Abstract—Depth maps captured by the range imaging sensors such as ToF (time of flight) camera and Kinect are stuck with limited spatial resolution and varieties of noises, which makes it difficult to be directly applied to 3D scene analysis. In this paper, we address these issues via an extended weighted mode filter (EWMF). In view of the impressive feature of the noise-aware filter in noise suppression, the proposed method synergistically combines standard weighted mode filter (WMF) and the noise-aware filter to achieve a better noise suppression performance. Different from conventional filtering-based methods with a fixed support window, a refined adaptive support window (RASW) is designed. The proposed filter with RASW can well capture local structure details better. Experimental results demonstrate that the proposed method outperforms several state-of-the-art super-resolution techniques in terms of bad pixel rate and root mean square error.

Index Terms—super-resolution; weighted mode filter; noise suppression; refined adaptive support window

I. INTRODUCTION

Limited by the current imaging technology of depth sensors, it’s difficult to achieve a precise and high-resolution depth image directly. Depth maps captured by depth sensors, like ToF cameras are stuck with the spatial resolution and noise. These deficiencies have impeded further applications of similar range sensors. The subject of depth map enhancement and super resolution has attracted more and more attention.

Lots of related research works have been done and previous attempts can be classified into two groups: optimization-based methods and filtering-based methods. In the first group, optimization-based methods can be deemed energy minimization approaches. This category shares similarity that an energy function is designed and minimized. A landmark work was done in [1]. Markov random fields (MRF) were employed to handle the problem of generating high-resolution range images. The object function they built contains a data and smooth term. After this, many works [2, 3, 4] based on the MRF framework were published. Park et al. [5] described another regularization framework with weight least square (WLS). They expanded the regularization term with an additional edge weighting scheme. Different from the above two frameworks, a novel convex optimization framework utilizing higher order regularization was presented in [6]. The strategy benefited from its parallel mode and ran at a video rate. Among filtering-based methods, a pioneering work is joint bilateral upsampling (JBU) [7] which extended from the bilateral filter [8]. Since then, a series of remarkably filter such as geodesic filter [9], guided filter [10] and its extended version [11] were presented. Min et al. [12] proposed a weighted mode filter based on the joint histogram for increasing the spatial resolution and suppressing the noise.

Our proposed method is inspired by the weighted mode filter [12]. The main contributions of this paper can be summarized as follows:

• Based on the relation between WMF and JBU in essence, we propose an enlightening proposition that standard WMF and JBU share some special-purpose improvement features.

• Inspired by the noise-restraining mechanism of the noise-aware filter (NAF) [13], we synergistically combine standard WMF and NAF to improve the noise suppression performance.

• A refined adaptive support window similar to [14] is presented to compute the similarity weight, which is effective to capture local structure details.

The rest of the paper is organized as follows: In Section 2, we present a simple revisit from JBU to WMF. Limitations of standard WMF are also discussed in this section. The details of EWMF and RASW are described in Section 3. Finally, we present experimental results and conclusions in Section 4 and 5, respectively.

II. WEIGHTED MODE FILTER

The WMF based on the joint histogram is originally proposed by Min et al. They derived the relation with JBU and point out that WMF seeks global mode on joint histogram by leveraging similarity measure between neighboring pixels in a fixed support region. To better understand the following content, we give a simple revisit of the derivation.

A. A revisit from JBU to WMF

A bilateral filter has two filter kernels, which are known as the spatial filter kernel and range filter kernel for distance measuring. The core idea of JBU is that its range filter is computed from the high-resolution guidance image. That is, a spatial filter $f(\cdot)$ and range filter $g(\cdot)$ are operated at two different resolutions simultaneously. Thus, the result of the filter with a support area $N(p)$ around $p$ is:

$$\hat{D}(p) = \frac{1}{k(p)} \sum_{q \in N(p)} d(q, f[p, q], g[\|f(p) - I(q)\|])$$  \hspace{1cm} (1)
where \( p \) and \( q \) denote the corresponding coordinates of pixels in the low resolution (LR) depth map \( d(\cdot) \), \( I(p) \) and \( I(q) \) represent the pixel values located at \( p \) and \( q \) of the high resolution image. \( k(p) \) is the normalizing factor.

