A Low-Complexity End-to-End Stereo Matching Pipeline From Raw Bayer Pattern Images to Disparity Maps

SHENGYU GAO, (Graduate Student Member, IEEE), HONGYU WANG, (Member, IEEE)

1School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China
2Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China
3University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Xin Lou (louxin@shanghaitech.edu.cn)

This work was supported by the Natural Science Foundation of China under Grant 61801292.

ABSTRACT Conventional computer vision algorithms, including stereo matching algorithms, take finely rendered color images as input. However, existing image signal processing (ISP) pipelines for color image generation are designed for photography with a goal of generating pleasing images for human eyes. This paper describes a new end-to-end pipeline for stereo matching from raw Bayer pattern images to disparity maps with customized ISP. Unlike conventional stereo matching systems which need a complete ISP module to render full-size standard RGB (sRGB) images, a subsampling-based demosaicing-downsampling (SDD) operation is introduced in the proposed pipeline to demosaic and downsample the Bayer pattern images. The resultant half-size color image pairs are processed with simple denoising and tone mapping algorithms to generate the final input images of stereo matching algorithms. It is found that the simple nearest neighbor upsampling method is good enough to generate the final full-size disparity maps. Experimental results show that the proposed pipeline is capable of generating comparable or even better stereo matching results than the conventional pipeline. By skipping most of the unnecessary ISP steps and reducing the size of input images, the computational complexity of the end-to-end stereo matching pipeline is significantly reduced.

INDEX TERMS Stereo matching, Bayer image, image signal processing (ISP), low-complexity.

I. INTRODUCTION Stereo vision has received great attention over the past several decades. It is widely applied in various applications such as autonomous driving, robots and navigation. Stereo matching is one of the fundamental problems in stereo vision. Many sophisticated algorithms have been proposed to solve the matching problem based on rectified images. Generally, all these stereo matching algorithms can be categorized into four types, which are local methods [1]–[3], global methods [4]–[6], semi-global methods [7], [8] and neural network-based methods [9]–[11].

It is well known that stereo matching algorithms are time and power hungry, making it challenging to be deployed in resource-limited embedded devices. For most of the stereo matching algorithms, computational complexity is proportional to the size of the input image pairs, i.e., higher resolution input images consume more computation. While on the other hand, higher resolution inputs do not necessarily result in better matching accuracy. According to the rankings on the Middlebury stereo benchmark, higher resolution may even lead to degraded results for some of the stereo matching algorithms [7], [9], [11]. For example, the semi-global block matching (SGBM) implemented in OpenCV returns bad 2.0 pixel error of 27.9%, 23.8%, 28.4% for quarter-size, half-size and full-size images respectively, on the Middlebury datasets [12]. Therefore, it is proposed in [7] to downsample high resolution image pairs first and calculate disparity on low resolution pairs. The resultant disparity maps are then upsampled to generate the corresponding high resolution ones.

Most modern digital cameras utilize color filter array (CFA) to capture color images. Among the existing CFAs, Bayer array is one of the most popular patterns.
As shown in Fig. 1, there is only one color component for each pixel in a raw Bayer pattern image. To render the final color image, a series of image signal processing (ISP) steps are applied [13]. Note that although the specific design of the ISP pipeline may vary for different vendors, they are almost all optimized to generate pleasing and recognizable images for human consumption. It has been shown that the ISP pipeline may introduce errors and harm the original information from image sensors [14]. Recent studies have found that some stages in the conventional ISP pipelines are redundant for modern computer vision algorithms, such that they can be skipped to simplify the vision system designs [15]–[17]. Moreover, it has been shown in [18] that some stages in the ISP pipeline can be modified or even redesigned to improve the vision performance. It is well known that ISP algorithms are computation intensive and consume a significant amount of processing time and power. Therefore, we believe the ISP pipeline for stereo matching should be designed specifically (instead of using existing ISP for photography) to not only save computation, but also (possibly) improve the stereo matching results.

