IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 15, 2018, accepted September 23, 2018, date of publication October 11, 2018,
date of current version January 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2873567

Efficient Confidence-Based Hierarchical Stereo Disparity Upsampling for Noisy Inputs

XIANG-BING MENG[1,2,3], MEI ZHANG[2,3], ZHAO-XING ZHANG[2], RONG WANG[1,2,3], ZHENG GENG[2], (Senior Member, IEEE), AND FEI-YUE WANG[2,3], (Fellow, IEEE)

[1]University of Chinese Academy of Sciences, Beijing 100190, China
[2]State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[3]Parallel Optics Technology Innovation Center, Qingdao Academy of Intelligent Industries, Qingdao 266000, China

Corresponding author: Mei Zhang (mei.zhang@ia.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61605240, Grant 61702519, and Grant 61309024, in part by the Key Program of the National Natural Science Foundation of China under Grant 61533019, Grant 71232006, and Grant 61233001, and in part by the Hunan Provincial Key Research and Development Program under Grant 2018SK2129.

ABSTRACT Disparity upsampling techniques aim to restore high-resolution disparity maps from low-resolution disparity inputs. These inputs must be of high quality and are often obtained via complicated passive or active 3-D reconstruction methods. Each pixel in the input disparity maps guides the disparity assignment in the upsampling process. The quality of the upsampled results will decreases if the initial disparity inputs are noisy, as the upsampled results are closely related to the initial inputs. We herein propose a hierarchical confidence-based upsampling framework that can be used to obtain relatively high quality upsampled results even under the noisy inputs. Specifically designed confidence measuring schemes are employed in our upsampling process, allowing the disparity assignment of only high-confidence pixels. For an effective depth quality evaluation, we present a novel classification of the confidence according to depth- and texture-related information and develop a confidence examination method with improved precision by combining multiple depth confidence evaluation methods. Our hierarchical pipeline contains three steps: confidence-based upsampling, confidence-based fine-tuning, and confidence-based optimization. The upsampling combines multichannel information. Fine-tuning is carried out using the stereo texture information. Optimization is conducted utilizing the Markov random field method. All these proposed methods work together to suppress the low-confidence pixels and propagate the high-confidence pixels in the upsampling process. The cumulative error distribution is further analyzed, revealing the effectiveness of our confidence evaluation. Extensive comparison experiments are also performed using both the ground truth and stereo matching disparity maps as inputs to demonstrate the advantage of our framework over state-of-the-art upsampling methods.

INDEX TERMS Disparity upsampling, confidence evaluation, noise, hierarchical structure, multichannel upsampling.

I. INTRODUCTION

Real-time high-resolution and high-quality 3D reconstruction has been one of the most significant issues in the field of computer vision. It is widely applied in 3D display technology [1], [2], augmented reality (AR) [3]–[5] and simultaneous localization and mapping (SLAM) [6].

The existing 3D reconstruction techniques are generally classified into passive and active methods. For the passive methods, depth is estimated via correspondence from different images. The stereo matching method, one of the most common passive methods, retrieves depth information from two rectified camera images [7], [8]. It can be further divided into local and global methods. Local methods are faster but usually produce relatively low-quality results. These low-quality results usually occur for poorly textured (textureless, repeatedly textured or occluded) areas or regions that do not satisfy the fronto-parallel assumption in the matching window (such as a depth edge). In contrast, global methods carry out the matching process by solving a global energy function. Their results are better than the local ones but at the cost of additional computation time. For active methods such as the structured light scanner [9] and

time-of-flight (TOF) [10], the depth information can be directly acquired using the active information. In the former, surface information is retrieved via the deformation of preknown projected image patterns, while in the latter, one obtains depth information via TOF-based range estimation. Unlike the passive 3D reconstruction techniques, these methods are intrinsically robust against the poorly textured areas of a reconstructed object and can produce accurate depth maps. Apart from the noise, however, the limited resolution yielded by these passive techniques is the major disadvantage preventing their utilization in areas where high-resolution depth maps are required.

The trade-off between high resolution and high quality to obtain depth maps in real time has led to the development of various depth upsampling strategies [11]–[13]. One prevailing strategy is based on the Markov random field (MRF) [14], [15]. This strategy uses the MRF framework and solves the upsampling problem either via a graph cut or gradient descent. Jang and Ho [16] combined stereo texture information with depth information obtained via TOF for upsampling. Li et al. [17] applied a hierarchical structure to gradually upsample the depth to the target scale. Their method yields better upsampled results and has been proven to be robust with respect to noise. Furthermore, the hierarchical structure is efficient because of its pyramid arrangement. However, according to our experimental results, these state-of-the-art methods fail to perform well when the inputs are of low quality. In Fig. 1, the left part illustrates the direct upsampling process, where the noise from the initial disparity maps induces more disparity-biased points in the high-resolution maps during the direct upsampling process. The right part shows that the noise information is suppressed in the process of upsampling, while the correct information seems to be propagated effectively. The above comparison illustrates that effective confidence evaluation and low-quality points restoration are essential to achieve depth denoising in the upsampling process.

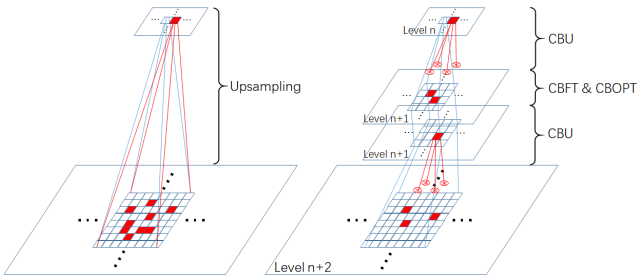


FIGURE 1. Framework comparison between traditional upsampling method and our proposed method. The left picture shows the traditional depth upsampling procedure without confidence evaluation, while the right one is based on our proposed framework. CBU, CBFT and CBOPT are the main modules used in our framework, representing confidence-based upsampling, confidence-based fine-tuning and confidence-based optimization, respectively. Red pixels represent low-quality pixels, and the circled "x" indicates the discontinuation of these low-quality pixels.

In this work, a confidence-based hierarchical upsampling framework that can reduce noise originating from the inputs via a carefully designed pipeline is proposed. First, a confidence evaluation is designed as the core module. Many state-of-the-art confidence evaluation strategies [18] and [19] are based on the stereo matching cost. However, none of these confidence evaluation strategies are effective for poorly textured areas. Therefore, coarse-to-fine confidence evaluation is proposed to solve this issue, where confidence is classified into two classes: well-describable confidence (WDC) measuring the well textured disparity points and non-describable confidence (NDC) measuring the poorly textured disparity points. Then, confidence is applied to guide the entire pipeline of our hierarchical structure to propagate those points with high confidence while restoring those points with low confidence. Second, based on the computational efficiency and complementarity in processing various texture conditions (WDC and NDC), multiple upsampling methods are proposed and combined according to the confidence guidance. Third, as described above, the depth/disparity[1] of passive reconstruction approaches such as the stereo matching methods is usually noisy compared with that of the active approaches. Therefore, disparity fine-tuning is proposed in an attempt to correct those points with low confidence. Finally, a confidence-based MRF strategy is applied to refine the restored data.

In summary, our contributions in this article include the following:

1) We propose a novel confidence-based hierarchical disparity upsampling and enhancement framework, in which confidence plays a vital role. The experimental results verify the high robustness and accuracy of our system under noisy inputs.
2) We design a coarse-to-fine confidence evaluation method, which includes a more precise confidence classification and evaluation compared with other methods.
3) We design a confidence-based upsampling method using multichannel information. Fine-tuning is applied to restore low-confidence pixels. The noisy information is greatly suppressed and corrected after carrying out these processes.
4) We realize all the key algorithms in our framework using a GPU and achieve near-real-time disparity upsampling.

II. RELATED WORKS

The state-of-the-art disparity upsampling methods and confidence-based multiscale strategy are presented in the following parts.

A. CURRENT UPSAMPLING METHODS

According to the core algorithm, upsampling methods can be categorized into one of the following four classes.

[1] In general, disparity being inversely proportional to depth is defined as the difference along the horizontal coordinate between the corresponding points in the rectified stereo maps.

1) INTERPOLATION-BASED METHODS

Early simple upsampling methods, such as bicubic interpolation (BCI) [20], use curve fitting for upsampling and purely rely on depth maps. Fukushima *et al.* [21] provided better upsampled results when they further used information related to patchwise similarities within the input image itself. Moreover, they combined local texture information with local depth information via linear interpolation for depth upsampling. He *et al.* [22] used the distribution of texture images as guidance to generate better upsampled results via interpolation.

2) WEIGHTED-FILTER-BASED METHODS

Among those upsampling methods that use information related to depth and the corresponding texture image, joint bilateral upsampling (JBU) [23] leverages both space distance and color distance during depth assignment. Yang *et al.* [24] proposed 3D-JBU, using cost volume to optimize and restore detailed depth information. The noise-aware filter for depth upsampling (NAFDU) [25] extended JBU by using both high-scale and low-scale texture information to preserve the beneficial properties of bilateral upsampling in those areas where standard bilateral upsampling performed well and to prevent artifacts in those areas where standard bilateral upsampling was likely to cause the texture copy problem. Xiao *et al.* [26] applied JBU to the upsampling image pyramid with ameliorated kernel shapes. Comparatively, joint geodesic upsampling (JGU) [27] replaces the Euclidean distance with the geodesic distance as the guide for upsampling, demonstrating better results in depth discontinuity regions with similar colors. Weighted mode filtering (WMF) [28] adopts the weighted voting method instead of the weighted average method to obtain the final result, achieving excellent performance in terms of edge preservation.

3) MRF-BASED METHODS

Based on the common MRF-based methods, Liu *et al.* [14] developed the iterative MRF-based upsampling (MBU) method and added the self-generated depth cues from previous calculations to the depth propagation weights, which has been proven to be more robust against depth discontinuities. Jung and Ho [15] took the difference between the depth from BCI and color-guided upsampling to evaluate the depth quality and optimized the MRF-based objective function by a graph cut. Li *et al.* [17] proposed a fast MRF-based upsampling (FGI) with a hierarchical structure. At each level, the fusion of two-channel information and a simple depth quantity evaluation such as [15] is applied to address the noisy source and provide a precise result.

