

# Stereo Matching with Color and Monochrome Cameras in Low-light Conditions

Hae-Gon Jeon<sup>1</sup>   Joon-Young Lee<sup>2</sup>   Sunghoon Im<sup>1</sup>   Hyowon Ha<sup>1</sup>   In So Kweon<sup>1</sup>  
 hgjeon@rcv.kaist.ac.kr   jolee@adobe.com   shim@rcv.kaist.ac.kr   hwaha@rcv.kaist.ac.kr   iskweon@kaist.ac.kr

<sup>1</sup> Robotics and Computer Vision Lab., KAIST   <sup>2</sup> Adobe Research

## Abstract

Consumer devices with stereo cameras have become popular because of their low-cost depth sensing capability. However, those systems usually suffer from low imaging quality and inaccurate depth acquisition under low-light conditions. To address the problem, we present a new stereo matching method with a color and monochrome camera pair. We focus on the fundamental trade-off that monochrome cameras have much better light-efficiency than color-filtered cameras. Our key ideas involve compensating for the radiometric difference between two cross-spectral images and taking full advantage of complementary data. Consequently, our method produces both an accurate depth map and high-quality images, which are applicable for various depth-aware image processing. Our method is evaluated using various datasets and the performance of our depth estimation consistently outperforms state-of-the-art methods.

## 1. Introduction

Stereo camera systems allow us to estimate depth information and have many advantages over active range sensors such as ToF (time-of-flight) cameras and laser scanners. Stereo cameras are cost-effective and can work in both indoor and outdoor environments. This is why they have been widely used in the computer vision and robotics fields for several decades. Recently, consumer devices having a stereo camera [2, 4] have been released for depth-aware image editing applications.

Despite the advantages, estimating accurate depth map in low-light conditions leads to severe image noise; thus is still challenging and limits the usefulness of stereo systems. Although a long exposure time or a flash light may alleviate the problem, they can induce other imaging problems such as motion blur or specular reflections [19]. To overcome these issues, multi-modal and multi-spectral imaging approaches such as a color and infrared camera pair [16] and cross-channel matching [26] have been proposed. However,

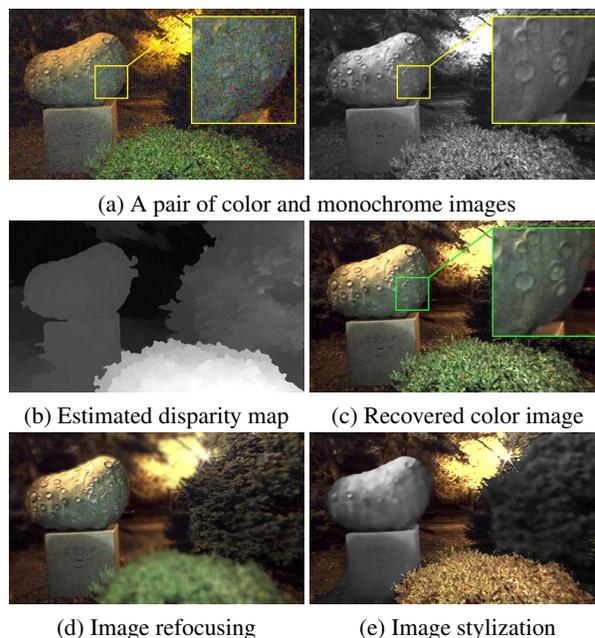


Figure 1. Given the pair of images (a), we estimate the accurate disparity map (b) and recover the high-quality color image (c). Our result is applicable to various depth-aware image processing (d, e).

these approaches require high manufacturing cost and specialized hardware.

In this paper, we present a stereo matching framework with a color and monochrome image pair (Fig. 1(a)). Our system is designed to estimate an accurate depth map under low-light conditions without additional light sources (Fig. 1(b)). In order to obtain reliable correspondence, we exploit the fundamental trade-off between color sensing capability and light efficiency of color cameras and monochrome cameras, respectively. Because monochrome cameras respond to all colors of light, they have much better light efficiency than Bayer-filtered color cameras [7, 3].

In general, image luminance recorded from a color camera is not consistent with that from a monochrome camera due to spatially-varying illumination and different spectral sensitivities of the cameras. This degrades the performance of stereo matching. To solve this problem,

we sample appropriate decolorization parameters of a color image and perform locally adaptive radiometric alignment and correspondences augmentation iteratively. After estimating a depth map, we recover a high-quality color image by colorizing the image from a light-efficient monochrome camera. For accurate colorization, we introduce a local chrominance consistency measure. Herein, we demonstrate the superior stereo matching performance over state-of-the-art methods. In addition, we show our result can be applied to depth-aware image processing algorithms such as image stylization and digital refocusing (Fig. 1(c)).