In (1), the spatial filter and range filter generally utilize a truncated Gaussian to measure the spatial and range distance. Let \( w(p,q) = G_r(p_i - q_i)G_s(I(p) - I(q)) \). Here, \( G_r \) and \( G_s \) represent the spatial and range kernel, respectively. Thus, JBU can be rewritten as

\[
\hat{D}(p) = \sum_{q \in N(p)} w(p,q)\delta(q_i) / \sum_{q \in N(p)} w(p,q)
\]

(2)

We slightly adjusted the original WMF to a better comparison. \( G_r \) (a Gaussian function with smoothing parameters \( \sigma_r \)) in original WMF is replaced with two order polynomial. For a pixel \( p \) in the HR image, similarity weight between \( p \) and its neighboring pixels can be expressed as a relaxed histogram. By seeking the global mode on histogram, the final solution \( \hat{D}(p) \) at \( \delta \)th bin can be computed as follows:

\[
H_G(p,d) = \sum_{q \in N(p)} w_G(p,q)\left(af(q)^2 + b(d)f(q) + c(d)\right)
\]

\[
\hat{D}(p) = \arg \max H_G(p,d)
\]

where \( a \) is a constant, \( b(d) = 2ad \) and \( c(d) = -ad^2 \).

To maximize (3), we take the first derivative with respect to \( d \) and get

\[
\frac{\partial H_G(p,d)}{\partial d} = 2a \sum_{q \in N(p)} w(p,q)(f(q) - d)
\]

(4)

By solving this derivative equation, we can get a solution share a similar expression with (2). That is, WMF can also be viewed as a variant of JBU. The main difference is WMF selects the optimal solution among all the depth candidates instead of summing all candidates [12].

B. Limitations of WMF

Although original WMF obtain visually pleasing results, there still remain some matters to iron out.

- Regardless of the multiscale color measure (MCM), the WMF itself doesn’t have an adaptive mechanism to suppress noise.
- As a filtering-based method, the support window of standard WMF is fixed. Obviously, it can’t realize the optimal weight assignments.

III. EXTENDED WEIGHTED MODE FILTER

Aiming at the issues described above, we proposed the extended WMF to improve them. Inspired by NAFDU [13] and its noise-aware mechanism, we combine the standard WMF and NAF to achieve a better noise suppression performance. Also, we design a refined adaptive support window (RASW) to capture local structure details better.

A. Fusion of Standard WMF and NAFDU

To address the above issues, an extended weighted mode filter is presented in the following section. Inspired by the theoretical derivation describing the relation between JBU and WMF, we reason that improvements designed for JBU can be applied to WMF. Before sharing details of EMWF, NAFDU [13] is introduced:

\[
\hat{D}(p) = \frac{1}{k_r \cdot \sum_{q \in N(p)} d(q_i)} f(\|p_i - q_i\|\alpha(\Lambda_n)g(\|p\| - I(q))

+ (1 - \alpha(\Lambda_n))h(\|p_i - d(q_i)\|)
\]

(5)

where \( \alpha(\cdot) \) denotes the blending function and \( \Lambda_n \) is the depth difference between the minimum and maximum depth values in the each filter kernel. \( h(\cdot) \) is a modified range term.

Following the regulatory mechanism of NAFDU, we get the extended WMF histogram:

\[
H_e(p,d) = \sum_{q \in N(p)} w_e(p,q)G_r(d - D(q))
\]

with the extended weight is computed as follows:

\[
w_e(p,q) = \alpha(\Lambda_n)f(p_i - q_i)g(I(p) - I(q)) + \sum_{q \in N(p)} (1 - \alpha(\Lambda_n))f(p_i - q_i)h(d(p_i) - d(q_i))
\]

(7)

After the fusion of WMF and NAF, beneficial properties of bilateral upsampling in those areas where the color edge co-occurs with the depth discontinuity can be preserved. Meanwhile, the regulatory mechanism can prevent artifacts in those areas where standard WMF is likely to cause erroneous texture copy. The performance evaluation on noise suppression is given in Section 4.

