single linear regression and causes following problems.

1. If the outlier likelihood of  $k$  is investigated based on the linear residual of  $\Omega_k$ , then the outliers will undermine the fidelities of all its neighbors;
2. If the outlier likelihood of  $k$  is investigated based on the distances between  $p_k$  and the optimal fitted result of  $\Omega_k$ , then the residual will be small if  $p_k \approx (\bar{p})_k$  and  $\mathbf{I}_k \approx (\bar{\mathbf{I}})_k$  even though  $p_k$  might be an obvious outlier.

The first problem can be seen in Fig. 1d, where the blue belt is over wide. As a comparison, the belt in Fig. 1c exactly cover the misaligned region. The second problem can be seen in Fig. 1i, where the pixels in the middle of the halo have high fidelities. As shown in Fig. 1h, with the equality constraint, these fake inliers disappear because the result crossing them can not fit other samples well.

### 4.3 vs. Edge Strength Ratio

The smoothness term of WLS [11] might be the simplest and most straightforward definition of edge consistency. It can be implemented as

$$w_k^{\text{ESR}} = \frac{|G(q_k)| + \epsilon}{|G(\mathbf{I}_k)| + \epsilon}, \quad (15)$$

where  $|G(x)|$  is the gradient module of  $x$ . As shown in Fig. 1,  $w^{\text{ESR}}$  reveals the misaligned edge in  $p_{\text{shift}}$  and the halo in  $p_{\text{halo}}$ . However, the result Fig. 1e can not help guided refinement algorithms to improve edge consistency. As a comparison,  $w^{\text{ALRe}}$  correctly reveals all the pixels need to be changed.

### 4.4 vs. SSIM

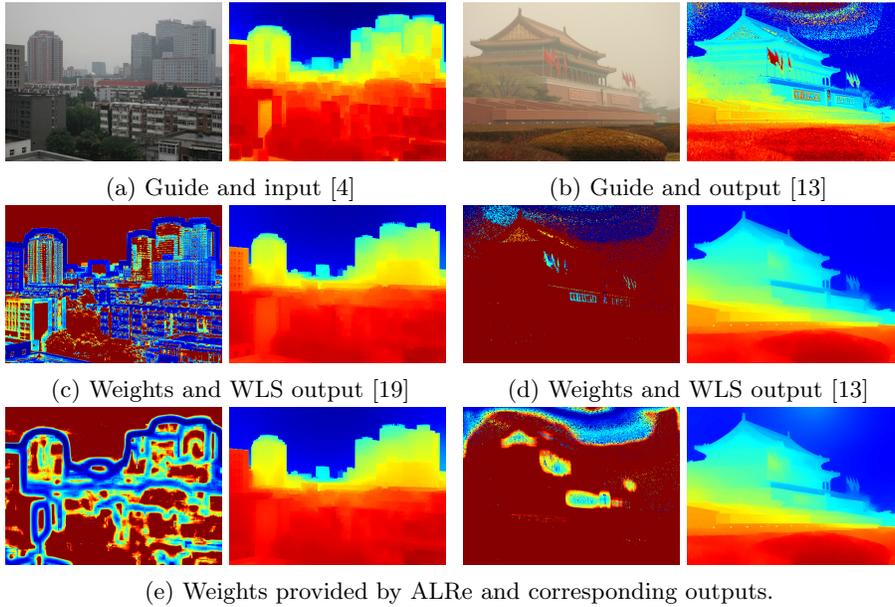
Structure similarity is one of the three similarity indexes of SSIM [14] defined as

$$w_k^{\text{SS}} = 1 - \frac{\sigma_k^{pI} + \epsilon}{\sigma_k^p \sigma_k^I + \epsilon}, \quad (16)$$

where  $(\sigma_k^p, \sigma_k^I)$  are standard deviations of  $p$  and  $I$  (single channel) in  $\Omega_k$ ,  $\sigma_k^{pI}$  is cross-covariance. Despite the outstanding performance in various fields, SSIM is not the right method for detecting outliers of edge consistency. Firstly, it lacks the feature of asymmetry. Secondly, outliers will undermine the SSIM of all its neighbors as shown in Fig. 1f and Fig. 1k, where the blue belts are over wide.