4) LEARNING-BASED METHODS

With the rapid development of learning-based methods [29]–[31], machine-learning-based upsampling has attracted more and more attention from researchers. Zhang and Cham [32] proposed a learning-based framework in the discrete cosine transform domain for face image upsampling. The methods proposed by Yang *et al.* Yang, Timofte *et al.* [34] and Kwon *et al.* [35] adopt sparse representation or dictionaries trained from existing data to perform depth upsampling and restoration. Dong *et al.* [36] proposed an end-to-end convolutional neural network (CNN) for image restoration. Hui *et al.* [37] introduced a CNN-like framework to address the problem of depth upsampling using a multiscale guided convolutional neural network (MSG-Net) associated with image information.

B. CONFIDENCE-BASED MULTISCALE STRATEGY

As mentioned above, FGI also uses the hierarchical structure strategy for fast upsampling. However, noise information will not be suppressed effectively, since the confidence information on the quality of the disparity is not used as the guidance of whole procedure and the method on confidence measurement is proved to be of low precision [38] and no fine-tune strategy is proposed to process the detail information of the disparity map.

Furthermore, the confidence-based multiscale strategy has also been used in reconstruction tasks involving stereo matching methods [19], [39]–[41], with excellent results being achieved according to these contributions. However, large differences in the tasks, processing method and inputs exist between the mentioned works and the proposed upsampling method: those works reconstruct 3D information using texture images through optimizing the cost volume [19], while upsampling aims at obtaining higher-resolution disparity maps by processing the existing lower-resolution ones.

As described above, a majority of the upsampling methods use a texture image to guide the depth denosing process, except for [15], [17], where the confidence is calculated independently from single-view information and used for denoising. Different from the mentioned upsampling methods, the matching information from stereo geometry and multiple confidence evaluation strategies are used to obtain the effective confidence evaluation result in this work. Furthermore, unlike the work of Bastian *et al.* [17], our hierarchical structure with the effective confidence evaluation can not only carry out denosing using single-view information but also correct low-quality points by fine-tuning via stereo matching.

III. PROPOSED METHOD

As shown in Fig. 1, traditional upsampling methods directly obtain a high-resolution disparity map, which is possibly of low quality because of the existing noise in the input low-resolution map. Thus, we propose a confidence-based hierarchical framework for disparity upsampling. It is based on both the hierarchical upsampling structure and a quality evaluation of disparity information. In the following subsections, we introduce our proposed system in detail.

Readers should note that the variables in a function have values corresponding to the same level of the proposed framework, unless otherwise noted.

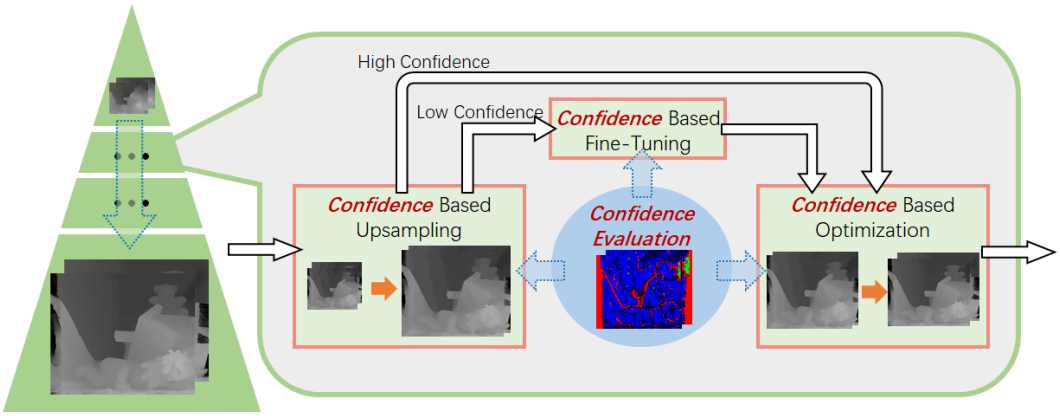


FIGURE 2. The overview of our system. Our system adopts the coarse-to-fine pyramid structure. Each level follows the same procedure, which consists of three modules: upsampling, fine-tuning and optimization, all of which are guided by the evaluated confidence. First, disparity maps from the previous level are upsampled using the confidence-based upsampling module, which fuses multichannel information from different upsampling methods. Then, the low-confidence points in the upsampled maps are updated via the confidence-based fine-tuning module. Finally, confidence-based optimization is executed to carry out disparity denoising and smoothing.

A. SYSTEM OVERVIEW

An overview of our proposed system is shown in Fig. 2. An iterative hierarchical strategy is adopted in the proposed framework. Each layer of the framework follows the same procedure, which consists of 3 modules: confidence-based upsampling (CBU), confidence-based fine-tuning (CBFT) and confidence-based optimization (CBOPT). These are guided by the disparity confidence obtained from the confidence evaluation. Next, we briefly introduce our system.

At each level, the initial sources from the last level, including stereo disparity maps with their confidence and texture maps, are considered as the inputs of the CBU module, which performs confidence-based disparity upsampling based on different upsampling methods. Next, low-confidence pixels in the upsampled disparity maps are updated by the CBFT module via a simple stereo matching method. Finally, the disparity maps are refined using the CBOPT module.

The advantages of our proposed pipeline are summarized from two aspects: first, high-quality disparity information is maintained at a high level via the iterative hierarchical strategy; second, low-quality disparity information is suppressed and corrected during the propagation process.

B. CONFIDENCE EVALUATION

High precision is essential to achieve properly designed confidence evaluation, as it takes the core position in this framework according to the aforementioned description. Multiple confidence measures based on the cost volume (CV) of stereo matching [18] are combined to achieve a satisfactory result with high precision. The effectiveness of this strategy was proven by Sun *et al.* [19]. However, a serious reduction in the precision of these CV-based methods will occur because of the poorly textured areas in depth maps and the limitations of the upsampling task, as discussed in the following sections. Coarse-to-fine confidence evaluation is proposed to overcome these issues:

$$C_{final} = C_{corr}\,(\lambda C_{ndc} + (1-\lambda)C_{wdc}), \tag{1}$$

where confidence is classified into two classes: well-describable confidence(WDC) defined in Section III-B3 and non-describable confidence (NDC) defined in Section III-B2. C_{corr} represents the confidence obtained from the simple confidence evaluation discussed in Section III-B1, C_{ndc} is the confidence generated by the NDC points classification and evaluation discussed in Section III-B2, C_{wdc} is the improved confidence at WDC points from the WDC points evaluation discussed in Section III-B3, and λ is a binary classification selector between NDC and WDC. As shown in (1) and Fig. 3 (a), C_{corr} is used for the initial coarse examination, which utilizes the correspondence information to measure the disparity rapidly. Then, the disparity is classified by λ into WDC and NDC. Finally, C_{ndc} or C_{wdc} is performed to achieve fine measurement. After this process, the uncertain confidence of NDC points is isolated, and the fine measurement of WDC points is obtained; thus, confidence-based precise upsampling is achieved. The experimental results illustrate that this strategy improves the final result. Next, the confidence evaluation is presented in detail.

1) SIMPLE CONFIDENCE EVALUATION

This module, which carries out the coarse evaluation, aims to isolate the obvious low-quality points from the disparity map. The correspondences between the left and right maps in terms of the disparity (LRDC) and intensity of the color (LRCC) are used to evaluate the confidence of a disparity directly:

$$C_{corr} = \begin{cases} 1, & (C_{lrdc} > \tau_{lrdc}) \\ & \wedge (C_{lrcc} > \tau_{lrcc}), \\ \tau_{C_{min}}, & otherwise \end{cases} \tag{2}$$

$$C_{lrdc,x} = -\left|D_{L,x} - D_{R,\tilde{x}}\right| \tag{3}$$

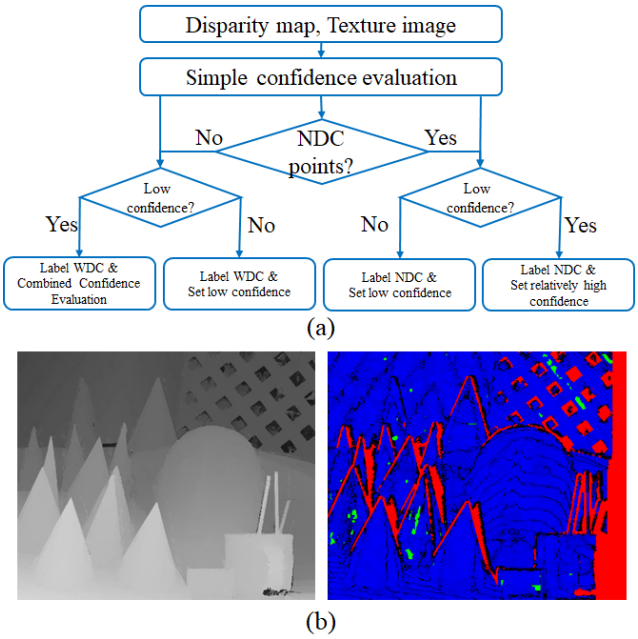


FIGURE 3. Overview of the confidence evaluation. (a) shows the flowchart of our confidence evaluation. In our measures, disparity pixels are divided into either well-describable-confidence or non-describable-confidence. (b) shows one example of a confidence map (right) and its disparity map (left) under our measures. Blue pixels in the confidence map represent disparity pixels whose confidence can be described. Brighter means that the disparity pixel presents more confidence. Red and green pixels, which are generally located at disparity edges and in textureless areas, respectively, belong to non-describable-confidence points.

$$C_{lrcc,x} = 1 - \frac{1}{Z_{I_{max}}} \left\| I_{L,x} - I_{R,\tilde{x}} \right\|_1, \tag{4}$$

where $Z_{I_{max}}$ is the normalization term, D_L, D_R and I_L, I_R are the left and right disparity or color map, respectively, and $\tilde{x}$ is the point corresponding to x via disparity D_x. τ_{lrcc} and τ_{lrdc} are thresholds for the confidence, with $\tau_{C_{min}}$ being a low confidence, as shown in Fig. 4. Obvious bad disparity points can be detected effectively, according to [18], [38]. This method is utilized at the very start in view of the complexity and precision of the framework, as it is simple and effective for detecting bad disparities without increasing the cost volume. After this simple evaluation, disparities with low confidence no longer need to be evaluated.

2) NDC POINTS CLASSIFICATION AND EVALUATION

Cost curves representing different texture conditions according to [42] are presented in Fig. 5. According to [18], the CV-based method is based on the hypothesis that the disparity to be evaluated has the minimal cost on the curve, such as point P_a in Fig. 5(a). As shown in Fig. 5(b)-(d), the poor texture characterized by the stereo matching cost will result in several ambiguous cost curves presenting either many minimal values (Fig. 5(b), (d)) or incorrect minimal values (Fig. 5(c)). Therefore, CV-based confidence evaluation methods will yield false judgments and low precision. In our proposed method, points in poorly textured areas are classified as NDC type and isolated from the confidence map.