B. Refined adaptive support window

Key factors of the local filter-based methods are the support region and support weight calculation. However, most local methods [7, 8, 10, 11, 12, 13] utilize square supports which are very likely to blur the depth discontinuity. We cite a simple example to illustrate the issue. The original depth map may be contaminated by severe noise, which makes a big change to some depth values in the support window. Standard WMF is no exception, and a fixed neighboring window is employed to compute the weight similarity. Based on the above, RASW is designed with inspiration from cross-based local multipoint filter (CLMF) [14]. In original CLMF, the order of performing horizontal or vertical extension is determined subjectively without available base. Different from previous method, we design a reasonable order based on color similarity. Before presenting the proposed strategy, we reiterate that RASW is designed to ensure depth values of pixels inside the support region are as similar as possible. For the center pixel \( p \), we use the binary mask \( M(p) = 1 \) to denote vertical extension first. Based on the heuristic assumption that pixels with similar color have similar depth values, \( M(p) \) can be determined as follows:

\[
M(p) = \begin{cases} 
1, & S_y - S_y > 0 \\
0, & \text{otherwise}
\end{cases}
\]

(8)
where \( S_p \) and \( S_q \) represent the custom variance measures how far the pixels in boundary of square window are spread out with respect to the center pixel \( p \).

The next thing is to determine the cross-based adaptive kernel. Four support arms of the corresponding center pixel \( p \) are computed based on color similarity. The support arms extend horizontally and vertically until a critical condition is triggered:

\[
|I_q(p) - I_c(p)| \leq \tau, \quad c \in \{R,G,B\}, \quad q \in W_p, \quad (9)
\]

where \( I_q \) is the intensity of the color band \( c \) of the \( 3 \times 3 \) media smoothed input image \( I_c \), \( q \) represents a pixel located at the corresponding arm, \( W_p \) is a square window centered at \( p \) with the size of \((2r+1) \times (2r+1)\), \( \tau \) controls the confidence level of the color similarity.

Suppose \( M(p) = 1 \). As shown in Fig. 1(a), extension is performed horizontally then vertically. Another case is illustrated in Fig. 1(b). Let \( H(p) \) and \( V(p) \) denote all the pixels covered by the horizontal and vertical arms of the pixel \( p \), respectively. The lengths of vertical arms \( V^L_p \) and \( V^R_p \) are determined first. Then the horizontal arms are extended until terminal pixels \( H^B_p \) and \( H^L_p \) are reached. Thus, the support arms can be described as follows:

\[
H(p) = \bigg\{ (u,v) \bigg| u \in [u_p - H^B_p, u_p + H^B_p], v = v_p \bigg\} \quad V(p) = \bigg\{ (u,v) \bigg| v \in [v_p - V^L_p, v_p + V^L_p], u = u_p \bigg\} \quad (10)
\]

where \( u_p \) and \( v_p \) represent the image coordinates in the square window.

The horizontal arms \( H(p) \) slides along the vertical arms \( V(q) \), forming the final RASW (the region inside the red boundary in Fig. 1):

\[
\Omega_p = \bigcup_{q \in V(p) \cap M(p) \neq M(p) \cap H(p)} V(q)(1-M(p)) + H(p)M(p) \quad (11)
\]

![Fig. 1](image-url)

Fig. 1  The modified cross-based adaptive kernel. (a) The extension is performed first in the vertical direction. (b) The extension is performed horizontally then vertically.

### IV. EXPERIMENTAL RESULTS

In this section, we test the proposed method on the Middlebury benchmark [15]. To evaluate the effectiveness of the proposed method in terms of edge-preserving and noise suppression performance, we compare it with some existing popular methods, i.e., JBU [7], NAFDU [13], and standard WMF (WMF) [12]. The root mean squared error (RMSE) [12] and bad pixel rate (BP\%) [15] are employed as the indicator for quantitatively assessing the performance. Note that the error threshold of BP is set at 1 and its calculations can reference [15]. We downsample the ground truth depth maps into LR depth maps for super-resolution experiments.