## 5 Applications

ALRe is tested by four applications. The guided refinement algorithms in Sec. 2 are employed. ALRe improves these algorithms by offering per-pixel inlier fidelities as their data weights  $w$  (the ones marked by triangles).



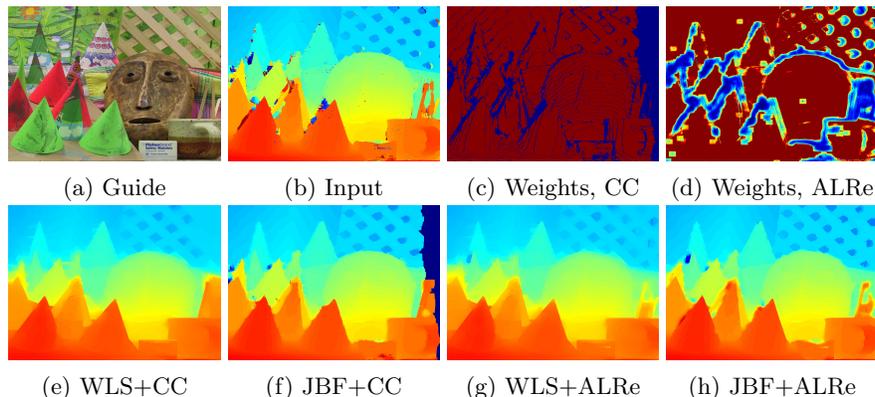
**Fig. 2. Transmission map refinements.** (a,b) Hazy images, and the initial transmission maps based on dark channel prior [4] and haze-line model [13] respectively; (c,d) The weight maps and the refined results based on Zhu *et al.* [19] and Berman *et al.* [13] respectively; (e) The weight maps based on ALRe and corresponding results.

### 5.1 Transmission Refinement for Haze Removal

In the field of haze removal, hazy images are considered as haze-free images attenuated by atmospheric lights, and transmission maps represent the attenuation ratios. With evenly dispersed haze, attenuation ratios are related to scene depths. Therefore, transmission edges should be consistent with depth edges. Since depth edges mostly happens on color edges, transmission maps are expected to have edge consistency with hazy images.

Limited by existing technologies, transmission maps usually have unsatisfactory edge consistency, thus a refinement guided by hazy images is popular [4], [13], [16]. In Fig. 2a, the *input image* is produced based on the dark channel prior [4]. It has outliers named block effect, which indicates the over-estimated transmissions in the vicinity of large depth jumps. Zhu *et al.* [19] detects these outliers based on an improved local constant assumption. The result is shown in Fig. 2c. In Fig. 2b, the *input image* is based on the haze-line model [13]. Outliers exist due to short haze-lines, which are traceable as shown in Fig. 2d. Benefiting from the weights, WLS erases these outliers well.

ALRe is able to detect these outliers without any prior knowledge of block effect or haze-line model. In Fig. 2e, ALRe predicts both kinds of outliers correctly. The results are almost the same with Fig. 2c and Fig. 2d.



**Fig. 3. Disparity map refinements.** (a) Color image; (b,c) The initial disparity map based on Hosni *et al.* [5] and the weight map based on cross check (both require an image from another view); (d) The weight map based on ALRe; (e,f) The refined results based on WLS and JBF with cross check. (g,h) Corresponding results with ALRe.

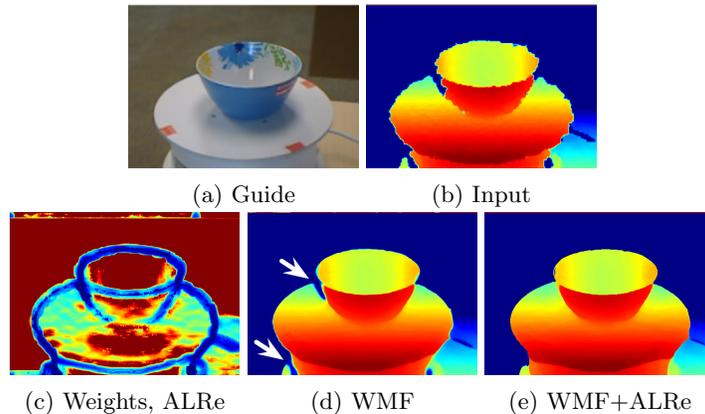
## 5.2 Disparity Refinement for Depth Estimation