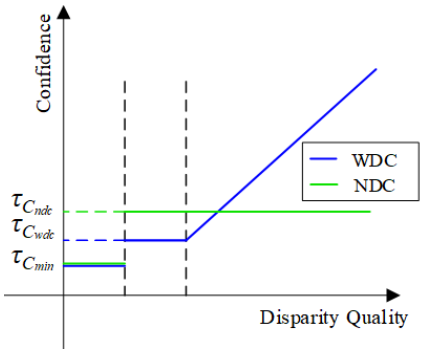


FIGURE 4. The ideal mapping between the proposed confidence and the real disparity quality. The blue line represents the confidence of WDC. The green line represents the confidence of NDC. The left vertical dashed line represents the threshold of the simple confidence evaluation; the right one represents the threshold of the combined confidence evaluation.

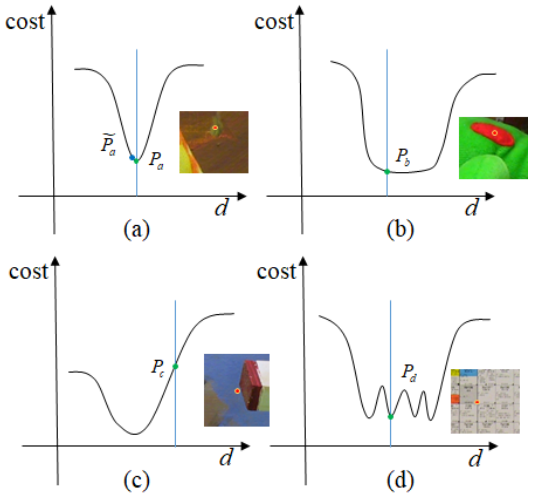


FIGURE 5. The cost curves for points in areas under different texture conditions, including a well-conditioned area (a), textureless area (b), disparity edge area (c) and repeatedly textured area (d). The horizontal axis represents the disparity of a point; the vertical axis represents the matching cost. The true disparity of a point in the images on the right is indicated by a green point on the curves.

The points in the repeatedly textured areas shown in Fig. 5(d) are not classified into NDC points, as the curve also locally has only one minimum. Fortunately, these points can be measured precisely using the simple confidence evaluation module. Therefore, textureless points (Fig. 5(b)), disparity edge points (Fig. 5(c)), including occluded edge points and non-occluded edge points, are detected as NDC points. More details on detecting NDC points are given below.

a: Textureless Points

Textureless points exist in areas with a low color gradient. The Sobel operator is used to calculate the gradient of gray images; points with gradients less than 6/255 are shown in Fig. 3(b) (the green points).

b: Occluded Edge Points

Occlusion is caused by a change in perspective and often occurs at an image border or depth edge, as shown in Fig. 3 (the red points on the right edge and the margin of the right part). Image border occlusion is detected using forward or backward warps. Occluded edge points are detected using the method presented in [43].

c: Non-Occluded Edge Points

As shown in Fig. 3(b), the red points at the left edge represent non-occluded edge points. First, the Canny edge detector [44] is applied to find the edge of a disparity. Then, isolated noise points are filtered out. Finally, the edges are expanded by a margin, the width of which is the same as the window size defined in (6).

Finally, the confidence of NDC points, C_{ndc}, is defined as:

$$\begin{cases} \tau_{C_{ndc}}, & \text{if } x_c \in \text{NDC} \quad (5a) \\ \frac{1}{2}(\tau_{C_{min}} + \tau_{C_{wdc}}), & \text{if } x_c \in \text{WDC} \quad (5b) \end{cases}$$

where x_c is the target point and x is the neighbor of x_c. $\tau_{C_{ndc}}$, $\tau_{C_{wdc}}$ and $\tau_{C_{min}}$ are the thresholds shown in Fig. 4. This strategy is proposed to prevent introducing noise to WDC points from NDC points for $C_{ndc} < \tau_{C_{wdc}}$ (5b) and to safely filter out the noise at the NDC points for $C_{ndc} > \tau_{C_{wdc}}$ (5a).

3) WDC POINTS EVALUATION

The WDC points are further evaluated using the CV-based method when they have passed the examination of the module described in Section III-B1, as a simple confidence evaluation cannot provide finer evaluation for the relatively good WDC points. Zero-mean normalized cross-correlation (ZNCC) is adopted to estimate the cost; it is defined as:

$$\begin{aligned} \tilde{c}_{d,x_c} &= \frac{N \cdot S_{IJ} - S_I \cdot S_J}{\sqrt{(N \cdot S_{II} - S_I^2)(N \cdot S_{JJ} - S_J^2)}}, \\ c_{d,x_c} &= 1 - T(\tilde{c}_{d,x_c})\tilde{c}_{d,x_c}, \end{aligned} \quad (6)$$

where $T(p)$ is set to 1 if $p > 0$ and set to 0 otherwise. N is the number of pixels in the window centered at x_c, and $S_I = \sum_{x \in N_{x_c}} I_{L,x}$, $S_J = \sum_{x \in N_{x_c}} I_{R,x-d}$, $S_{II} = \sum_{x \in N_{x_c}} I_{L,x}^2$, $S_{JJ} = \sum_{x \in N_{x_c}} I_{R,x-d}^2$, and $S_{IJ} = \sum_{x \in N_{x_c}} I_{L,x} I_{R,x-d}$.

However, the above process is not sufficient for the upsampling task to support the hypothesis mentioned in Section III-B2. The specific disparity obtained via the upsampling process will not always take the minimal cost, according to the point $\tilde{P}_a$ shown in Fig. 5(a), and doubtless unstable evaluation results will be yielded. Thus, additional strategies are proposed, the details regarding which are as follows.

a: Matching Score Measure (MSM)

The matching cost, as the simplest measure [18], is not dependent on the hypothesis of minimal cost. The confidence is defined as:

$$\tilde{c}_{msm} = max(\tau_{msm} - c_d, 0), \quad (7)$$

where τ_{msm} represents the truncated value of the maximum MSM.

b: Curvature (CUR)

CUR, defined in [18], is equivalent to the sum of the absolute values of the left and right cost gradients at the point with minimal cost. However, the evaluation result will be incorrect when the cost of the current disparity deviates the minimal one slightly. For example, the confidence of $\tilde{P}_a$ in Fig. 5(a) will be very small or negative. Therefore, the improvement defined in (8) is proposed to fit this task:

$$\begin{aligned} \tilde{c}_{cur} &= max(c_{cur}, 2min(|c_{d-1} - c_d|, |c_{d+1} - c_d|)), \\ \tilde{c}_{cur} &= min(\tilde{c}_{cur}, \tau_{cur}), \end{aligned} \quad (8)$$

where c_{cur} is the original CUR, defined as $c_{d-1} + c_{d+1} - 2c_d$, and τ_{cur} represents the truncated value of the maximum CUR.

c: Naive Peak Ratio (PKRN)

The PKRN, defined in [18], is used to evaluate the margin between points with the first and second minimal cost. Similarly, failure of the hypothesis will lead to the wrong confidence. Thus, the improved PKRN is defined as shown in (9):

$$\tilde{c}_{pkrn} = min(f(c_{d-1}/c_d), f(c_{d+1}/c_d), \tau_{pkrn}), \quad (9)$$

where $f(p) = T(p-1)p + (1 - T(p-1))/p$, $c_{pkrn} = c_2/c_d$ is the original PKRN, and c_2 is the second minimal cost. τ_{pkrn} represents the truncated value of the maximum PKRN.

d: Left-Right Difference (LRD)

The LRD is the best measure for carrying out the disparity evaluation [18]. For the same reason mentioned regarding both the CUR and PKRN, the LRD also needs to be improved:

$$\tilde{c}_{lrd} = min(\frac{min(|c_{d-1} - c_d|, |c_{d+1} - c_d|)}{|c_d - \tilde{c}_{\tilde{d},\tilde{x}}|}, \tau_{lrd}), \quad (10)$$

where $\tilde{c}_{\tilde{d},\tilde{x}}$ is the cost of correspondence pixel $\tilde{x}$ in the target image at the current target disparity. τ_{lrd} represents the truncated value of the maximum LRD.

Then, all the evaluations are combined (see (11)), with the minimal value being set to the final calculated result to limit the confidence to within a reasonable range.

$$\begin{aligned} C_{wdc} &= \frac{\tilde{c}_{msm} + \tilde{c}_{cur}}{\tilde{c}_{msm} + \tilde{c}_{cur} + \tau_{msm} + \tau_{cur}} \\ &\quad \cdot \frac{\tilde{c}_{pkrn} * \tau_{pkrn}}{\tilde{c}_{pkrn} * \tau_{pkrn}} \cdot \frac{\tilde{c}_{lrd} * \tau_{lrd}}{\tilde{c}_{lrd} * \tau_{lrd}} \\ C_{wdc} &= max(\tilde{C}_{wdc}, \tau_{C_{wdc}}), \end{aligned} \quad (11)$$

where $\tau_{C_{wdc}}$, shown in Fig. 4, is the minimum WDC, which is greater than $\tau_{C_{min}}$, as the points have passed the evaluation of Section III-B1.

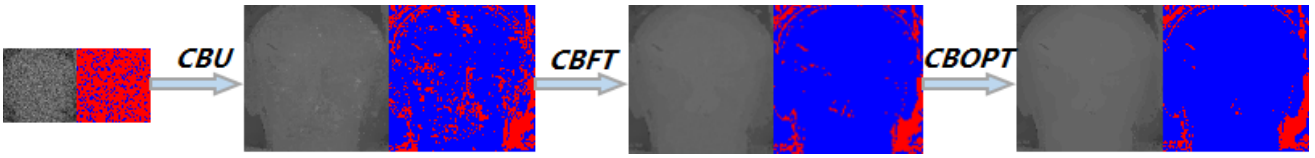


FIGURE 6. The upsampled results of our modules in the framework from one layer in the structure. This experiment uses disparity maps with Gaussian noise as the original inputs. The first set of images represents the results of the disparity maps. The second set of images presents the error maps, where the blue and red points are correct and incorrect disparity points, respectively, with a threshold of 4 pixels at the highest resolution.