All four methods are implemented in MATLAB and the multiscale color model (MCM) in original work [12] is discarded to merely compare the performance of filter. That is, the sparse original depth pixels mapped into the color camera coordinate and the holes are filled by WMF. In our experiments, the fixed support window is chosen as a \( 10 \times 10 \) square patch, the standard deviations in the EWMF are set as: range sigma \( \sigma_r = 2 \), spatial sigma \( \sigma_s = 20 \) and smoothing parameter \( \sigma_f = 5 \). We set \( \sigma_r = 8 \) in the presence of noise. The corresponding parameters in the other three methods are same with EWMF. All four methods are tested with the fixed parameters for all images selected. The comparison results of four datasets, Teddy, Aloe, Midd1 and Baby2 are illustrated in Fig. 2. Summary of comprehensive comparison in terms of RMSE and BP\% of different images is given in Table I and Table II.

Further experiments are performed to evaluate the effectiveness of the depth value rectification of the proposed EWMF. In follow-up experiments, additive salt&pepper noise (the noise density being 0.01) is added to the datasets. The corresponding quantitative comparison is given in Table III and Table IV. Note that the superscript of EMWF 1 and 2 represent the method utilizing the fixed support window and RASW, respectively.

| Method | Teddy | Aloe | Midd1 | Baby2 |
|--------|-------|------|-------|-------|
| JBU    | 3.291 | 6.253| 3.080 | 3.279 |
| NAFDU  | 3.093 | 6.081| 3.315 | 2.879 |
| WMF    | 3.060 | 6.429| 3.676 | 2.891 |
| EWMF\(^1\) | 2.896 | 6.042| 3.208 | 2.863 |
| EWMF\(^2\) | 2.788 | 5.759| 3.094 | 2.604 |

| Method | Teddy | Aloe | Midd1 | Baby2 |
|--------|-------|------|-------|-------|
| JBU    | 9.327 | 11.029| 11.404| 5.039 |
| NAFDU  | 9.307 | 10.322| 12.093| 4.671 |
| WMF    | 7.915 | 8.992 | 10.454| 3.724 |
| EWMF\(^1\) | 7.728 | 8.417| 10.323| 3.487 |
| EWMF\(^2\) | 7.657 | 7.423| 7.892 | 3.401 |

### TABLE I

**Quantitative comparison on the data from the Middlebury dataset in terms of RMSE. The upsampling factor chosen is 4.**

| Method | Teddy | Aloe | Midd1 | Baby2 |
|--------|-------|------|-------|-------|
| JBU    | 3.291 | 6.253| 3.080 | 3.279 |
| NAFDU  | 3.093 | 6.081| 3.315 | 2.879 |
| WMF    | 3.060 | 6.429| 3.676 | 2.891 |
| EWMF\(^1\) | 2.896 | 6.042| 3.208 | 2.863 |
| EWMF\(^2\) | 2.788 | 5.759| 3.094 | 2.604 |

### TABLE II

**Performance comparisons in terms of BP\% (for all pixels).**

| Method | Teddy | Aloe | Midd1 | Baby2 |
|--------|-------|------|-------|-------|
| JBU    | 9.327 | 11.029| 11.404| 5.039 |
| NAFDU  | 9.307 | 10.322| 12.093| 4.671 |
| WMF    | 7.915 | 8.992 | 10.454| 3.724 |
| EWMF\(^1\) | 7.728 | 8.417| 10.323| 3.487 |
| EWMF\(^2\) | 7.657 | 7.423| 7.892 | 3.401 |
In this paper, we first tease out the relationship between weighted mode filter and joint bilateral filter. WMF has the joint bilateral kernel and provides an optimal solution with the help of joint histogram. For the residual issues in original WMF, we have presented the extended WMF for LR depth map super-resolution. EWMF extends previous weighted mode filter with the combination of NAF. The mechanism of noise-aware is grafted onto the original WMF. Different from conventional fixed support window based methods, we design an adaptive refined window inspired by the cross-based multipoint filter. The extension order in RASW is determined based on a custom variance measures how far the pixels in the window boundary are spread out. The EWMF with RASW is more noise-aware than original WMF. Experiment results validate the effectiveness of proposed method.