In this paper, we propose an end-to-end pipeline for stereo matching from Bayer pattern images to disparity maps. Although modern deep learning methods achieve relatively better performance on different stereo matching benchmarks, they are obsessed by the huge computation and domain gap [20]. Huge computation means these methods are not suitable for resource constraint embedded platforms. Domain gap makes deep learning methods lose generalization ability once they overfit the models to specific domains. In this paper, we focus on traditional stereo matching algorithms, which are more possible to be implemented on mobile and embedded devices compared to network-based algorithms. As many of the stereo matching algorithms utilize the color information of images, it is reasonable to convert single-channel raw Bayer images to three-channel RGB images. But it is well-known that modern demosaicing algorithms are computation intensive [15]. In contrast to conventional stereo matching systems as illustrated in Fig. 2(a), the demosaicing and downsampling operations are merged into a single step, denoted by subsampling-based demosaicing-downsampling (SDD) operation, in the proposed pipeline. After the proposed SDD operation, a tone mapping step, which is necessary according to [15] is further applied. The resultant half-size image pairs are processed by stereo matching algorithms to generate a half-size disparity map. The half-size disparity map is then upsampled to generate the final full-resolution disparity. Since the size of matching images are reduced to half and a lot of unnecessary ISP stages are removed, the computational complexity are significantly reduced. Moreover, as will be shown in the experiments, the proposed end-to-end stereo matching pipeline is able to generate comparable or even better matching results than the conventional pipeline, even if it skips many steps in conventional ISP.

The rest of this paper is organized as follows. Section II provides a brief overview of the existing stereo matching pipelines. Section III presents the proposed pipeline from raw Bayer images to full-resolution disparity maps. Section IV analyses the effects of some operations in the proposed pipeline and provides a qualitative and quantitative comparison with existing stereo matching pipelines. Finally, the conclusions are drawn in Section V.

II. RELATED WORK

Stereo matching has been a classic problem in computer vision. Traditional methods usually formulate an energy function about a label \( f \) as

\[
E(f) = E_{\text{smooth}}(f) + E_{\text{data}}(f)
\]

and seek the label \( f \) by solving the energy minimization problem. Algorithms that ignore the smoothness term are called local methods [1]–[3]. Otherwise, they are referred to as global methods [4]–[6] or semi-global methods [7], [8]. Although a lot of efforts have been devoted to speed up the stereo matching algorithms, they are still time-consuming, especially with high-resolution input image pairs. Besides, it is found in [21] that those algorithms surprisingly perform worse on high-resolution images than on low-resolution inputs. This may be caused by more pronounced miscalibration effects at higher resolutions [21].

Moreover, for some resource-limited embedded systems, algorithms can only handle low-resolution image pairs [22] to meet the requirement of real-time performance. As shown in 2(a), for practical applications, it is common to downsample the full-size images and compute low-resolution disparity on that. Then, some upsampling methods [23]–[25] are employed to generate high-resolution disparity maps.

Nowadays, most of computer vision algorithms, including stereo matching, are developed based on sRGB color space. As shown in Fig. 2(a), a series of ISP stages are applied on raw Bayer pattern images to generate sRGB color images to be consumed by stereo matching algorithms.

Recently, researchers have realized that it is possible to utilize raw Bayer images after simple processing as inputs for computer vision algorithms. Liu et al. found it is feasible to only perform demosaicing and gamma correction for face detection [26]. Zhou et al. demonstrated the possibility of generating gradient-based features from raw Bayer pattern images [16]. Buckler et al. found that many traditional ISP stages are not necessary for most of the computer vision analysis of some operations in the proposed pipeline. Finally, the conclusions are drawn in Section V.
algorithms they tested [15]. For stereo matching, they chose SGBM implemented in OpenCV for testing, and found that only denoising, demosaicing and tone mapping have significant impacts on the matching results. Based on that, they propose a minimal pipeline for stereo matching as shown in Fig. 2(b).

III. PROPOSED METHOD

It has been shown that, generally, the performance of stereo matching algorithms drops when working on higher resolution images [21]. Besides, for some resource-constrained embedded systems, it is necessary to downsample the input image pairs to meet the requirement of real-time performance.