Disparity refers to the difference in image locations of a point seen by different cameras. Disparity maps are inversely proportional to depth maps, whose edges are mostly consistent with color edges. Therefore, they are also expected to have edge consistency with color images. Disparities can be estimated by stereo matching [5]. However, it might be false or invalid on several pixels due to occlusions. An example of Middlebury dataset 2003 [20] is shown in Fig. 3a. The initial disparity map in Fig. 3b is generated by Hosni *et al.* [5]. Outliers against edge consistency can be seen near the edges. Hosni *et al.* [5] traces them by cross check, which requires the disparity map from another view.

ALRe could detect these outliers without another view. The result is displayed in Fig. 3d, where low weights are in the right positions referring to the binary result of cross check in Fig. 3c. The refined results are similar with each other. Clearly, cross check is more reliable than ALRe because of the extra information. However, extra information are not always available. For example, RGB-D camera usually provides depth maps with misaligned edges as shown in Fig. 4b. This problem can be solved by WMF. As shown in Fig. 4d, the winding edges are well regularized, but the values of the pointed regions are wrong picked. These regions have zero values because they are invisible to the depth camera, thus should not be considered in median calculation. With the help of ALRe, these regions are trivial in WMF and a more convincing result is achieved.

## 5.3 Guided Feathering

Guided feathering produces an alpha matte of complex object boundaries based on rough binary mask, which can be obtained manually or from other segmentation methods. GF is an efficient tool for this task but not error-tolerant enough,



**Fig. 4. Depth map refinements.** (a) Color image; (b) Rough depth map; (c) The weight map based on ALRe; (d) The refined result based on WMF without ALRe; (e) Corresponding result with ALRe.

thus masks with large errors might lead to halo-effects. As shown in Fig. 5d, the result of GF inherits the over-estimated and under-estimated errors marked by A and B. As displayed in Fig. 5c, the weights are all low near the boundaries. Furthermore, phantom of the edges from both images can be observed, and the regions between them have almost zero fidelities. With this message, a more convincing matte is produced as shown in Fig. 8e.

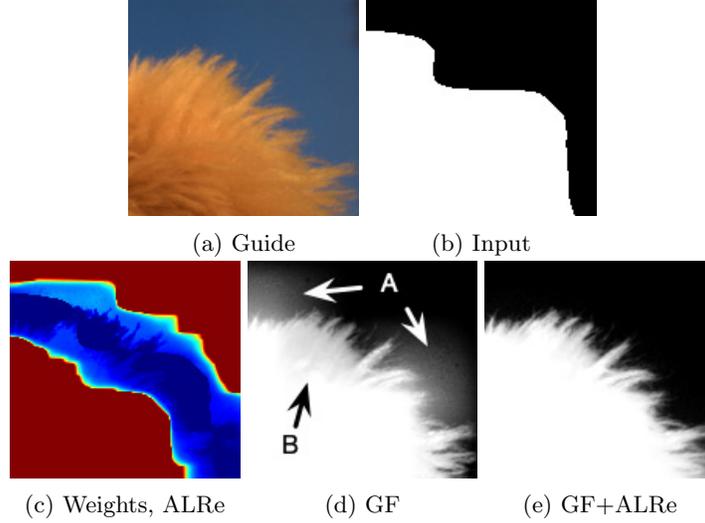
#### 5.4 Edge-preserving smoothing

As an edge-preserving filter, GF has the problem of halo-effects. Various methods have been proposed to solve this problem by introducing weights [8], [9], [10], as listed in Eq. 3. In this paper, we address edge-preserving smoothing in the view of guided refinement. The sharp input is firstly smoothed by Gaussian low-pass filter. Then, the smoothed result is refined back by GF in the guidance of the sharp input. Since weak noises are mostly erased by low-pass filter, they are not enhanced in guided refinement. On the other hand, sharp edges are turned into weak edges and halos, leaving clues for GF to enhance them back. ALRe could recognize these sharp edges based on the halos, as shown in Fig. 6c. The result is shown in Fig. 6e, where the halo-effects are trivial.