C. CONFIDENCE-BASED UPSAMPLING

To achieve robust and highly accurate results, three complementary schemes are combined in this module: bicubic interpolation upsampling (BIU), confidence-based joint cubic upsampling (CJCU) and weighted voting upsampling (WVU). This weighted combination strategy is inspired by the work proposed by Zhang *at el.* [45], which composites multiple images with different exposure according to the quality assessment. BIU uses interpolation to retain the original distribution of the disparity information. CJCU is performed under the assumption that the noise of the disparity obeys a Gaussian distribution. It takes advantage of the texture, space and disparity information and uses the confidence as a guide to suppress the noise. WVU does not take the assumption of CJCU and instead uses the weighted voting result as a guide to suppress the noise. Thus, the original information is preserved by BIU, and the noise is suppressed by CJCU and WVU. More details are given below.

1) BICUBIC INTERPOLATION UPSAMPLING

BIU is applied to directly obtain the upsampling disparity from the low-level disparity maps. The original information from the lower level is well preserved, as it is based only on the interpolation.

2) CONFIDENCE-BASED JOINT CUBIC UPSAMPLING

The original JBU method [23] adopts both spatial and color distances as weights in the upsampling process. Since JBU is relatively sensitive to noise, we further add the disparity distance and disparity confidence as the weighting terms to suppress the influence from inappropriate disparities on the object point. The overall weighting term is given in (12):

$$D_{cjcu}^{i+1} = \frac{1}{Z_{cjcu}} \sum_{x^{i+1} \in N_{x_c}^{i+1}} w_d^i w_c^{i+1} w_s^{i+1} D_{x^i}^i, \tag{12}$$

where Z_{cjcu} is the normalized factor. The superscript i represents the ith layer in the framework. In addition, $x^i = x^{i+1}/S$, where S represents the upsampling scale between two layers in the framework. Additionally, the color-related weight $w_c = \exp\left(\|I_x - I_{x_c}\|/3\sigma_c^2\right)$, the space-related weight $w_s = \exp\left(\|x - x_c\|/\sigma_s^2\right)$, and both the confidence and the disparity-related weight $w_d = C_x \exp\left(C_{x_c}\|D_x - D_{x_c}\|/\sigma_d^2\right)$ are utilized to perform CJCU. The terms w_s, w_c and w_d in the following parts have the same formulas as those presented above.

3) WEIGHTED VOTING UPSAMPLING

This voting post-processing method is inspired by [46] and proposed to produce "confidence" for denoising, as the proposed confidence cannot be used to evaluate the NDC points. The guided weight information from the texture images is applied to build the histogram of the disparity and computed using (13):

$$h_{d_i} = \sum_{x \in N_{x_c}} T(|D_x - d_i| < l_{bin}) w_c w_s \tag{13}$$

$$l_{bin} = max((d_{max} - d_{min})/\tau_{h_{maxnum}} + 1, \tau_{l_{bin}}), \tag{14}$$

where d_i is the center disparity of the ith bin, the width of which is defined as l_{bin} according to (14). d_{max} and d_{min} are the maximal and minimal disparities in N_{x_c}, respectively. $\tau_{h_{maxnum}}$ represents the maximal number of bins in the histogram. $\tau_{l_{bin}}$ is the minimal width of each bin. Finally, the disparity is updated using (15), similar to (12), to perform denoising and smoothing in the WVU process:

$$D_{wvu}^{i+1} = \frac{1}{Z_{wvu}} \sum_{x^{i+1} \in N_{x_c}^{i+1}} T_d^i w_c^{i+1} w_s^{i+1} D_{x^i}^i, \tag{15}$$

where Z_{wvu} is the normalized factor, $T_d = T(|D_x - \tilde{D}_{x_c}| < l_{bin})$ represents the disparity guidance, and $\tilde{D}_{x_c}$ is the disparity with the highest weight in the histogram, defined as $\arg\max_{d_i \in [d_{min}, d_{max}]} h_{d_i}$.

4) COMBINED UPSAMPLING

After the initial upsampling, we obtain 3 upsampled maps: $D_{h,biu}$, $D_{h,cjcu}$ and $D_{h,wvu}$. Then, confidence evaluation is conducted to obtain their corresponding confidence maps. Considering the efficiency of the algorithm, the NDC points will not be detected for $D_{h,biu}$ and $D_{h,cjcu}$. The final disparity result can be obtained using (16):

$$D_x^{i+1} = \begin{cases} D_{wvu}^{i+1}, & \text{if } x \in \text{NDC} \\ D_{confMax}^{i+1}, & \text{otherwise,} \end{cases} \tag{16}$$

where $D_{confMax}$ is the disparity with the highest confidence among the 3 initial upsampled results. The disparity of an NDC point is assigned by WVU, since the weighted voting strategy can be regarded as the confidence for denoising, as discussed in Section III-C3. The winner-take-all strategy is applied to calculate the final disparity for the WDC points. Actually, the confidence-based weighted combination method is also tested; however, relatively worse results are obtained.

As shown in Fig. 6, most of the low-quality disparities are corrected using the CBU module. The remaining noise is removed by the other two modules.

D. CONFIDENCE-BASED FINE-TUNING

This module is introduced to effectively suppress the disparity error caused by CBU or existing in the original inputs. The NDC points do not undergo this process, as we cannot precisely evaluate their quality. Since the main purpose of our framework is to upsample and enhance the disparity rather than reconstruction, CBFT performs fine-tuning only around the initial disparity under the condition that no extra noise is added to the target maps.

1) MASK GENERATION FOR CBFT

The WDC points with confidence lower than τ_{cbft} are processed by the CBFT module. Considering the speed and quality, we set τ_{cbft} as $\tau_{C_{wdc}}$ at high levels (low resolution) and $\tau_{C_{min}}$ at low levels (high resolution) in our framework.

2) FINE-TUNING

This lightweight fine-tuning offers a high possibility to obtain the correct disparity using our strategy. Furthermore, it is computationally efficient, as only low-confidence points are processed by this module. First, based on the existing stereo disparity maps, the center of the disparity tuning range for every marked point is determined by disparities of all the other confident points. As defined in (17), a neighboring window is set for those points whose tuning range $[\tilde{d}_{x_c} - \tau_{range}, \tilde{d}_{x_c} + \tau_{range}]$ is determined by a weight consisting of the color distance, point distance and confidence.

$$\tilde{d}_{x_c} = \frac{1}{Z_d} \sum_{x \in N_{x_c}} w_c w_s T(C_x - \tau_{cbft}) D_x \tag{17}$$

For those points whose neighboring windows lack a sufficient number of high-confidence points, we set $\tilde{d}_{x_c} = (d_{x_c} + \max\{d_x | x \in N_{x_c}\})/2$. All the estimated tuning ranges are checked to ensure that they are reasonable for each level. Then, ZNCC, defined in (6), is used to compute the cost within the determined tuning ranges $d \in [\tilde{d}_{x_c} - \tau_{range}, \tilde{d}_{x_c} + \tau_{range}]$. Finally, the best disparity with the minimal cost is selected as the final result.

3) CONFIDENCE-GUIDED ROLL-BACK

After applying the CBFT module, the stereo confidence and disparity maps are updated again only for the marked points. As the fine-tuning method is simple, some points under the bad texture condition may not be corrected, and additional noise will be introduced. Thus, the disparity and confidence values prior to CBFT will be reassigned if the confidence of the point decreases (we call this the ''roll-back'' operation).

As shown in Fig. 6, most of the noise points (in WDC areas) are removed via the CBFT module. The final optimization module described next performs further filtering.

E. CONFIDENCE-BASED OPTIMIZATION

Finally, optimization is executed based on the initial disparity data obtained after performing CBU and CBFT. This module aims at removing the noise induced by the CBFT module and refining the initial disparity. We combine the confidence with another state-of-the-art MBU method to propagate the disparity based on the confidence. The energy function is presented as:

$$E = \sum_{x_c \in \Omega} C^0_{x_c} (D_{x_c} - D^0_{x_c})^2 + \lambda \sum_{x_c \in \Omega} \sum_{x \in N_{x_c}} C_x w_{x,x_c} \psi_C, \tag{18}$$

where Ω represents the entire map area, D_x and C_x represent the disparity and confidence of point x, respectively, the superscript represents the data in the *nth* iteration, ψ_C is defined as $\frac{2{\sigma_d}^2}{C_{x_c}+\epsilon_\psi}\left(1 - \exp\left(-\frac{C_{x_c}\|D_{x_c} - D_x\|_2^2}{2{\sigma_d}^2}\right)\right)$, and w_{x,x_c} is defined as $w_c w_s$. We obtain the final solution to (18) using the iterative model, referring to [14], as shown in (19):

$$D^{n+1}_{x_c} = \frac{1}{Z_D}\left(C^0_{x_c} D^0_{x_c} + \lambda \sum_{x \in N_{x_c}} \tilde{w}_{x,x_c} D^n_x\right), \tag{19}$$

where $\tilde{w}_{x,x_c}$ is defined as $w_d w_c w_s$, which is similar to (12). Z_D is the normalized value, and λ is the parameter used to balance the relative weight between the initial information and optimization information.

As shown in Fig. 6, most of the points with NSP-like noise are removed using the CBOPT module.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the proposed framework is evaluated on datasets from Middlebury online [47]–[49]. Five different sources were used as input disparity maps: data with no noise (GT), data with salt and pepper noise (NSP), data with Gaussian noise (NGS), data obtained via the AdCensus stereo matching method [46] (AdC) and data obtained via the MeshStereo stereo matching method [50] (MS). We compared our proposed framework with other upsampling methods, the source codes of which are available online. The experiments were performed on a PC with an Intel Xeon CPU (2.60 GHz) and an NVIDIA GeForce GTX TITAN X GPU. Our framework was implemented in a MATLAB, C++, and CUDA hybrid programming language. The error ratio, which was measured at the highest scale/resolution, was applied to evaluate all the experimental results. Considering the error ratio of the initial disparity maps and without loss of generality (all the methods were measured against the same standard), the 4-disparity bias was used as the threshold of the error ratio. Moreover, the initial disparity maps in all the experiments were upsampled to their highest resolution at all upsampling rates. Downsampling was executed to obtain the lower disparity and color maps for the GT dataset.

A. PARAMETER SETTINGS

The following parameters correspond to a 16× upsampling with a 5-layer framework. The parameters for 8× and 4× can be inferred accordingly.