In this section, the proposed SDD operation is introduced. On the basis of that, an end-to-end stereo matching processing pipeline from raw Bayer images to full-resolution disparity maps are proposed. Fig. 3 illustrates the concept of the proposed end-to-end pipeline. As shown in Fig. 3, half-size images are generated by the proposed SDD operation. The tone mapped half-size image pairs are then processed by stereo matching algorithms to generate a half-size disparity map. After that, the half-size disparity map is upsampled to produce the final full-resolution disparity.

A. BAYER SUBSAMPLING

In the proposed stereo matching pipeline, the demosaicing step and downsampling step are merged into a single SDD operation, where the color difference constancy [19] is assumed. More specifically, as shown in Fig. 4, we take a super pixel consisting of RGGB pixels from a Bayer pattern image and convert it into a single pixel with RGB three channels. In this process, the resolution in both horizontal and vertical directions are reduced by half. As will be shown later

---

**FIGURE 2.** Existing stereo matching pipelines.

(a) Conventional pipeline

(b) Minimal pipeline [15]

**FIGURE 3.** Proposed pipeline.
in this subsection, the downsampling operation is equivalent to bilinear interpolation as long as the color difference constancy holds, whereas the computation is negligible.

Let us consider downsampling an image by half with bilinear interpolation. For example in Fig. 4, the estimated pixel values, denoted by \( \tilde{R}, \tilde{G} \) and \( \tilde{B} \), can be expressed as

\[
\begin{align*}
\tilde{R} &= \frac{1}{4} (R_{1,1} + R_{1,2} + R_{2,1} + R_{2,2}) \\
\tilde{G} &= \frac{1}{4} (G_{1,1} + G_{1,2} + G_{2,1} + G_{2,2}) \\
\tilde{B} &= \frac{1}{4} (B_{1,1} + B_{1,2} + B_{2,1} + B_{2,2}).
\end{align*}
\]

For demosaicing, one of the main hypothesis is the color difference constancy [19], which assumes that the difference between channels of a pixel is constant within a small pixel neighborhood. According to [19], the color difference constancy assumption holds for most of the natural images. Thus, for a super-pixel, as shown in Fig. 4, we have

\[
R_{i,j} - G_{i,j} = c, \quad [i,j] \in \{1,2\},
\]

where \( c \) is a constant. In this work, we take \( 1/2(G_{1,2}+G_{2,1}) \) as the estimated value for the missing components of the green channel, i.e.,

\[
\begin{align*}
G_{1,1} &\approx \frac{1}{2} (G_{1,2} + G_{2,1}) \\
G_{2,2} &\approx \frac{1}{2} (G_{1,2} + G_{2,1}).
\end{align*}
\]

By combining (2), (3) and (4), we can get the value of the downsampled \( \tilde{R} \) channel as

\[
\begin{align*}
\tilde{R} &= \frac{1}{4} (R_{1,1} + R_{1,2} + R_{2,1} + R_{2,2}) \\
&= \frac{1}{4} \left( R_{1,1} + G_{1,2} + c + G_{2,1} + c + G_{2,2} + c \right) \\
&= \frac{1}{4} \left( R_{1,1} + G_{1,2} + G_{2,1} + G_{2,2} + 3 (R_{1,1} - G_{1,1}) \right) \\
&\approx \frac{1}{4} \left( 4R_{1,1} + G_{1,2} + G_{2,1} + \frac{1}{2} (G_{1,2} + G_{2,1}) \right. \\
&\quad - 3 \times \frac{1}{2} (G_{1,2} + G_{2,1}) \\
&= R_{1,1}.
\end{align*}
\]

This means the value of \( \tilde{R} \) channel of the demosaiced and downsampled image is equal to the value of the \( R \) channel in the original super pixel, as long as the color difference holds. Similarly, the value of the \( \tilde{B} \) channel of the demosaiced and downsampled image is \( \tilde{B} = B_{2,2} \). Therefore, as long as the color difference constancy holds, we can generate a demosaiced half-size color image from the corresponding Bayer pattern image as

\[
\begin{align*}
\tilde{R} &= R_{1,1} \\
\tilde{G} &= \frac{1}{2} (G_{1,2} + G_{2,1}) \\
\tilde{B} &= B_{2,2}.
\end{align*}
\]

Compared with the conventional pipeline, the proposed SDD operation does not need any complex computation. As will be shown in the experiments, the proposed SDD operation leads to comparable or even better results than the conventional ISP approaches.