## 2. Related Work

Our method is related to cross-spectral stereo matching and colorization. Prior to introducing previous studies, we refer the reader to [14] for a comprehensive discussion of stereo matching with radiometric and noise variation.

Cross-spectral stereo matching has been studied extensively to find correspondence between multi-modal and color-inconsistent stereo images. Heo *et al.* [12] analyzed a color formation model and proposed an adaptive normalized cross correlation for stereo matching, that would be robust to various radiometric changes. This was extended in [13], which presented an iterative framework to simultaneously achieve both depth estimation and color consistency. Pinggera *et al.* [23] presented depth map estimation with cross-spectral stereo images, which uses dense gradient features based on the HOG descriptor [9]. Kim *et al.* [18] designed a dense descriptor for multi-modal correspondences by leveraging a measure of adaptive self-correlation and randomized receptive field pooling. Holloway *et al.* [15] proposed an assorted camera array and a cross-channel point correspondence measure using normalized gradient cost.

Colorization is a process of adding color channels to a grayscale image and video. Levin *et al.* [21] presented a user-guided colorization method that takes partial color information from user scribbles and automatically propagates the given seed color to make a complete color image. Yatziv and Sapiro [30] proposed a fast colorization method using the geodesic distance between neighboring pixels. Gastal and Oliveira [10] introduced an edge-aware filter in a transformed domain and showed a colorization result as one of its applications. Irony *et al.* [17] proposed an example-based colorization based on an assumption that similarly textured regions have similar colors.

In this study, we focused on simultaneously reconstructing an accurate depth-map and a noise-free color image using a color + monochrome image pair. We achieve this by taking the advantage of our cross-spectral stereo system. We created a locally adaptive spectral alignment algorithm that allows us to estimate an accurate disparity map without complex optimizations for intensity

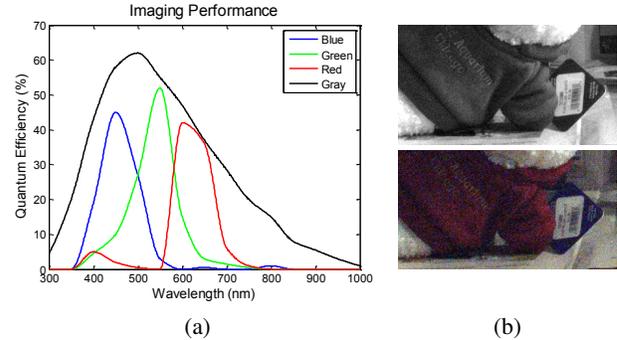


Figure 2. (a) Spectral sensitivity of the color and monochrome camera [1] used in our prototype stereo system. (b) An example image pair captured by our stereo system. Note that there is visible difference of image noise due to the gap of light-efficiency between two cameras.

equalization. As will be demonstrated in the experimental section (Sec. 7), our method is highly effective for accurate disparity estimation and significantly outperforms the state-of-the-art algorithms [12, 13, 18, 15]. In colorization, most approaches concentrate on propagating limited numbers of user-defined seeds, while we have lots of seed pixels with outliers around occlusion boundaries. To handle this issue, we introduced a new weighting term to correct inaccurate seed pixels and successfully recover a high-quality color image.

Recent work by [7] presented the concept of an alternative camera sensor that samples color information very sparsely. They recover a full color image by propagating the sparsely sampled colors into an entire image. This work shares the same philosophy with our work that takes the advantage of light-efficient monochrome sensors, but the concept may suffer from color noise that leads to an erroneous color image. Moreover, we adopt the idea in a stereo system and obtain an accurate depth-map and a noise-free color image simultaneously.

## 3. Stereo System with Color and Mono Cameras

Most color cameras use a color filter array called a Bayer array to capture color information. The Bayer array is positioned over the pixels of an image sensor and separates the incoming light into one of three primary colors (red, green, or blue) by filtering the light spectra according to wavelength range. This process is effective for capturing color information, but it amplifies image noise under low-light conditions because the array occludes a lot of incoming light. It may also reduce image sharpness as a result of using an anti-aliasing filter or optical low-pass filter to avoid aliasing or moiré artifacts during the demosaicing process.