### V. CONCLUSIONS

### REFERENCES

[1] J. Diebel and S. Thrun, “An application of markov random fields to range sensing,” in NIPS, 2005, pp. 291-298.

[2] W. Freeman and C. Liu, “Markov random fields for super-resolution and texture synthesis,” Advances in Markov Random Fields for Vision and Image Processing, vol. 1, 2011.

[3] S. Zheng, P. An, Y. Zuo, X. Zou, and J. Wang, “Depth Map Upsampling Using Segmentation and Edge Information,” in Image and Graphics, ed: Springer, 2015, pp. 116-126.

[4] W. Liu, S. Ia, P. Li, X. Chen, J. Yang, and Q. Wu, “An MRF-based depth upsampling: upsample the depth map with its own property,” Signal Processing Letters, IEEE, vol. 22, pp. 1708-1712, 2015.

[5] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, “High quality depth map upsampling for 3D-TOF cameras,” in Computer Vision (ICCV), 2011 IEEE International Conference on, 2011, pp. 1623-1630.

[6] D. Ferstl, C. Reinhacker, R. Ranftl, M. Rüther, and H. Bischof, “Image guided depth upsampling using anisotropic total generalized variation,” in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 993-1000.

[7] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, “Joint bilateral upsampling,” ACM Transactions on Graphics (TOG), vol. 26, p. 96, 2007.

[8] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in Computer Vision, 1998. Sixth International Conference on, 1998, pp. 839-846.

[9] M.-Y. Liu, O. Tuzel, and Y. Taguchi, “Joint geodesic upsampling of depth images,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 169-176.

[10] K. He, J. Sun, and X. Tang, “Guided image filtering,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 35, pp. 1397-1409, 2013.

[11] K.-L. Hua, K.-H. Lo, and Y.-C. Wang. “Extended Guided Filtering for Depth Map Upsampling,” IEEE MultiMedia 1:1-1.

[12] D. Min, J. Lu, and M. N. Do, “Depth video enhancement based on weighted mode filtering,” Image Processing, IEEE Transactions on, vol. 21, pp. 1176-1190, 2012.

[13] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, “A noise-aware filter for real-time depth upsampling,” in Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SF2 2008, 2008.

[14] J. Lu, K. Shi, D. Min, L. Lin, and M. N. Do, “Cross-based local multipoint filtering,” in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 430-437.

[15] “Middlebury stereo,” http://vision.middlebury.edu/stereo/data/.

---

**TABLE III**

**PERFORMANCE COMPARISONS IN TERMS OF RMSE WITH ADDITIONAL NOISE.**

| Method | Teddy | Aloe | Midd1 | Baby2 |
|--------|-------|------|-------|-------|
| JBU    | 8.598 | 10.991 | 6.853 | 3.576 |
| NAFDU  | 9.076 | 11.006 | 7.181 | 3.957 |
| WMF    | 7.385 | 7.687 | 7.953 | 3.576 |
| EWMF²  | 5.251 | 5.825 | 4.161 | 5.474 |

**TABLE IV**

**PERFORMANCE COMPARISONS IN TERMS OF BP% WITH ADDITIONAL NOISE.**

| Method | Teddy | Aloe | Midd1 | Baby2 |
|--------|-------|------|-------|-------|
| JBU    | 11.692 | 13.810 | 14.968 | 6.814 |
| NAFDU  | 11.540 | 13.193 | 13.330 | 6.853 |
| WMF    | 8.151 | 8.154 | 8.681 | 4.057 |
| EWMF²  | 7.835 | 7.687 | 7.953 | 3.576 |
| EWMF²  | 6.999 | 6.812 | 7.588 | 3.220 |

---

![Fig. 2](image-url) The super-resolution (4x) results of Teddy (top) and Aloe (bottom) with two regions of interest (ROI). (a) The original HR color image, (b) the ground truth depth map, the corresponding results of (c) JBU, (d) NAFDU, (e) original WMF, proposed EWMF utilizing (f) fixed square window and (g) RASW.