\section*{B. DISPARITY UPSAMPLING}

There are many upsampling methods in existing literatures. Since there is no full-size intensity image to guide the disparity upsampling, some sophisticated methods such as guided filtering [23], [24], [27] are not applicable in the proposed method to generate the full-size disparity map from the half-size one. To figure out which upsampling method is suitable for the proposed pipeline, we carry out disparity upsampling using different methods and compare their performance.

Three commonly used upsampling methods, which are nearest neighbor interpolation (NEAREST), bilinear interpolation (BILINEAR) and the Self-guided (SG) method [25] are evaluated in this work. To address the absence of high-resolution intensity images, the SG method uses the upsampled images as the guiding image and interpolates the residual to generate the high-resolution disparity map. As shown in the results, the NEAREST method not only achieves the best results but also consumes the least computation.

\section*{C. POST-PROCESSING}

It is common for stereo matching to incorporate post-processing to obtain refined dense disparity maps. In the conventional pipeline, post-processing is applied on full-size disparity maps, which are the direct output of stereo matching algorithms. However, the output of matching algorithms is half-size in the proposed pipeline. Such that there are two choices for the sequence of post-processing and upsampling, i.e., post-processing first on half-size disparity maps followed by upsampling or upsampling first followed by post-processing on full-size disparity maps. These two different configurations are studied in the following section.

In this work, a widely employed post-processing method [7], which uses left-right consistency check, peak filtering, disparity interpolation and median filtering, is applied to improve the disparity map.
TABLE 1. Results of different upsampling methods for 2-Pixel accuracy.

| Method    | GF0        | SGBM0   | LocalExp0 | GF1        | SGBM1   | LocalExp1 |
|-----------|------------|---------|-----------|------------|---------|-----------|
|           | nonocc    | all     | nonocc    | all        | nonocc  | all        |
| NEAREST   | 15.09     | 25.64  | 11.70     | 25.00      | 10.97   | 20.92      |
| LINEAR    | 17.00     | 27.71  | 13.23     | 26.32      | 11.89   | 21.96      |
| SG [25]   | 18.57     | 28.99  | 12.80     | 25.95      | 11.52   | 21.58      |

Average error rate using bad 2.0 metric on non-occluded region (nonocc) and all region. The suffix 0 after the name of the algorithm represents the origin disparity without post-processing and 1 represents the refined disparity by post-processing. E.g., LocalExp0 represents disparity maps generated by LocalExp without post-processing, LocalExp1 represents disparity maps generated by LocalExp with post-processing.

IV. EXPERIMENTS

In the experiments, three well-known stereo matching algorithms, which are a) guided filter (GF) [3] (a local method), b) SGBM that implemented in OpenCV (a semi-global method) and c) LocalExp [9] (a PathMatch-based method, which has the best performance for the Middlebury dataset) are used as benchmark algorithms to demonstrate the effectiveness of the proposed stereo matching pipeline from raw Bayer pattern images to disparity maps.

Since the algorithms selected are designed for early datasets, we use the Middlebury 2003 and 2005 [28], [29] as datasets in our experiments. The complete testing dataset contains eight image pairs, which are Cones, Teddy, Art, Books, Dolls, Laundry, Moebius and Reindeer. The toolbox introduced in [15] is used to convert the Middlebury dataset to its corresponding Bayer version.

For GF, we set \( r_{GF} = 20 \) for half-size images, and \( r_{GF} = 40 \) for full-size images. All other parameters are the same as in [3]. For SGBM, the parameters recommended in Middlebury leaderboard are used. For LocalExp, the default V2 version is used [9].