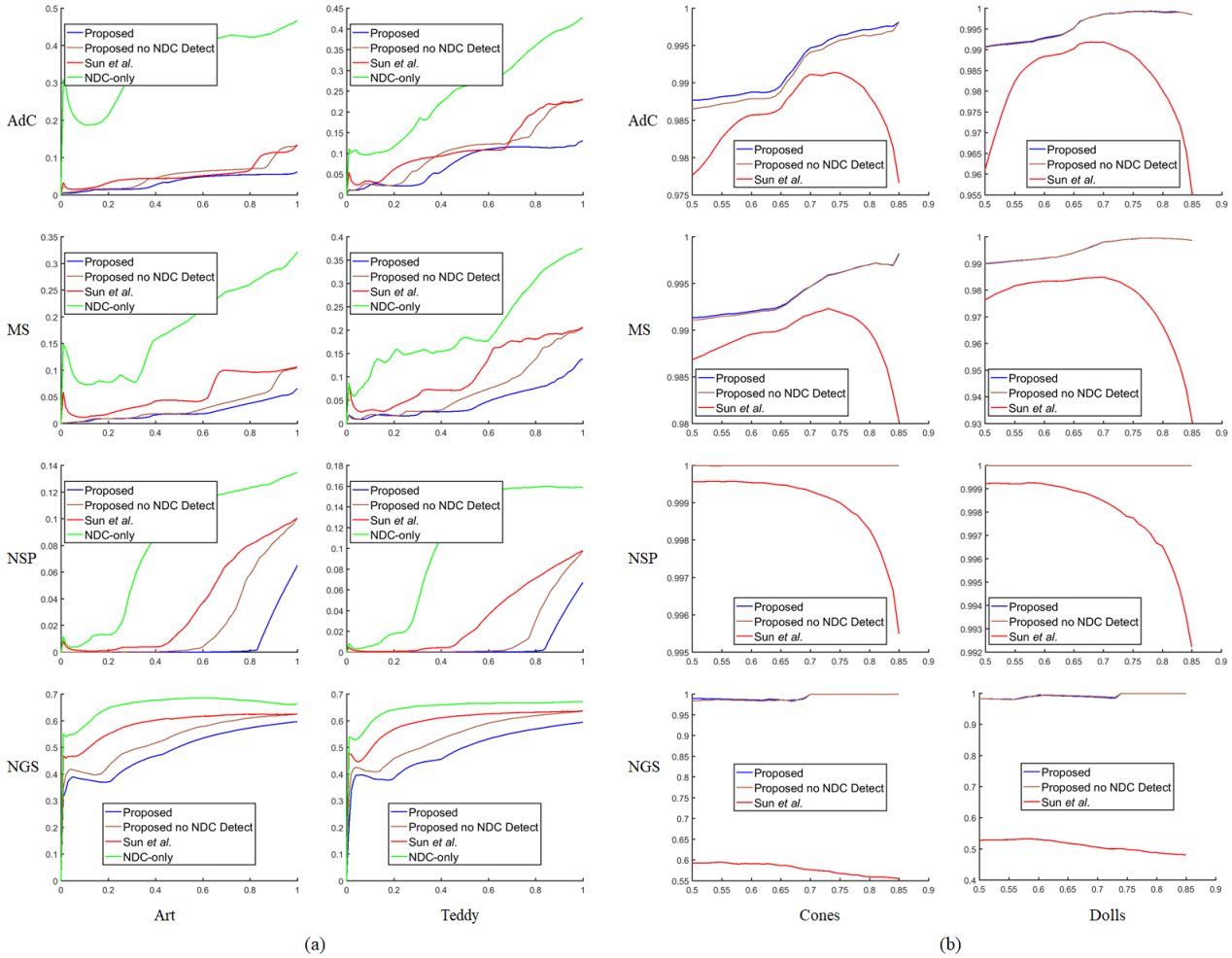


FIGURE 7. Comparison of the ROCs in terms of the error rate (a) and precision (b) under different confidence evaluation strategies. Art, Teddy, Cones and Dolls, which are listed in the columns, are used as the test disparity maps. AdC, MS, NSP and NGS, which are listed in the rows, are used as the test disparity sources. The ROCs, which are from our proposed method, our method without NDC detection, the method proposed by Sun *et al.* [19] and our method only in the NDC area, are plotted in different colors. Each point on the lines in (b) represents the precision under the condition that disparity with confidence higher than the threshold (horizontal axis) is determined to be high-confidence disparity.

1) CONFIDENCE PARAMETERS

As discussed in Section III-B, we set $\tau_{C_{min}} = 0.001$, $\tau_{C_{wdc}} = 0.01$ and $\tau_{C_{ndc}} = 0.1$. In the simple confidence evaluation, we set $\tau_{lrcc} = \frac{15}{255}$ and $\tau_{lrdc} = [1, 2, 4, 8, 10]$. Additionally, we set $\tau_{msm} = 0.7$, $\tau_{cur} = 0.5$, $\tau_{pkrn} = 3$, and $\tau_{lrd} = 7$. The window sizes of the NDC-TS and NDC-ND calculations were the same as those for the ZNCC cost calculation: [1, 2, 3, 4, 5].

2) OTHER PARAMETERS

We set $\tau_{h_{maxnum}} = 15$ and $\tau_{l_{bin}} = [1, 2, 4, 8]$, with a trade-off being made between speed and accuracy. Since the scales of disparity were different in different layers of our framework, we set $\tau_{range} = [1, 2, 4, 6, 8]$ in the CBFT module. Regarding the guided parameters, σ_c was set as $\frac{9}{255}$ for all modules, σ_d was set as $\frac{70}{255}$ for the CBU module and $\frac{10}{255}$ for the CBOPT module, and σ_s was set according to the window size. α in the CBOPT module was set to 0.9.

B. CONFIDENCE ANALYSIS

In this experiment, the accuracy and precision of our confidence evaluation method were analyzed. The cost value in the confidence evaluation was computed based on the ZNCC defined in (6). The comparison in terms of the NSP map between the method of Sun *et al.* and our method is shown in Fig. 8. First, our proposed method performed better in detecting the NDC points according to Fig. 8(b), (d), as there was more noise in the NDC areas corresponding to the results of Sun *et al.* and it was proved that setting the confidence of NDC points with low value was an effecitve strategy. Then, our method achieved a more accurate result in the confidence evaluation of WDC points according to Fig. 8(c), as our results in WDC areas were smoother than those of Sun *et al.*. Furthermore, we analyzed our method in terms of accuracy and precision in more detail. The NSP, NGS, MS and AdC sources were used to provide target disparity maps with error points.

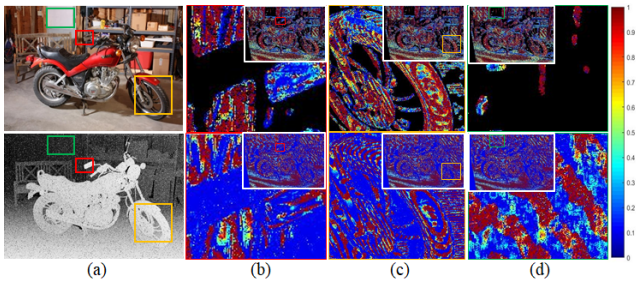


FIGURE 8. Comparison between the confidence maps of the method of Sun *et al.* and our method. The maps in (a) represent the input disparity map and its color map. The maps in (b)-(d) represent the confidence jet color maps evaluated using the method of Sun *et al.* (bottom) and our method (top). They include the occlusion area, well-conditioned area and textureless area. The black points in the upper maps represent the NDC points.

1) CUMULATIVE ERROR DISTRIBUTION ANALYSIS

Receiver operating characteristic (ROC) curves of the error rate were examined to analyze the error distribution in the confidence evaluation [18], [51], [52]. Each point $P(C, R)$ on the ROC curve represents the error rate (R, the vertical axes) of the region in the target disparity map, with the confidences of points in this region being higher than C (the horizontal axes). The area under the ROC curve (AUC) measures the ability of the confidence to predict correct matches. According to [18], the smaller the AUC, the better the method.

As shown in Fig. 7(a), our proposed method outperformed that of Sun *et al.*, as the ROC curve of our method was at the bottom and had a smaller AUC. More conclusions are listed below. First, our modifications yielded obvious improvements according to the red and brown curves. Second, the confidence in the NDC area had the worst performance according to the green curves. In contrast, the blue curves, which correspond to the confidence evaluated using our proposed method in WDC areas, indicate the best performance. The final error ratios of the blue curves were considerably lower than those of the green curves. Thus, it is necessary to separate NDC points from WDC areas and take them as low-confidence points, as discussed in Section III-B.

2) PRECISION ANALYSIS

High precision[2] is necessary for confidence to play a guiding role, as the error will be propagated during upsampling if the error points are evaluated as high confidence. As shown in Fig. 7(b), our proposed method presented the highest precision. Since it did not utilize sufficient stereo information, the method proposed by Sun *et al.* performed worse under some particular conditions. Furthermore, the precision presented a positive correlation with our proposed confidence in most cases. This illustrates the efficient guidance our confidence offers in most cases.

[2]Precision is defined as the ratio of the number of correct disparity points with high confidence to the number of all high confidence points.

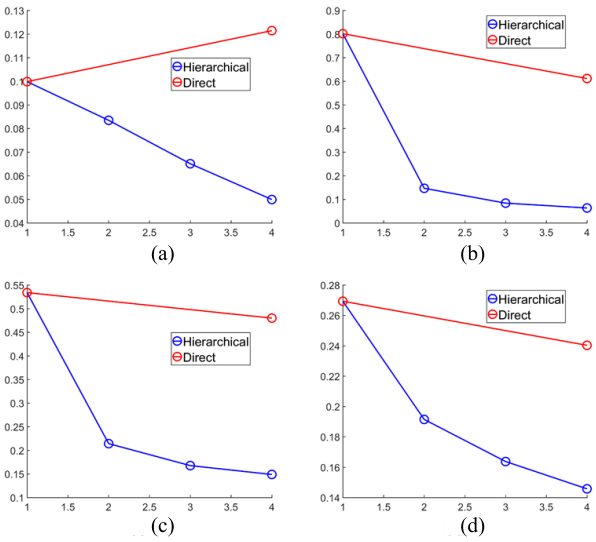


FIGURE 9. Comparison (in terms of the mean error ratio) between our proposed hierarchical structure and the direct structure at an 8× upsampling rate. Horizontal axis represents the layer number; vertical axis represents the error ratio. (a)-(d) use the datasets from NSP, NGS, AdC and MS, respectively. The hierarchical structure is plotted by the blue line. The direct structure is plotted by the red line.

C. FRAMEWORK ANALYSIS

In this section, our framework is analyzed in detail. The analysis considers both the role of modules and the impact of the hierarchical structure. In the following tables, the results in bold, underline and italic formats represent the best results, the second-best results and the worst results, respectively.

1) HIERARCHICAL STRUCTURE ANALYSIS

The direct upsampling method (no hierarchical structure) was compared with our hierarchical structure. As shown in Fig. 9, the hierarchical strategy outperformed the direct strategy on all datasets. Furthermore, the error ratio of the disparity decreased layer by layer in the hierarchical structure, as our proposed modules in each layer were combined to perform denoising under the condition of lossless original accuracy.