Unlike color cameras, monochrome cameras receive all the incoming light at each pixel and need no demosaicing process. Therefore, they have much better light efficiency

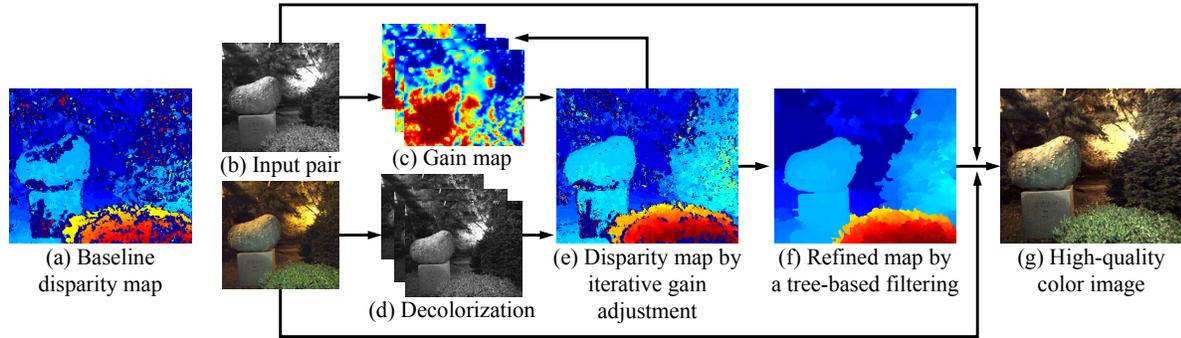


Figure 3. The overview of our algorithm. Our method produces the accurate depth map (f) and the high-quality color image (g), while the baseline stereo matching method with the luminance channels of the input images results in the poor disparity map (a).

and provide shaper images. In Fig. 2, we compare the imaging quality of a color and a monochrome camera. The comparison of spectral sensitivity (Fig. 2(a)) and the example image-pair captured under the same conditions (Fig. 2(b)) prove the large difference in light efficiency and image quality between the two types of cameras. That is, a color + monochrome camera pair is highly suitable for achieving a noise-free color image in addition to accurate depth estimation.

#### 4. Overview

Our stereo setup is a kind of cross-spectral system that includes a color and a monochrome (RGB-W) camera. In this section, we present an overview of our framework for the RGB-W stereo setup. Our key ideas are compensating for the spectral/radiometric difference between RGB-W images by locally adaptive radiometric alignment, and aggregating reliable correspondences by robust and noise-tolerant stereo matching. From these, we are able to utilize a disparity map to reconstruct high-quality color and depth images with applications to depth-aware image processing.

Fig. 3 depicts an overview of our framework. To account for spectral differences between a pair of RGB-W images, we first decolorize the color input image (Sec. 5.1). Because two cameras have different spectral sensitivities and viewpoints, there is no global mapping function that can explain the radiometric difference between two images without any assumptions. Moreover, estimation of local mapping functions is unstable and tends to diverge without reliable and dense correspondences. Instead, we use a candidate set of global decolorization functions that serve to preserve contrast distinctiveness and to suppress noise amplification simultaneously.

Then, we estimate disparities based on brightness constancy and edge similarity constraints, and each constraint term is designed to be robust to image noise and non-linear intensity changes, respectively (Sec. 5.2). Because the estimated disparity may contain outliers due to radiometric

difference or image noise, we retain reliable correspondences with a left-right consistency check and aggregate them from all candidate decolorized images.

After that, we have a set of reliable correspondences and use them to augment additional correspondences by iterative gain compensation and disparity estimation. Given the grayscale input and aligned decolorized image, we match the brightness of the input image to the decolorized image by estimating a local gain map (Sec. 5.3). Because our decolorization is performed to preserve the contrast distinctiveness of a color image, it can capture important local edges better than the grayscale input image, where edges may be ambiguous due to the mixing of spectral information. Therefore, this iterative process provides increases the number of reliable correspondences.

Last, we fuse the RGB-W stereo input image with the estimated disparity map and obtain a high-quality color image (Sec. 6). We show experimental validation and additional applications of our method in Sec. 7.

### 5. Stereo Matching with RGB-W Images

#### 5.1. Image Decolorization

A decolorization is a dimension-reduction process that converts three dimensional data to one dimension in the same range. Existing decolorization studies [25, 22] mostly considered contrast preservation and visual distinctiveness. We propose a decolorization method for our cross-spectral stereo setup by considering contrast preservation and noise suppression together.