A. COMPARISON OF DIFFERENT UPSAMPLING METHODS

To study the impact of disparity upsampling methods on the proposed pipeline, three upsampling methods, i.e., NEAREST, BILINEAR and SG [25], are tested. In this experiment, post-processing is applied on half-size disparity maps and full-size disparities are generated by upsampling the refined half-size disparity maps.

Table 1 presents the average error rates over all image pairs using different upsampling methods. It is observed that the simplest NEAREST method achieves the best performance for all stereo matching algorithms. Fig. 5 visualize the comparison of different upsampling methods. As shown in Fig. 5(a) and (b), the NEAREST method outperforms BILINEAR and SG at the edges of disparity maps. From Fig. 5(c)-(f), we can see that LINEAR and SG smooth the edges due to interpolation whereas the NEAREST method keeps the sharp edges. In real world, the discontinuity of disparity exists at the interface of foreground and background, which results in sharp transitions. This property is well preserved by NEAREST, whereas for the other two methods, the edges are smoothed.

Recalling that the proposed SDD operation uses the concept of super-pixel which converts four Bayer pixels to one.
TABLE 2. Results of different post-processing and upsampling sequences for 2-Pixel accuracy.

| Method      | GF nonocc | SGBM nonocc | LocalExp nonocc | GF all | SGBM all | LocalExp all |
|-------------|-----------|-------------|-----------------|--------|----------|--------------|
|             | H0F1      | H1F0        | H0F1            | H1F0   | H0F1     | H1F0         |
| NEAREST     | 14.56     | 14.54       | 10.46           | 10.53  | 10.36    | 10.25        |
| LINEAR      | 16.15     | 15.21       | 11.15           | 11.21  | 11.16    | 10.85        |
| SG [25]     | 15.88     | 17.03       | 10.70           | 11.10  | 10.71    | 10.53        |

Average error rate using bad 2.0 metric on non-occluded (nonocc) region and all region. H0F1 represents upsampling first and then post-processing. H1F0 represents post-processing first and then upsampling.

B. COMPARISON OF POST-PROCESSING SEQUENCES

In this section, we evaluate the impact of the sequence of post-processing and upsampling. Two different configurations of post-processing and upsampling sequences, i.e., a) upsampling first and perform post-processing on full-size disparity maps (denoted by H0F1) and b) post-processing on half-size disparity first and perform upsampling to generate full-size disparity maps (denoted by H1F0) are experimented.

Table 2 shows the error rates of different configurations of sequence of post-processing and upsampling methods. Note that in this experiment, three different upsampling methods in the previous subsection are evaluated. In general, no matter which upsampling method is used, the results of H0F1 and H1F0 are similar for the same stereo matching algorithm. More specifically, for NEAREST, the difference between two configurations is very small, which is less than 0.2%.

Therefore, according to our experiments, the sequence of post-processing has negligible effect on final results. Considering the computation complexity of post-processing, which is $O(W \times H)$, it is beneficial to perform post-processing on half-size disparity maps before upsampling for saving computation.

C. EFFECT OF COLOR ARTIFACTS

In the proposed end-to-end stereo matching pipeline, demosaicing and downsampling are merged into a single step, where artifacts [19] may be introduced, especially at the high frequency regions. As shown in Fig. 6(a)-(d), although the proposed SDD operation is based on the color difference constancy assumption, there are color artifacts around the edges, where the assumption may fail. To evaluate the effect of these color artifacts, we generate half-size images using
two different methods: a) scale the original full-size sRGB images generated by a conventional ISP pipeline to half-size (ordinary demosaicing, denoted by OD) and b) demosaic and downsample the Bayer images by the proposed SDD and perform all other ISP stages (approximate demosaicing, denoted by AD). Stereo matching algorithms are applied on these half-size images and disparity upsampling are perform to generated the full-size disparity maps.

Table 3 shows the comparison results of the OD and AD-based disparity maps. In general, OD performs a little bit better than AD, but the differences are negligible (less than 1%). Fig. 6(e)-(f) illustrate the locations where differences exist between OD and AD-based results. As we can see, at the high frequency regions, OD does sometimes perform better than AD, whereas we also note that the AD also has advantages over OD in the same region. In other words, although AD introduces some color artifacts, it also brings some positive effects, and it is hard to distinguish them clearly. In summary, AD does not have significantly adverse impacts on the final results, whereas tremendous computations can be saved.