2) UPSAMPLING MODULE ANALYSIS

To evaluate the upsampling module, we designed the experiments by adjusting the upsampling module in our framework via the method mentioned in Section III-C. The disparity upsampled results obtained at the 8× upsampling rate for different source inputs are shown with the mean error ratio in Table 3. In this table, UpBIU used only the strategy described in Section III-C1, UpCJCU used only the strategy described in Section III-C2, and UpWVU used only the strategy described in Section III-C3. The last row represents our complete upsampling module. From this table, we can draw the following conclusions. First, the BIU strategy performed well on the GT datasets and maintained the original noise distribution. Second, our proposed CJCU strategy successfully addressed the upsampling task with a complex uncertain

TABLE 1. Quantitative upsampled results of our proposed and other strategies with the GT inputs at three upsampling rates.

	Motorcycle			Recycle			Cones			Reindeer			Mean		
	4x	8x	16x	4x	8x	16x	4x	8x	16x	4x	8x	16x	4x	8x	16x
GIF	20.97	22.33	25.34	8.32	8.78	9.95	13.61	14.77	17.52	11.16	12.14	14.26	13.94	14.98	17.33
BIU	7.62	12.17	18.88	3.31	5.42	8.45	4.02	6.88	11.91	3.69	6.22	10.88	4.92	7.96	12.72
JBU	16.66	23.88	39.55	7.28	10.31	18.56	8.65	12.71	19.83	8.49	14.75	27.97	11.85	17.57	29.00
JGU	6.80	11.03	18.03	3.08	4.98	8.48	3.59	6.24	11.54	3.13	5.84	11.82	4.55	7.48	12.83
MBU	7.62	12.16	18.84	3.31	5.42	8.43	4.02	6.85	11.71	3.68	6.21	10.76	4.92	7.95	12.66
WMF	5.59	11.08	19.22	2.16	4.78	9.93	3.23	6.42	-	2.93	6.54	-	3.14	6.74	45.47
NAFDU	6.87	11.39	18.34	2.96	5.05	9.30	3.62	6.38	11.57	3.34	6.24	11.60	3.76	6.66	11.98
MSG-Net	8.51	12.60	24.97	3.44	5.11	15.32	**1.64**	2.81	5.21	**1.38**	**2.09**	**3.71**	2.80	**4.26**	8.12
FGI	7.70	12.87	23.42	3.75	6.33	11.50	3.07	5.16	10.44	2.90	5.80	16.56	21.91	51.60	77.27
Proposed	**3.35**	**5.69**	**9.81**	**1.53**	**3.27**	**6.06**	1.69	**2.48**	**4.68**	2.32	4.18	7.78	**2.63**	4.55	**7.16**

TABLE 2. Quantitative upsampled results of our proposed and other strategies for inputs with different noises at three upsampling rates.

	AdC			MS			NSP			NGS		
	4x	8x	16x	4x	8x	16x	4x	8x	16x	4x	8x	16x
origin	30.66	57.34	75.39	18.18	29.22	52.00	10.01	9.93	9.96	80.22	80.18	80.31
GIF	30.89	55.57	76.70	18.12	26.83	48.61	53.41	60.16	65.57	16.36	29.06	49.83
BIU	27.69	54.94	75.40	15.45	25.69	49.06	44.44	43.62	46.41	76.00	76.24	76.83
JBU	30.01	56.68	79.28	17.38	28.11	50.55	51.52	55.07	61.90	13.58	21.31	34.62
JGU	26.66	54.03	74.88	15.24	25.11	47.30	26.45	25.78	26.35	64.12	67.51	71.95
MBU	27.67	54.92	75.41	15.44	25.68	49.01	44.44	43.59	46.31	76.00	76.24	76.83
WMF	21.72	53.62	84.80	13.57	25.40	69.93	5.54	12.70	19.57	78.13	80.49	81.78
NAFDU	22.08	53.25	75.80	13.71	25.27	52.21	11.06	13.60	18.32	77.34	78.76	79.83
MSG-Net	24.35	53.42	75.50	14.81	25.90	51.16	62.94	63.09	63.42	81.40	81.29	81.78
FGI	21.91	51.60	77.27	14.29	25.35	48.22	46.29	48.06	52.11	20.90	25.62	34.69
Proposed	**16.70**	**17.58**	**19.86**	**12.27**	**15.98**	**20.61**	**2.93**	**5.34**	**7.70**	**5.20**	**6.84**	**9.92**

noise distribution. Third, our proposed WVU strategy was good at addressing the upsampling task with a certain noise distribution. Finally, our complete CBU strategy yielded the best results on datasets with different noise distributions, as the confidence guidance was used to combine the advantages of the strategies listed above effectively.

3) OTHER MODULE ANALYSIS

To analyze the modules and the hierarchical structure used in our framework, experiments in which the module to be analyzed was removed from the whole framework were designed. The resulting upsampled disparity maps at the 8× upsampling rate for different input sources are shown in terms of the mean error ratio in Table 4. In this table, NoConf represents the upsampling strategy without confidence guidance[3]; NoCBFT represents the upsampling strategy without the CBFT process; and NoCBOPT represents the upsampling strategy without the CBOPT process. Now, the following conclusions can be drawn based on Table 4. First, the confidence held the most vital position in the proposed framework, as the worst performance was achieved without this guidance. Second, the CBFT module was essential for the datasets with complex noise distributions. Furthermore, no extra noise was added by CBFT to the results of GT. Third, the CBOPT module had the effect of refining the results from the CBU and CBFT modules under any inputs. Finally, when the strategies were combined, our framework was robust against datasets with all the types of noise distributions mentioned in the experiments.

[3]Note that the CBFT was also invalid since the confidences of all points were set to 1.

TABLE 3. Quantitative upsampled results (in terms of mean error ratio) of our proposed strategy with different upsampling modules.

	GT	AdC	MS	NSP	NGS
UpBIU	5.87	19.97	17.88	36.14	42.52
UpCJCU	6.43	16.33	14.66	10.17	7.79
UpWVU	4.02	16.43	16.30	5.08	6.87
complete	**3.75**	**14.88**	**14.57**	**4.75**	**6.34**

TABLE 4. Quantitative upsampled results (in terms of mean error ratio) of our proposed strategy without a specific module.

	GT	AdC	MS	NSP	NGS
NoConf	*5.42*	*48.28*	*24.10*	*6.37*	*16.87*
NoCBFT	3.76	32.78	19.28	**4.74**	6.97
NoCBOPT	4.02	15.87	15.52	5.14	6.88
complete	**3.75**	**14.88**	**14.57**	4.75	**6.34**

D. COMPARISONS OF THE DISPARITY RESULTS

The classical and state-of-the-art upsampling methods were compared with our proposed method. Among these methods, BIU [20], GIF [22], JGU [27], WMF [28], NAFDU [25], MSG-Net [37] and FGI [17] were executed using source codes found online, while JBU [23] and MBU [14] were implemented by us using CUDA. And MSG-Net was learning-based method. BIU and GIF were

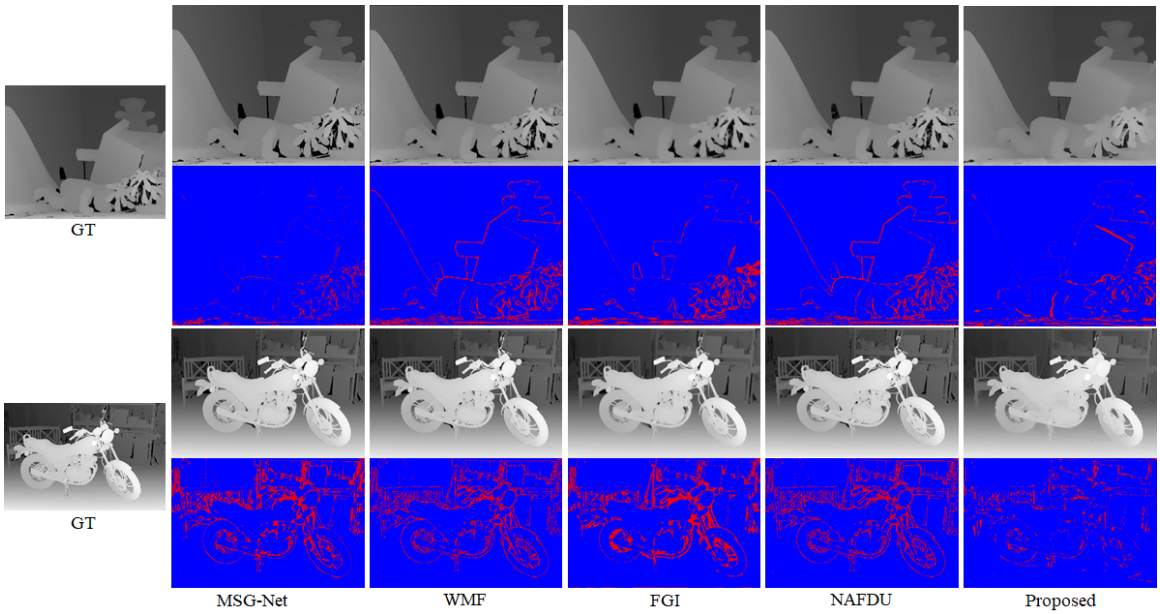


FIGURE 10. The upsampled results obtained on the GT datasets. The first column represents the input sources; other columns contain the upsampled results of different methods.

interpolation-based methods. JBU, JGU, WMF and NAFDU were weighted-filter-based methods. FGI and MBU were MRF-based methods. Ten Middlebury datasets with different environments were used for the comparison. Five different input sources with different noise distributions were applied to measure the abilities of the upsampling methods. The comparisons of the methods are detailed in Tables 1 and 2 and Fig. 10, 11 and 12, where the gray and red-blue images represent the disparity maps and error maps,[4] respectively. The results of the proposed method, together with the top four results of the comparison methods, are listed in the figures.

1) COMPARISON ON GT DATASETS

The input source without noise was applied to measure the ability of the methods to maintain the original accuracy in the upsampling process. The results of different methods for the GT input source are listed in Table 1. As shown, the error ratios of some of the maps (Motorcycle, Recycle, Cones and Reindeer) are listed; the mean results listed in the last column are the mean error ratios of all maps.[5] According to the table, our proposed method and the learning-based MSG-Net method performed the best. Moreover, our method was more robust, as MSG-Net achieved poor results on the Motorcycle and Recycle maps.