We assume that a decolorized image  $I^\gamma$  is constructed by the weighted sum of three color channels of an input color image  $I$  as:

$$I^\gamma = \omega_r I_r + \omega_g I_g + \omega_b I_b$$

$$\text{s.t. } \omega_r + \omega_g + \omega_b = 1, \omega_r \geq 0, \omega_g \geq 0, \omega_b \geq 0, \quad (1)$$

where  $I_r$ ,  $I_g$  and  $I_b$  are three color channels, and  $\omega_r$ ,  $\omega_g$  and  $\omega_b$  are their weighting parameters.

We discretize each of  $\omega_r$ ,  $\omega_g$  and  $\omega_b$  with an interval



(a) Gradient sparsity by  $L_1$  norm (b) Normalized sparsity

Figure 4. Decolorized images with two different gradient sparsity measures.

of 0.1, and make 64 candidate parameter sets  $\Gamma_n$  where  $n \in \{1, 2, \dots, 64\}$ , because finer discretization produces indistinguishable differences in output for most cases, as shown in [25]. Then, we choose a set of appropriate parameters with contrast preservation and noise suppression constraints for accurate stereo matching.

**Contrast Preservation.** Contrast preservation is a key property in decolorization because it is effective for preserving rich color information and for reducing the perceptual difference between color and decolorized images. To estimate good decolorization parameters, we adopt the contrast preserving measure in [25], as follows.

To measure the contrast difference between a color image  $I$  and its decolorized image  $I^\gamma$ , which is also robust to image noise, we reconstruct a color image  $\tilde{I}^\gamma$  having contrast of  $I^\gamma$  by applying the guided filter [11] at each pixel  $i$  as:

$$\tilde{I}_i^\gamma = G_i(I, I^\gamma) = \sum_j \frac{1}{|\Omega|^2} w_{ij}(I^\gamma) I_j, \quad (2)$$

$$w_{ij}(I^\gamma) = \sum_{k|(i,j) \in \Omega_k} \left( 1 + \frac{(I_i^\gamma - \mu_k)(I_j^\gamma - \mu_k)}{\sigma_k^2 + \epsilon} \right),$$

where  $|\Omega|$  is the number of pixels in  $\Omega_k$ ,  $\epsilon$  is a regularization parameter,  $\mu_k$  and  $\sigma_k$  are the mean and standard deviation of  $I^\gamma$  in a  $5 \times 5$  window  $\Omega_k$  centered at the pixel  $k$  respectively.

Then, we measure a contrast preserving cost  $E_c(\gamma)$  for each decolorization parameter  $\gamma$  as:

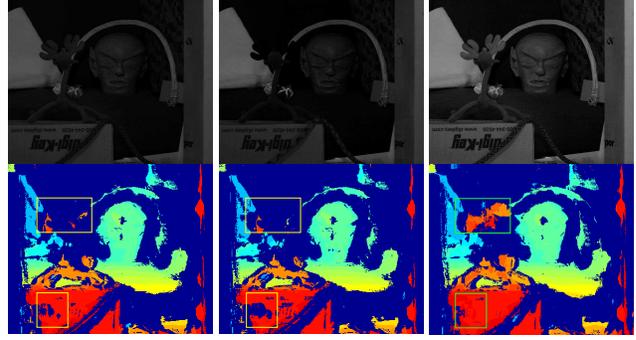
$$E_c(\gamma) = \|G(I, I) - \tilde{I}^\gamma\|_1, \quad \forall \gamma \in \Gamma, \quad (3)$$

where  $G(I, I)$  is the guided output image of the color input image  $I$  with a guidance of itself.

After computing the cost  $E_c(\gamma)$ , we linearly interpolate the scattered data in the  $\Gamma$  space and find the local minima from the interpolated cost map. We denote the set of weighting parameters at the local minima as  $\Gamma_1$ .

**Noise Suppression.** Because we consider low-light conditions where images suffer from large noise [8], the contrast preserving measure is not enough for producing properly decolorized images. As a complementary measure, we adopt the normalized sparsity measure, which was originally proposed for motion deblurring [20].