### D. COMPARISON OF DIFFERENT PIPELINES

In this section, we evaluate stereo matching performance based on the conventional ISP pipeline, minimal ISP pipeline (proposed in [15]) and the proposed end-to-end pipeline. The conventional pipeline computes the disparity directly on full-size sRGB images. The minimal pipeline only applies denoising, demosaicing and tone mapping on the Bayer pattern images to generate the input images.

Table 5 presents the performance of different pipelines using three different stereo matching algorithms. In general, all the three pipelines have similar stereo matching performance. It is observed that the proposed pipeline outperforms the minimal pipeline in GF and LocalExp. Compared with the conventional pipeline, the proposed pipeline produces slightly worse results with the biggest difference less than 1%. In some cases, such as LocalExp without post-processing, the proposed pipeline even achieves the best performance. That means in this case, better results can be achieved based on low-resolution images for LocalExp. But note that low-resolution is not always a benefit. In some cases, e.g., the resolution is extremely low, images needs to be upsampled for better stereo matching performance [30], [31].

It is interesting to note that the proposed pipeline achieves relatively better results when no post-processing is applied for SGBM and LocalExp. In other words, as illustrated in Fig. 7, although the results of the proposed pipeline are improved by post-processing, the improvement is not as much as that for the conventional pipeline. It is also found that the proposed pipeline has relatively poor results for Laundary, which has a lot of textureless areas with the same color. A possible reason is that the parameters of the stereo matching algorithms are not suitable for that since the proposed pipeline skips white balance and color transformation. But in general, the proposed pipeline achieves comparable results to the conventional pipeline, and even returns the lowest error in some cases.

### E. COMPARISON OF ISP COMPLEXITY

To profile the running time of major ISP steps, we implement different ISP pipelines in Python. For a fair comparison, a 768 × 512 Bayer image is processed by different ISP
FIGURE 7. Results for non-occluded regions. From left to right corresponding to GF, SGBM, LocalExp.

pipelines on the same workstation with Intel Xeon E5-2690 CPU and 64GB memory. The demosaicng method of directionally weighted gradient based interpolation [32] is used in the full pipeline, and the nearest-neighbor (NN) algorithm proposed in [15] is used in minimal ISP. Bilateral filter is applied for denoising in all ISP pipelines. The profiling results are shown in Table 5. It is obvious that the time consumption of ISP in the proposed pipeline is much less than the full ISP in the conventional pipeline, because it skips unnecessary steps compared with full ISP. Moreover, the proposed SDD step can substitute the domosaicing stage in the minimal ISP and downsample the image simultaneously, such that the tone mapping and denoising steps can run on half size images, which saves a lot of computation time compared with the other two pipelines. Therefore, the running time of the proposed ISP pipeline only accounts for 20.4% of the full ISP.

In summary, although the proposed pipeline skips most of the ISP stages and lowers the resolution of the input images to stereo matching algorithms, it achieves almost the same results as that based on the conventional pipeline. From the system perspective, the skipped ISP stages along with the lowered resolution of input images are beneficial for the processing speed and power consumption of the stereo matching systems. Furthermore, if the NEAREST method is used for disparity upsampling, the computation overhead introduced by the proposed pipeline is basically negligible.

V. CONCLUSION
In this paper, we propose an end-to-end pipeline for stereo matching from Bayer images to disparity maps. Unlike conventional pipeline that generate sRGB images using a series of ISP stages, the proposed pipeline skips most unnecessary ISP stages. The Bayer pattern images are converted to half-size sRGB images using a single SDD operation. Since stereo matching are performed on half-size input pairs and most of the ISP steps are skipped, the computational complexity of the proposed stereo matching pipeline is significantly less than that of the conventional pipeline. It is shown in the experiments, the proposed pipeline adapts to existing stereo matching algorithms and achieves almost the same results as conventional approach.