As shown in Fig. 10, our proposed method performed as well as the MSG-Net method [37]

and outperformed all the other methods. Furthermore, our method presented better results than those of the other three methods (WMF, FGI and NAFDU), especially at the edges.

[4]Red points denote the error points.

[5]The maps are Motorcycle, Recycle, Piano, Teddy, Cones, Art, Dolls, Baby1, Reindeer and Cloth1.

2) COMPARISON ON NSP AND NGS DATASETS

NSP and NGS source inputs were used to simulate the disparity from the RGBD cameras with specific noise information. The NSP datasets were created with salt and pepper noise, which satisfies the uniform distribution requirement. The NGS datasets were created with Gaussian noise. As shown in Table 2, the results for NSP and NGS listed in the last two columns represent the mean error ratios. According to the table, our proposed method yielded the best result among all the methods at all upsampling rates. The JBU method took the second-best position on the NGS datasets, as the JBU method assumes noise with a Gaussian distribution. The WMF and NAFDU methods performed well on the NSP datasets, as the median filter and the filter proposed by NAFDU were good at removing the salt and pepper noise. Then, learning-based method MSG-Net met with poor results, since it was sensitive to the content of inputs. Furthermore, with confidence guidance and the hierarchical structure, our proposed method presented a clear advantage. As shown in Fig. 11, the results of our proposed methods on NSP and NGS proved that our method can remove noise with a specific distribution, in contrast to the other state-of-the-art methods, since most noise points (except for the occlusion points) were corrected by our proposed method.

3) COMPARISON ON ADC AND MS DATASETS

AdC and MS source inputs were used to determine the disparity from stereo matching with unknown noise information. As shown in Table 2, the results for AdC and MS listed in the first two columns represent the mean error ratios. The disparity image results with the error maps are shown in Fig. 12. According to the table and the figures, our proposed method

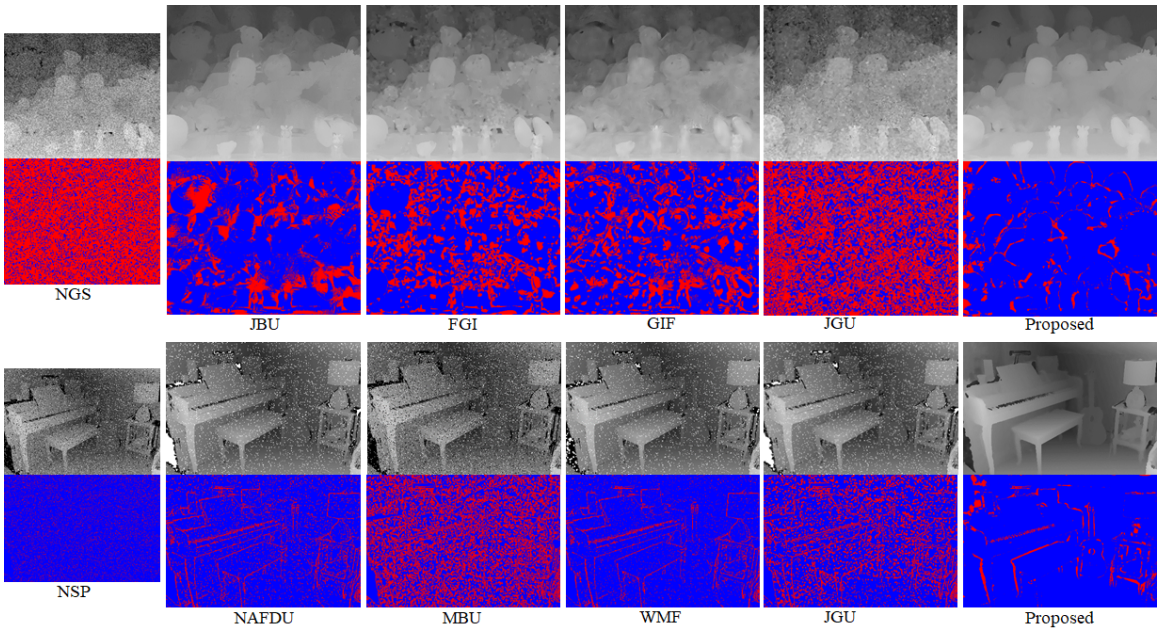


FIGURE 11. The upsampled results obtained on the NSP and NGS datasets. The first row shows the results obtained on the NGS datasets. The second row shows the results obtained on the NSP datasets. The first column represents the input sources; other columns contain the upsampled results of different methods.

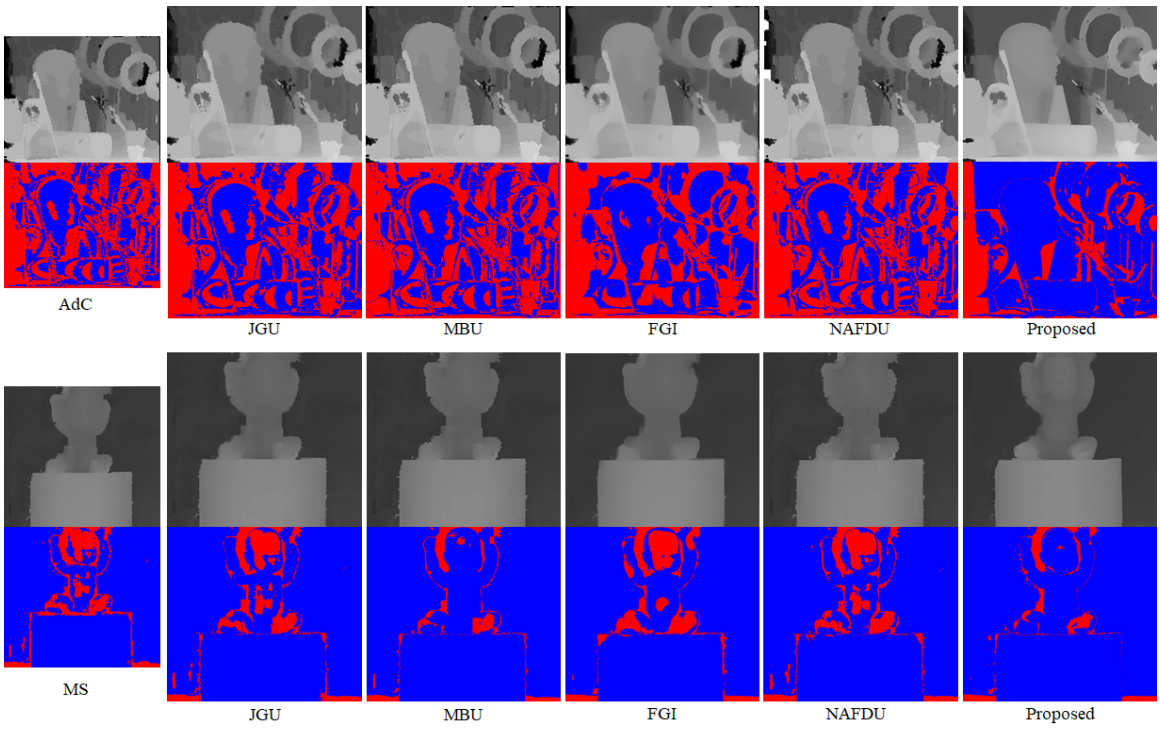


FIGURE 12. The upsampled results obtained on the AdC and MS datasets. The first row shows the results obtained on the AdC datasets. The second row shows the results obtained on the MS datasets. The first column represents the input sources; other columns contain the upsampled results of different methods.

performed the best, being clearly superior, especially at higher upsampling rates. Moreover, the results of other methods listed in the table are similar to the original source. Therefore, it was concluded that the other methods were less competitive in terms of denoising on datasets with unknown noise distributions. In contrast, the hierarchical structure and confidence-based modules of our method resulted in a robust noise suppression. Moreover, among the modules, the CBFT

module corrected the error disparity via robust fine-tune on the datasets with unknown noise distributions.

E. TIME CONSUMPTION

The time consumption was related to the upsampling rate and the target resolution while almost independent of the input sources. Thus, it was evaluated using the mean values at the 3 upsampling rates. As listed in Table 5, the time consumption of our framework had a positive correlation with the target upsampling resolution. Furthermore, the time consumption of CBFT at all upsampling rates was fairly consistent, as most error disparities were corrected at low levels in the hierarchical framework.

TABLE 5. Mean time consumption (in milliseconds, ms) of the modules for 2 resolutions at 3 upsampling rates.

	1792x1488/ms				1376x1104/ms			
	2x	4x	8x	16x	2x	4x	8x	16x
CBU	260	317	328	330	146	177	183	187
CBFT	152	186	179	175	73	85	87	87
CBOPT	81	160	161	161	48	92	93	94
all	492	663	668	666	267	354	363	368

V. FUTURE WORKS

According to the experimental results, the confidence evaluation plays a vital role in our framework. However, the confidence in the NDC area cannot be measured well through current strategy and more information, such as structural information in the image, should be used to measure confidence. In the future, emphasis will be placed on the establishment of an efficient confidence measuring scheme that is able to offer a more precise evaluation for every pixel in the input disparity maps. Advanced methods such as learning-based ones are promising, as a learned confidence value can be, on average, more accurate and robust. Improved specific upsampling strategies for stereo matching inputs will be considered, as the stereo information in this work was mainly used to measure the confidence.

VI. CONCLUSIONS

A novel, confidence-based, multistrategy hierarchical framework, which can handle an upsampling task with input low-resolution disparity corrupted with multidistribution noise, was proposed in this study. The proposed efficient confidence evaluation has an advantage in terms of disparity quality monitoring and plays a role in the absolute guidance of the framework. A hierarchical structure combined with several modules results in a robust and efficient framework for disparity upsampling and enhancement. This proposed method has clear advantages on noisy and noise-free datasets compared with previous state-of-the-art methods. Furthermore, the real-time speed of the framework makes further acceleration via code optimization and the use of an FPGA possible.