We use the normalized sparsity measure to estimate noise amplification during the decolorization process. It is



(a) rgb2gray (b) Only contrast (c) Proposed

Figure 5. Results of stereo matching between a gray reference image and decolorized images with three different methods: (a) the “rgb2gray” in MATLAB, (b) the contrast preserving measure in Eq. (3), and (c) the proposed measure. A detailed description for the dataset is presented in Sec. 7.1.

defined as:

$$E_n(\gamma) = \frac{\|\nabla_x I^\gamma\|_1 + \|\nabla_y I^\gamma\|_1}{\|\nabla_x I^\gamma\|_2 + \|\nabla_y I^\gamma\|_2}, \quad \forall \gamma \in \Gamma, \quad (4)$$

where  $I^\gamma$  is the decolorized image with a parameter  $\gamma$ ,  $\nabla$  is a gradient magnitude of  $x$  or  $y$  direction, and  $\|\cdot\|_1$  and  $\|\cdot\|_2$  are the  $L_1$  and  $L_2$  norm, respectively.

The normalized sparsity measure computes the normalized  $L_1$  norm of the image gradient which makes it scale-invariant, while the conventional  $L_1$  norm of gradient, widely used in denoising [27], imposes signal sparsity and is scale-variant that can be simply minimized by reducing the entire signal. We denote the set of decolorization parameters resulting in low normalized sparsity values in Eq. (4) as  $\Gamma_2$ . We empirically took the 20 percent of parameter subset in  $\Gamma$  as  $\Gamma_2$  for this paper.

Fig. 4 shows the comparison of two sparsity measures. While the  $L_1$  norm fails to select a good parameter because it favors low-intensity images, the normalized sparsity chooses a proper decolorized image that adjusts a good balance between signal power and image noise.

**Decolorization parameters.** Based on the two measures,  $E_c$  and  $E_n$ , we determine the candidate set of decolorization parameters  $\Gamma_d$  as the intersection of  $\Gamma_1$  and  $\Gamma_2$  ( $\Gamma_d = \Gamma_1 \cap \Gamma_2$ ). In our experiment, usually 5~7 decolorization parameters were selected through this process.

To validate the effectiveness of our decolorization process, we perform stereo matching without post-processing in Fig. 5. The decolorized image with our parameter produces a greater number of reliable correspondences than those with the baselines.

## 5.2. Disparity Estimation

In stereo matching, the brightness constancy assumption shows promising results in the presence of strong image

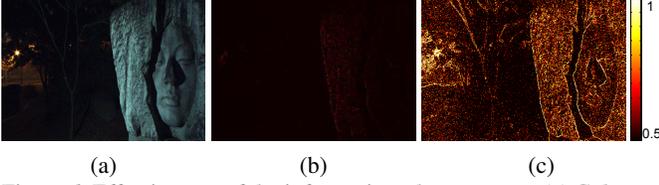


Figure 6. Effectiveness of the informative edge measure. (a) Color image. (b) Conventional gradient map. (c) Informative edge map.

noise because summing over a patch acts as a mean filter. Edge similarity assumption works well for the case of nonlinear intensity variations due to camera gain and changes in gamma [14].

To achieve robust stereo matching results, we combine two complementary costs; the sum of absolute differences (SAD) as a brightness constancy measure and the sum of informative edges (SIE) as an edge similarity measure. Our cost volume  $\mathcal{V}$  at pixel  $x$  is defined as:

$$\mathcal{V}(x, l) = \alpha \mathcal{V}_{SAD}(x, l) + (1 - \alpha) \mathcal{V}_{SIE}(x, l), \quad (5)$$

where  $l$  represents a cost label and  $\alpha \in [0, 1]$  is a balancing parameter between a brightness constancy term  $\mathcal{V}_{SAD}$  and an informative edges term  $\mathcal{V}_{SIE}$ .

The brightness constancy term  $\mathcal{V}_{SAD}$  is defined as:

$$\mathcal{V}_{SAD}(x, l) = \sum_{x \in \Omega_x} \min(|I_L(x) - I_R^\gamma(x + d)|, \tau_1), \quad (6)$$

where  $\Omega_x$  is a  $7 \times 7$  support window centered at  $x$ ,  $I_L$  is a monochrome input image,  $I_R^\gamma$  is a decolorized image from a color input image,  $d$  is a disparity, and  $\tau_1$  is a truncation value for robustness.