## 6 Experiment

In this section, we compare the outlier detection accuracies of WMF and ALRe on synthetic images. The IoU (intersection over union) of outlier detection result and groundtruth is used for evaluation, as

$$\text{IoU}(\text{GT}, \text{MASK}) = \frac{|\text{GT} \cap \text{MASK}|}{|\text{GT} \cup \text{MASK}|}. \quad (17)$$



**Fig. 5. Feathering.** (a) Color image; (b) Rough mask; (c) The weight map based on ALRe; (d) The result based on GF without ALRe; (e) Corresponding result with ALRe.

MASK and GT are binary images, whose pixel  $x$  equals 1 if  $p(x)$  is asserted as an outlier. The MASK of ALRe is

$$\text{MASK}_{\text{ALRe}}(p, \mathbf{I})_k = \begin{cases} 1, & w(p, \mathbf{I})_k < 0.05 \\ 0, & w(p, \mathbf{I})_k \geq 0.05 \end{cases}. \quad (18)$$

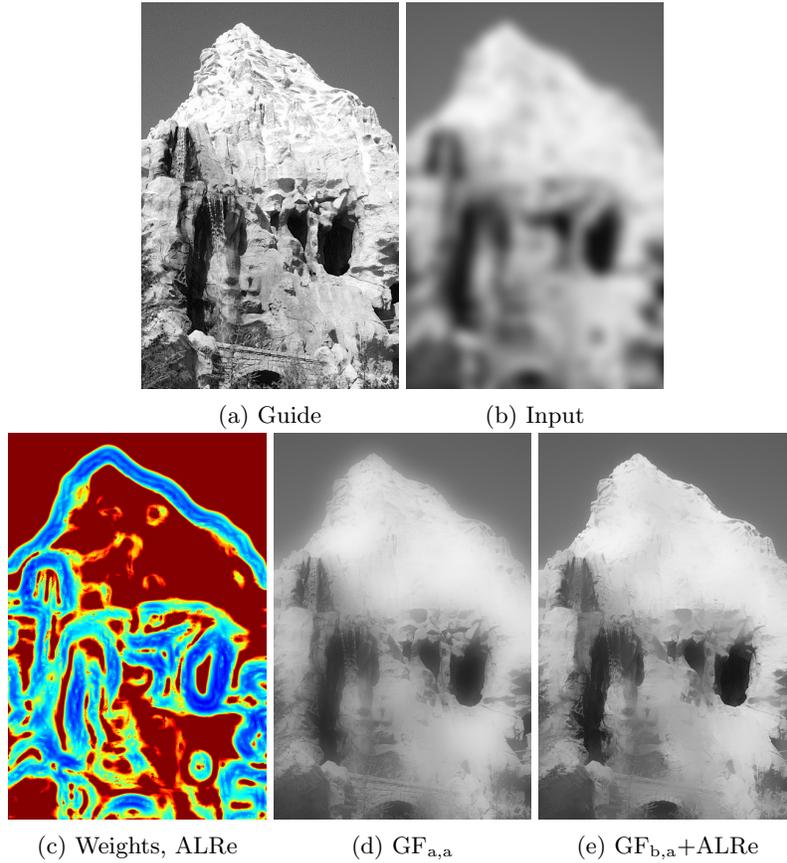
The MASK of WMF is produced based on the intuition that pixels greatly changed after filtering are outliers, thus

$$\text{MASK}_{\text{WMF}}(p, \mathbf{I})_k = \begin{cases} 1, & |\text{WMF}(p, \mathbf{I}) - p|_k > 0.3 \\ 0, & |\text{WMF}(p, \mathbf{I}) - p|_k \leq 0.3 \end{cases}. \quad (19)$$

Input images  $\mathbf{I}$  are provided by Middlebury dataset 2014 [21]. Contaminated input  $p$  and GT are produced by following steps.