REFERENCES

[1] J. Geng, "Three-dimensional display technologies," *Adv. Opt. Photon.*, vol. 5, no. 4, pp. 456–535, Dec. 2013. [Online]. Available: http://aop.osa.org/abstract.cfm?URI=aop-5-4-456
[2] R. Pei, Z. Geng, Z. X. Zhang, X. Cao, and R. Wang, "A novel optimization method for lenticular 3-D display based on light field decomposion," *J. Display Technol.*, vol. 12, no. 7, pp. 727–735, 2016.
[3] R. Wang, Z. Geng, Z. Zhang, and R. Pei, "Visualization techniques for augmented reality in endoscopic surgery," in *Proc. Int. Conf. Med. Imag. Augmented Reality*, 2016, pp. 129–138.
[4] S. Nicolau, L. Soler, D. Mutter, and J. Marescaux, "Augmented reality in laparoscopic surgical oncology," *Surgical Oncol.*, vol. 20, no. 3, pp. 189–201, 2011.
[5] S. Bernhardt, S. A. Nicolau, L. Soler, and C. Doignon, "The status of augmented reality in laparoscopic surgery as of 2016," *Med. Image Anal.*, vol. 37, pp. 66–90, Apr. 2017.
[6] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2320–2327.
[7] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.
[8] M. Bleyer and C. Breiteneder, *Stereo Matching—State-of-the-Art and Research Challenges*. London, U.K.: Springer, 2013, pp. 143–179.
[9] J. Geng, "Structured-light 3D surface imaging: A tutorial," *Adv. Opt. Photon.*, vol. 3, no. 2, pp. 128–160, 2011.
[10] M. Hansard, S. Lee, O. Choi, and R. P. Horaud, *Time-of-Flight Cameras: Principles, Methods and Applications*. Springer, 2012.
[11] I. Eichhardt, D. Chetverikov, and Z. Jankó, "Image-guided ToF depth upsampling: A survey," *Mach. Vis. Appl.*, vol. 28, nos. 3–4, pp. 267–282, 2017, doi: 10.1007/s00138-017-0831-9.
[12] D. Csetverikov, I. Eichhardt, and Z. Jankó, "A brief survey of image-based depth upsampling," in *Proc. KÉPAF*, 2015, pp. 279–294.
[13] L. P. J. Vosters, C. Varekamp, and G. de Haan, "Evaluation of efficient high quality depth upsampling methods for 3DTV," *Proc. SPIE*, vol. 8650, p. 865005, Mar. 2013.
[14] W. Liu, S. Jia, P. Li, X. Chen, J. Yang, and Q. Wu, "An MRF-based depth upsampling: Upsample the depth map with its own property," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1708–1712, Oct. 2015.
[15] J. I. Jung and Y. S. Ho, "Depth image interpolation using confidence-based Markov random field," *IEEE Trans. Consum. Electron.*, vol. 58, no. 4, pp. 1399–1402, Nov. 2012.
[16] W.-S. Jang and Y.-S. Ho, "Efficient disparity map generation using stereo and time-of-flight depth cameras," in *Proc. Pacific Rim Conf. Multimedia*, vol. 9315. New York, NY, USA: Springer-Verlag, 2015, pp. 623–631, doi: 10.1007/978-3-319-24078-7_64.
[17] Y. Li, D. Min, M. N. Do, and J. Lu, "Fast guided global interpolation for depth and motion," in *Computer Vision—ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 717–733, doi: 10.1007/978-3-319-46487-9_44.
[18] X. Hu and P. Mordohai, "A quantitative evaluation of confidence measures for stereo vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2121–2133, Nov. 2012.
[19] L. Sun, K. Chen, M. Song, D. Tao, G. Chen, and C. Chen, "Robust, efficient depth reconstruction with hierarchical confidence-based matching," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3331–3343, Jul. 2017.
[20] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.
[21] N. Fukushima, K. Takeuchi, and A. Kojima, "Self-similarity matching with predictive linear upsampling for depth map," in *Proc. 3DTV-Conf., True Vis., Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jul. 2016, pp. 1–4.
[22] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[23] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, p. 96, Jul. 2007.
[24] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[25] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noise-aware filter for real-time depth upsampling," in *Proc. Workshop Multi-Camera Multi-Modal Sensor Fusion Algorithms Appl.*, 2008, pp. 1–13.

[26] C. Xiao, Y. Nie, W. Hua, and W. Zheng, "Fast multi-scale joint bilateral texture upsampling," *Vis. Comput.*, vol. 26, no. 4, pp. 263–275, Apr. 2010, doi: 10.1007/s00371-009-0409-2.

[27] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. CVPR*, Jun. 2013, pp. 169–176.

[28] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1176–1190, Mar. 2012.

[29] W. Zhang, W. Zhang, K. Liu, and J. Gu, "A feature descriptor based on local normalized difference for real-world texture classification," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 880–888, Apr. 2018.

[30] W. Zhang, X. Yu, and X. He, "Learning bidirectional temporal cues for video-based person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2017.2718188.

[31] W. Zhang, Q. Chen, W. Zhang, and X. He, "Long-range Terrain perception using convolutional neural networks," *Neurocomputing*, vol. 275, pp. 781–787, Jan. 2018.

[32] W. Zhang and W.-K. Cham, "Hallucinating face in the DCT domain," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2769–2779, Oct. 2011.

[33] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[34] R. Timofte, V. De, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.

[35] H.-H. Kwon, Y.-W. Tai, and S. Lin, "Data-driven depth map refinement via multi-scale sparse representation," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 159–167.

[36] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[37] T.-W. Hui, C. C. Loy, and X. Tang, "Depth map super-resolution by deep multi-scale guidance," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 353–369.

[38] X. Meng, Z. Zhang, Z. Geng, and M. Zhang, "Mrf-based disparity upsampling using stereo confidence evaluations," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 561–565, Apr. 2018.

[39] A. Buades and G. Facciolo, "Reliable multiscale and multiwindow stereo matching," *SIAM J. Imag. Sci.*, vol. 8, no. 2, pp. 888–915, 2015, doi: 10.1137/140984269.

[40] S. Hermann and R. Klette, "Evaluation of a new coarse-to-fine strategy for fast semi-global stereo matching," in *Advances in Image and Video Technology*, Y.-S. Ho, Ed. Berlin, Germany: Springer, 2012, pp. 395–406.

[41] L. Tang, M. K. Garvin, K. Lee, W. L. M. Alward, Y. H. Kwon, and M. D. Abràmoff, "Robust multiscale stereo matching from fundus images with radiometric differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2245–2258, Nov. 2011.

[42] S. Mattoccia, S. M. Updates, and S. M. Outline, "Stereo vision: Algorithms and applications," Tech. Rep., 2011.

[43] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[44] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[45] W. Zhang and W.-K. Cham, "Gradient-directed multiexposure composition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2318–2323, Apr. 2012.

[46] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Proc. ICCV Workshops*, Nov. 2011, pp. 467–474.

[47] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2003, pp. I-195–I-202.

[48] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[49] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[50] C. Zhang, Z. Li, Y. Cheng, R. Cai, H. Chao, and Y. Rui, "Meshstereo: A global stereo model with mesh alignment regularization for view interpolation," in *Proc. ICCV*, Dec. 2015, pp. 2057–2065.

[51] M. Gong and Y.-H. Yang, "Fast unambiguous stereo matching using reliability-based dynamic programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 998–1003, Jun. 2005.

[52] P. Mordohai, "The self-aware matching measure for stereo," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1841–1848.

XIANG-BING MENG was born in Liaocheng, Shandong, China, in 1990. He received the B.S. degree in electric engineering from the Harbin Institute of Technology in 2014. He is currently pursuing the Ph.D. degree in computer vision with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

His research interests include image processing, 3-D reconstruction, 3-D displays, SLAM, and machine learning.

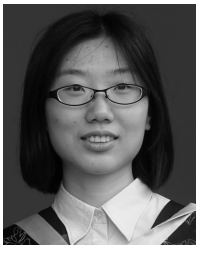
MEI ZHANG was born in Cangzhou, Heibei, China, in 1984. She received the B.S. degree in physics from Hebei University in 2004 and the Ph.D. degree from Nankai University in 2010. From 2007 to 2008, she was involved in some research with the Optometry School, Indiana University.

From 2010 to 2012, she held a post-doctoral position at the Institute of Automation, Chinese Academy of Science, where she has been an Associate Professor since 2013. She has authored over 20 articles. Her research interests include optical design, 3-D image acquisition and display, and vision science.

She received support from the National Nature Science Foundation of China Program and participated in the National High Technology Research and Development Program of China.

ZHAO-XING ZHANG received the B.S. and M.S. degrees in electronic engineering and photoelectric technology from the Nanjing University of Science and Technology in 2007 and 2009, respectively.

From 2009 to 2017, he was a Research Assistant with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision and machine learning.



RONG WANG received the B.S. degree in automation from Beihang University, Beijing, China, in 2013, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 2018.

She is currently a Researcher with the China Electronics Technology Group Corporation. Her research interests include augmented reality, 3-D tracking, 3-D registration, and SLAM.

ZHENG GENG received the B.S. and M.S. degrees from the Nanjing University of Science and Technology and the Ph.D. degree from The George Washington University in 1990.

In 1991, he has served as the Department Head of Intelligent Automatic Inc., USA. In 1994, he found Genex Technologies Inc., USA. In 2005, he served as a Professor with the Institute of Automation, Chinese Academy of Science. He published hundreds of papers in high-level journals. He received nearly 20 national invention patent grants. His research interests mainly focused on the biomedical engineering, 3-D imaging, and 3-D display.

He presided over hundreds of scientific research projects for research and technology industrialization, which were supported by the National Nature Science Foundation of China, the National High Technology Research and Development Program of China, and other local governments.

FEI-YUE WANG (F'03) was born in Qingdao, Shandong, China, in 1961. He received the B.S. degree in mechanism engineering from the Qingdao University of Science and Technology in 1982, the M.S. degree in mechanical engineering from Zhejiang University in 1984, and the Ph.D. degree in computer and systems engineering from the Rensselaer Polytechnic Institute in 1990.

Since 1990, he has been an Assistant Professor (1990), an Associate Professor (1995), and a Full Professor (1999) with The University of Arizona, where he was also the Director of the Robotics and Automation Laboratory in 1990. He was the Director of the Advanced Research Center for Complex Systems in 1999 and the Director of the China-U.S. Higher Research and Education Center in 2004. In 2002, he served as the Director of the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, where he was appointed as the Deputy Director of the Institute of Automation in 2006. In 2005, he served as the Dean of the School of Software, Xi'an Jiaotong University. He has published over 600 papers, over 20,000 times of Google citations, and an H-Index coefficient of 54. His research interests are mainly about the modeling for complex and intelligent systems, social computing, and parallel intelligence. He has been the Chairman of the IEEE ITS Institute since 2006.

In 2007, he was elected as a fellow of International Federation of Automation, American Association for Advancement of Sciences, and the American Society of Mechanical Engineers. He was elected as the Distinguished Scientist of Association for Computing Machinery in 2007. He received the Second Prize of the 2007 National Natural Science Award for the study in intelligent control in 2008. He was a recipient of the IEEE SMC Norbert Wiener Award in 2014.