The informative edges term  $\mathcal{V}_{SIE}$  is defined as:

$$\mathcal{V}_{SIE}(x, l) = \sum_{x \in \Omega_x} \min(|J(I_L(x)) - J(I_R^\gamma(x + d))|, \tau_2),$$

$$\text{s.t. } J(I) = \frac{|\sum_{x \in \Omega_x} \nabla I(x)|}{\sum_{x \in \Omega_x} |\nabla I(x)| + 0.5}, \quad (7)$$

where  $\tau_2$  is a truncation value. We adopt  $J$  as a criterion to represent informative edges, which was introduced in deblurring research [28]. In the definition of  $J$ , the sum of signed gradients in the numerator cancels out image noise, while the sum of absolute gradient magnitudes in the denominator, computes how strong the edges are, around a pixel location. The constant value (0.5) prevents production of a large edge response in homogeneous regions. The informative edge response  $J$  always results in a normalized value in the range of  $[0, 1]$ , and is robust to nonlinear intensity changes and image noise. Therefore, we use Eq. (7) as our edge similarity measure. Fig. 6 shows the effectiveness of the informative edge measure. While the conventional gradient map in Fig. 6(b) fails to detect distinctive edge responses, the informative edge map in Fig. 6(c) captures important edge responses even under low-light conditions.

As a sequential step, we refine every cost slice in Eq. (5) by applying an edge-preserving filter that aggregates labels over the monochrome guidance image  $I_L$  [24]. Then we determine a disparity map using the winner-takes-all strategy. We reject outliers by the left-right consistency check, which marks pixels out if the disparity of a pixel on the left view is not consistent with the disparity of the corresponding pixel on the right view. We add such reliable correspondences by iteratively performing disparity estimation and gain adjustment that will be explained in the following section, and finally optimize the disparity map using minimum spanning tree-based filtering [29].

### 5.3. Gain Adjustment

In order to locally match the intensity levels between  $I_L$  and  $I_R^\gamma$ , we estimate a local gain map that adjusts brightness of the monochrome input image  $I_L$ . We achieve this by solving a constrained linear least-squares problem.

Given two images,  $I_L$  and  $I_R^\gamma$ , we first divide  $I_L$  into unit blocks of  $15 \times 20$  pixels where each block is assumed to be a uniform disparity. Each block is assigned one gain value, therefore we need to estimate a gain map  $\Pi = \{\pi_1, \pi_2, \dots, \pi_{n_b}\}$  where  $n_b$  is the number of blocks. The gain map is computed by solving the problem:

$$\underset{\pi_\nu^t}{\operatorname{argmin}} \sum_{\nu} \delta_\nu(I_{RL}^\gamma) (\pi_\nu^t \beta_\nu(I_L) - \beta_\nu(I_{RL}^\gamma))^2$$

$$\text{s.t. } 0.8\pi_\nu^{t-1} \leq \pi_\nu^t \leq 1.2\pi_\nu^{t-1}, \quad \pi_\nu^0 = \frac{\mu(\beta_\nu(I_{RL}^\gamma))}{\mu(\beta_\nu(I_L))}, \quad (8)$$

where  $\nu$  is a block index,  $\beta_\nu$  is the set of intensities of correspondences in a block  $\nu$ ,  $I_{RL}^\gamma$  represents the aligned image of  $I_R^\gamma$  to  $I_L$  with its disparity, and  $t$  is an iteration index of the disparity estimation and gain adjustment loop.  $\delta_\nu(\cdot)$  is an indicator function activated when the number of correspondences in block  $\nu$  is larger than three. We use the indicator function to avoid over-fitting to outliers.

After solving Eq. (8), we estimate a dense and smooth gain map by propagating  $\Pi$  to the entire image  $I_L$  using the local affinity model [21].

## 6. High-quality Color Image Recovery

The next step of stereo matching is merging the RGB-W stereo image to recover a high-quality color image. This is straightforward because monochrome cameras have better light efficiency than color cameras as discussed in Sec. 3.

In this process, we use the  $YUV$  colorspace, which is composed of one luminance channel  $Y$  and two chrominance channels,  $U$  and  $V$ . We directly use the monochrome input image as the luminance channel of a recovered color image and reconstruct its color information by combining the chrominance channels of the color input image according to the estimated disparity.

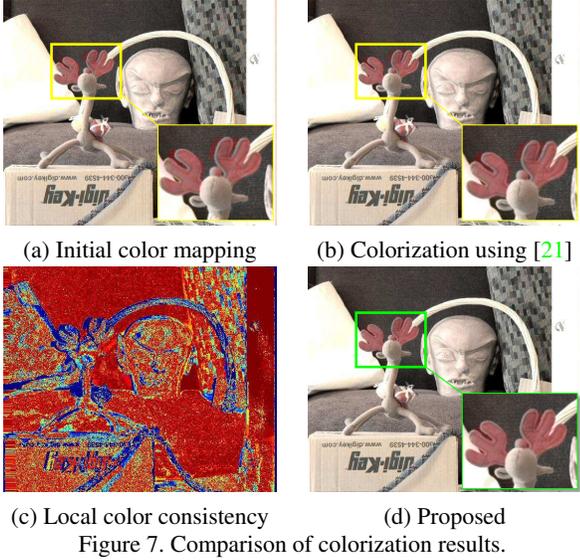


Figure 7. Comparison of colorization results.