1. generate random  $\mathbf{a}_k$  with elements in  $[0, 1/3]$  and random  $b_k$  in  $[-1, 1]$ ;
2. smooth  $\mathbf{a}$  and  $b$  based on  $R \times R$  boxfilter;
3. calculate  $q$  based on  $q_k = \mathbf{a}_k^T \mathbf{I}_k + b_k$ , rescale  $q$  to  $[0, 1]$ ;
4. generate  $M$  random regions with heights in  $[0, 15]$  and lengths in  $[5, 105]$  centered at random pixels, pixels of GT in these regions equal 1;
5.  $p_k = q_k + 0.5\text{GT}$  for  $q_k < 0.5$ , and  $p_k = q_k - 0.5\text{GT}$  for  $q_k \geq 0.5$ .

A small  $R$  leads to noisy  $\mathbf{a}$  and  $b$ , undermining the local linear assumption between  $q$  and  $\mathbf{I}$ . The number of random regions  $M$  affects the number of outliers. Therefore, the task is more challenge with smaller  $R$  and larger  $M$ . We test 22



**Fig. 6. Edge preserving filtering.** (a) Sharp image; (b) The smoothed image based on Gaussian low-pass filter; (c) The weight map of (b) based on ALRe; (d) The smoothed result of (a) guided by itself based on GF; (e) The enhanced result of (b) guided by (a) based on GF with ALRe.

kinds of  $R$  and 3 kinds of  $M$ , as  $R = [15, 19, 23, \dots, 99]$  and  $M = [50, 100, 200]$ . The size of  $\Omega$  for ALRe and WMF is  $25 \times 25$  (the inputs are  $640 \times 480$ ). Larger  $R$  leads to better local linear relationship, thus higher  $\text{mean}(w_{ALRe}(q, \mathbf{I}))$ . It is the basic of ALRe, therefore, we term it as ALRe expectation.

The results are illustrated in Fig. 7. As shown, both ALRe and WMF are barely affected by ALRe expectation, but degraded when  $M$  increases. The mean IoUs of ALRe corresponding to the increasing  $M$  are 0.978, 0.958 and 0.868, while the ones of WMF are 0.814, 0.727, 0.556. The gap is about 0.2. An example is displayed in Fig. 8.  $\text{MASK}_{ALRe}$  is almost the same with GT when  $R=99$ ,  $M=50$ . In the case of  $R=15$ ,  $M=200$ , the output is severely contaminated. However,  $\text{MASK}_{ALRe}$  is still plausible. As a comparison,  $\text{MASK}_{WMF}$  in both cases contain undesired hollows, where the outliers should be changed by 0.5 but less than 0.3.

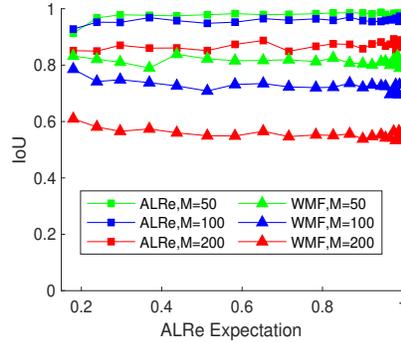


Fig. 7. Outlier detection accuracy.