A complete stereo matching system consists of three parts, which are two camera sensors, two ISP processors and a stereo matching processor (either application specific integrated circuits or general purpose processor). Nowadays, a typical mobile camera sensor consume power ranging from 137.1 mW to 338.6 mW [33] and an ISP processor consumes about 250 mW for processing 1080 videos at 60fps [34] and a stereo matching algorithm chip (based on SGM) consumes 836 mW for 1080P video at 30FPS [35]. It can be found that the two ISP processors consume a significant amount of power in a stereo matching system. By adopting simple ISP that skips unnecessary stages and applies Bayer subsampling without complex signal processing, the proposed pipeline can save most of the power in ISP processors. In addition, the stereo matching algorithms and post-processing are performed at half-size, which can save about three-quarter of the computation compared with traditional and minimal pipeline. This is also beneficial for power saving. Although extra disparity upsampling is needed in the proposed pipeline, the computation overhead of it is negligible. Therefore, the proposed pipeline is suitable for embedded applications where power is the main restricting factor.

REFERENCES
[1] S. Matteocci, S. Giardino, and A. Gambini, “Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering,” in Proc. ACCV. Berlin, Germany: Springer, 2009, pp. 371–380.
[2] L. Xu, O. C. Au, W. Sun, L. Fang, K. Tang, J. Li, and Y. Guo, “Stereo matching by adaptive weighting selection based cost aggregation,” in Proc. ISCAS, 2013, pp. 1420–1423.
[3] P. Tan and P. Monasse, “Stereo disparity through cost aggregation with guided filter,” Image Process. Line., vol. 4, pp. 252–275, Oct. 2014. [Online]. Available: http://www.ipol.im/pub/art/2014/78/
[4] Y. Qian, M. Gong, and Y.-H. Yang, “Stereo-based 3D reconstruction of dynamic fluid surfaces by global optimization,” in Proc. CVPR, 2017, pp. 1269–1278.
[5] L. Li, S. Zhang, X. Yu, and L. Zhang, “PMSC: PatchMatch-based superpixel cut for accurate stereo matching,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 679–692, Mar. 2018.

[6] C. Olsson, J. Ulen, and Y. Boykov, “In defense of 3D-label stereo,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1730–1737.

[7] 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.

[8] D. Scharstein, T. Taniai, and S. N. Sinha, “Semi-global stereo matching with surface orientation priors,” in *Proc. Int. Conf. 3D Vis. (3DV)*, Oct. 2017, pp. 215–224.

[9] T. Taniai, Y. Matsushita, Y. Sato, and T. Naemura, “Continuous 3D label stereo matching using local expansion moves,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 11, pp. 2725–2739, Nov. 2018.

[10] J. Zbontar and Y. LeCun, “Stereo matching by training a convolutional neural network to compare image patches,” *J. Mach. Learn. Res.*, vol. 17, nos. 1–32, p. 2, 2016.

[11] J.-R. Chang and Y.-S. Chen, “Pyramid stereo matching network,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5410–5418.

[12] Middlebury Stereo Evaluation-Version 3. Accessed: Jan. 23, 2021. [Online]. Available: https://vision.middlebury.edu/stereo/eval3/

[13] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, “Color image processing pipeline,” *IEEE Signal Process. Mag.*, vol. 22, no. 1, pp. 34–43, Jan. 2005.

[14] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pająk, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, and K. Pulli, “FlexISP: A flexible camera image processing framework,” *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–13, Nov. 2014.

[15] M. Buckler, S. Jayasuriya, and A. Sampson, “Reconfiguring the imaging pipeline for computer vision,” in *Proc. IEEE Int. Conf. Comput. Vis.* (ICCV), Oct. 2017, pp. 975–984.

[16] W. Zhou, S. Gao, L. Zhang, and X. Lou, “Histogram of oriented gradients feature extraction from raw bayer pattern images,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 5, pp. 946–950, May 2020.

[17] B. Magnier, A. Aberkane, and N. Gorriti, “A recursive edge detector for color filter array image,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 2543–2547.

[18] C.-T. Wu, L. F. Isikoglan, S. Rao, B. Nayak, T. Gerasimow, A. Sutic, L. Ain-Kedem, and G. Michael, “VisionISP: Repurposing the image signal processor for computer vision applications,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 4624–4628.