Fig. 7(a) shows the reconstructed color image by this initial color mapping. It already shows a pretty good result thanks to our accurate disparity estimation, but there are color bleeding errors in occluded regions and conventional colorization [21] cannot handle the problem because the algorithm [21] is specialized for color propagation not for color correction (see Fig. 7(b)).

To resolve the problem, we introduce a simple but effective weight term and modify the algorithm [21] to correct color bleeding errors. We segment the luminance channel into super-pixels [6] and compute the confidence of initial chrominance mapping at the pixel  $i$  as:

$$w_i^d = \exp\left(\sum_{C \in \{U, V\}} -\frac{(C_i - \text{median}_S(C_i))^2}{2\sigma_C^2}\right), \quad (9)$$

where  $\text{median}_S(C_i)$  is the median chrominance of a super pixel containing the pixel  $i$  and  $\sigma_C$  is a control parameter. Fig. 7(c) shows the confidence map computed from Fig. 7(a).

The confidence value is used as an additional weighting term of the colorization method [21]. Specifically, we recover color-corrected chrominance channels by minimizing an objective function defined as:

$$\begin{aligned} \operatorname{argmin}_{\hat{C} \in \{\hat{U}, \hat{V}\}} \sum_i w_i^d \left(\hat{C}_i - C_i\right)^2 + \lambda_s \sum_i \sum_{j \in N_i} \left(\hat{C}_i - \frac{w_{ij}^s}{W} \hat{C}_j\right)^2, \\ \text{s.t. } w_{ij}^s = \exp\left(-\frac{(Y_i - Y_j)^2}{2\sigma_{N_i}^2}\right), \quad W = \sum_{j \in N_i} w_{ij}^s, \end{aligned} \quad (10)$$

where  $\lambda_s$  is a balancing parameter between the data term and the smoothness term,  $N_i$  represents the eight neighboring pixels of  $i$ , and  $\sigma_{N_i}^2$  is the variance of the luminances in  $N_i$ . Following [21], this objective function can be

	color camera			monochrome camera		
	illum.	exp.	noise std.	illum.	exp.	noise std.
Setup 1	1	0	$0.03\sqrt{\kappa}$	2	1	$0.01\sqrt{\kappa}$
Setup 2	3	0	$0.07\sqrt{\kappa}$	1	2	$0.01\sqrt{\kappa}$

Table 1. Two setups of the Middlebury stereo benchmark. We simulate our stereo system by taking a image of a given illumination and exposure level from the Middlebury dataset. We add additional signal dependent Gaussian noise with a given standard deviation where  $\kappa$  represents the noise-free signal intensity [5]

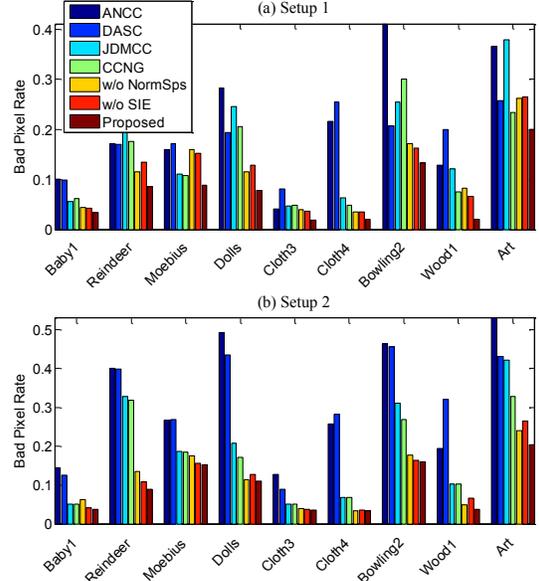


Figure 9. Quantitative evaluation results on the Middlebury stereo benchmark. The experimental setup is summarized in Table 1.

efficiently solved as:

$$(W^d + \lambda_s L)\hat{\mathbf{c}} = W^d \mathbf{c} \quad \text{s.t.} \quad L = \mathcal{I} - W^s, \quad (11)$$

where  $W^d$  is a diagonal matrix consisting of the data weight  $w_p^d$ ,  $L$  is a Laplacian matrix,  $\mathcal{I}$  is an identity matrix,  $W^s$  is a matrix form of the smoothness term, and  $\mathbf{c}$  and  $\hat{\mathbf{c}}$  are vectorized forms of  $C$  and  $\hat{C}$  respectively. Fig. 7(d) shows our colorization result in which color bleeding is recovered.