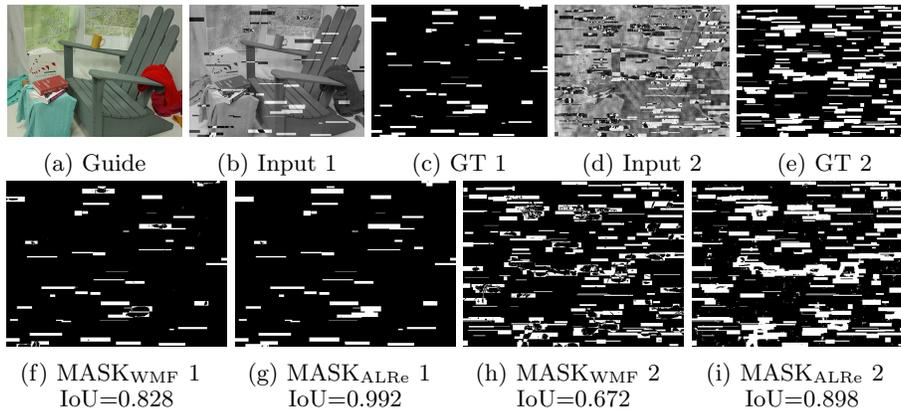


Fig. 8. Comparison of ALRe and WMF on outlier detection. (a) Guide image; (b) Contaminated input image ( $R = 99, M = 50$ ); (c, f, g) Corresponding GT, MASK<sub>WMF</sub> and MASK<sub>ALRe</sub>; (d) Contaminated input image ( $R = 15, M = 200$ ); (e, h, i) Corresponding GT, MASK<sub>WMF</sub> and MASK<sub>ALRe</sub>;

## 7 Conclusion

In this paper, we propose a general outlier detection method for guided refinement, termed ALRe. Different from the traditional model-based and hardware-based detections, ALRe produces meaningful predictions without any prior knowledge. Valuable features such as asymmetry and linear complexity are achieved. The effectiveness of ALRe is verified based on four applications and four guided refinement algorithms. It shows that the weight map provided by ALRe is valuable and even comparable to customized methods. Quantitative comparison reveals that ALRe could detect outliers accurately, even though the image and the local linear relationship are severely contaminated. Its accuracy represented by IoU is about 0.2 higher than weighted median filter.

## Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1312800 and in part by the National Natural Science Foundation of China under Grant U1909206.

## References

1. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(2) (2008) 228–242
2. Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., Toyama, K.: Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics* **23**(3) (2004) 664–672
3. Liu, W., Chen, X., Yang, J., Wu, Q.: Robust color guided depth map restoration. *IEEE Transactions on Image Processing* **26**(1) (2016) 315–327
4. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(12) (2011) 2341–2353
5. Hosni, A., Rhemann, C., Bleyer, M., Rother, C., Gelautz, M.: Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(2) (2012) 504–511
6. Park, J., Kim, H., Tai, Y.W., Brown, M.S., Kweon, I.: High quality depth map upsampling for 3d-tof cameras. In: *International Conference on Computer Vision*. (2011) 1623–1630
7. Yang, J., Ye, X., Li, K., Hou, C., Wang, Y.: Color-guided depth recovery from rgb-d data using an adaptive autoregressive model. *IEEE Transactions on Image Processing* **23**(8) (2014) 3443–3458
8. Li, Z., Zheng, J., Zhu, Z., Yao, W., Wu, S.: Weighted guided image filtering. *IEEE Transactions on Image Processing* **24**(1) (2014) 120–129
9. Dai, L., Yuan, M., Zhang, F., Zhang, X.: Fully connected guided image filtering. In: *IEEE International Conference on Computer Vision*. (2015) 352–360
10. Ochotorena, C.N., Yamashita, Y.: Anisotropic guided filtering. *IEEE Transactions on Image Processing* **29** (2019) 1397–1412
11. Farbman, Z., Fattal, R., Lischinski, D., Szeliski, R.: Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics* **27**(3) (2008) 1–10
12. He, K., Sun, J., Tang, X.: Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(6) (2013) 1397–1409
13. Berman, D., Avidan, S.: Non-local image dehazing. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2016) 1674–1682
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**(4) (2004) 600–612
15. Ma, Z., He, K., Wei, Y., Sun, J., Wu, E.: Constant time weighted median filtering for stereo matching and beyond. In: *IEEE International Conference on Computer Vision*. (2013) 49–56
16. Fattal, R.: Dehazing using color-lines. *ACM Transactions on Graphics* **34**(1) (2014) 1–14

17. Guo, X., Li, Y., Ma, J., Ling, H.: Mutually guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(3) (2018) 1283–1290
18. Li, Y., Huang, J.B., Ahuja, N., Yang, M.H.: Joint image filtering with deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**(8) (2019) 1909–1923
19. Zhu, M., He, B., Liu, J., Yu, J.: Boosting dark channel dehazing via weighted local constant assumption. *Signal Processing* **171** (2020) 107453
20. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2003) I–I
21. Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P.: High-resolution stereo datasets with subpixel-accurate ground truth. In: *German Conference on Pattern Recognition*. (2014) 31–42