[19] O. Losson, L. Macaire, and Y. Yang, *Comparison of Color Demosaicing Methods*, vol. 162, Amsterdam, The Netherlands: Elsevier, 2010, pp. 173–265.

[20] H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun, “A survey on deep learning techniques for stereo-based depth estimation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Oct. 20, 2020, doi: 10.1109/TPAMI.2020.3032602.

[21] C. LeGendre, K. Batsos, and P. Mordohai, “High-resolution stereo matching based on sampled photoconsistency computation,” in *Proc. BMVC*, 2017, pp. 1–13.

[22] B. Tippett, D. J. Lee, K. Lillywhite, and J. Archibald, “Review of stereo vision algorithms and their suitability for resource-limited systems,” *J. Real-Time Image Process.*, vol. 11, no. 1, pp. 5–25, Jan. 2016.

[23] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, “Joint bilateral upsampling,” in *Proc. ACM SIGGRAPH*, vol. 26, 2007, p. 96.

[24] D. Forsl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, “Image guided depth upsampling using anisotropic total generalized variation,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 993–1000.

[25] Y. Konno, M. Tanaka, M. Okutomi, Y. Yanagawa, K. Kinoshita, and M. Kawade, “Depth map upsampling by self-guided residual interpolation,” in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 1394–1399.

[26] Z. Liu, T. Park, H. Park, and N. S. Kim, “Ultra-low-power image signal processor for smart camera applications,” *Electron. Lett.*, vol. 51, no. 22, pp. 1778–1780, 2015.

[27] Y. Chen, J. U. Kang, G. Zhang, J. Cao, Q. Xie, and C. Kwan, “Real-time infrared image detail enhancement based on fast guided image filter and plateau equalization,” *Appl. Opt.*, vol. 59, no. 21, pp. 6407–6416, 2020.

[28] D. Scharstein and R. Szeliski, “High-accuracy stereo depth maps using structured light,” in *Proc. CVPR*, vol. 1, 2003, pp. 195–202.

[29] H. Hirschmuller and D. Scharstein, “Evaluation of stereo matching costs on images with radiometric differences,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1582–1599, Sep. 2009.

[30] B. Ayhan and C. Kwan, “Mastcam image resolution enhancement with application to disparity map generation for stereo images with different resolutions,” *Sensors*, vol. 19, no. 16, p. 3526, Aug. 2019.

[31] C. Kwan, B. Chou, and B. Ayhan, “Enhancing stereo image formation and depth map estimation for mastcam images,” in *Proc. UEMCON*, 2018, pp. 566–572.

[32] R. A. Maschul, Jr., S. S. Young, J. Reynolds, K. Krapels, J. Fanning, and T. Corbin, “Review of Bayer pattern color filter array (CFA) demosaicking with new quality assessment algorithms,” Army Res. Lab Adelphi Md Sennors Electron Devices Directorate, Adelphi, MD, USA, Tech. Rep. ARL-TR-5061, 2010.

[33] R. LiKamWa, B. Priyantha, M. Philipose, L. Zhong, and P. Bahl, “Energy characterization and optimization of image sensing toward continuous mobile vision,” in *Proc. MobiSys*, 2013, pp. 69–82.

[34] J. Hegarty, J. Brunhaer, Z. DeVito, J. Ragan-Kelley, N. Cohen, S. Bell, A. Vasilyev, M. Horowitz, and P. Hanrahan, “Darkroom: Compiling high-level image processing code into hardware pipelines,” *Proc. ACM SIGGRAPH*, vol. 33, no. 4, pp. 1–144, 2014.

[35] Z. Li, Q. Dong, M. Saligane, B. Kempe, L. Gong, Z. Zhang, R. Dreslinski, D. Sylvester, D. Blaauw, and H. Kim, “A 1920×1080 30-frames/s 2.3 TOPS/W stereo-depth processor for energy-efficient autonomous navigation of micro aerial vehicles,” *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 76–90, Sep. 2018.