## 7. Experiments

We implemented our method in MATLAB and it takes about 5 minutes to process one dataset of  $1390 \times 1110$  resolution on an I7 3.4GHz machine. Among all the steps, disparity estimation in Sec. 5.2 uses most of the processing time. We expect that the computational time can be significantly reduced using GPU parallelization.

For the evaluation, we compare our method with state-of-the-art methods of multi-spectral or cross-channel stereo matching; DASC [18], CCNG [15], ANCC [12] and JDMCC [13]. For a fair comparison, we used the original

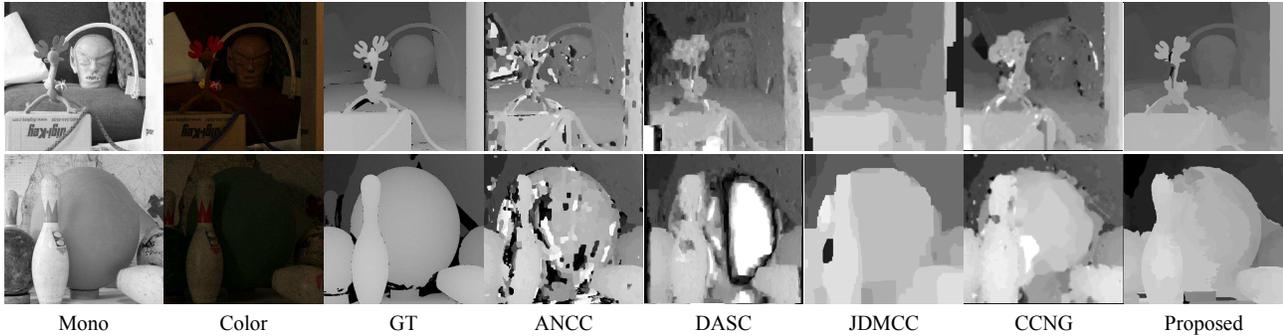


Figure 8. Comparison of estimated disparity maps under the “Setup 2” on the Middlebury Benchmark.

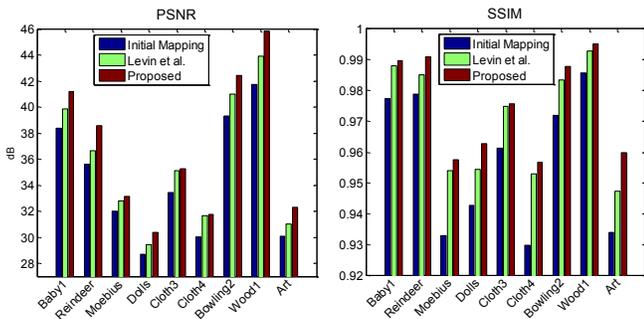


Figure 10. Evaluation for each of the colorization methods.

authors’ code and chose the best performing parameters using a parameter sweep. We used the same set of parameters  $\{\alpha, \tau_1, \tau_2, \sigma_C, \lambda_s\} = \{0.5, 0.1, 0.1, 0.01, 3\}$  to generate all the results of our method. Please refer to the supplementary material for more results and comparisons with state-of-the-art methods.

### 7.1. Middlebury Stereo Benchmark

We quantitatively evaluated our method using the Middlebury stereo benchmark [14]. For realistic simulations, we took two images captured under different illuminations to simulate different spectral sensitivities and add additional noise to simulate low-light conditions. To imitate the light-efficiency difference between color and monochrome cameras, we used longer exposure images as monochrome input images, and added more noise to the color input images. We configured two different setups for this experiment. The details are summarized in Table 1.

Two examples of stereo matching results are shown in Fig. 8. The quantitative comparison is presented in Fig. 9. We use the *bad pixel rate* as an evaluation criterion, which is defined as the percentage of pixels for which the absolute disparity error is greater than 1. In this experiment, our method largely outperformed all the competing methods for all the test datasets. The NCC (normalized cross correlation)-based methods [18, 15, 12] are vulnerable to low intensity level and severe noise Fig. 9(b), as demonstrated in [14]. The JDMCC [13] worked relatively well

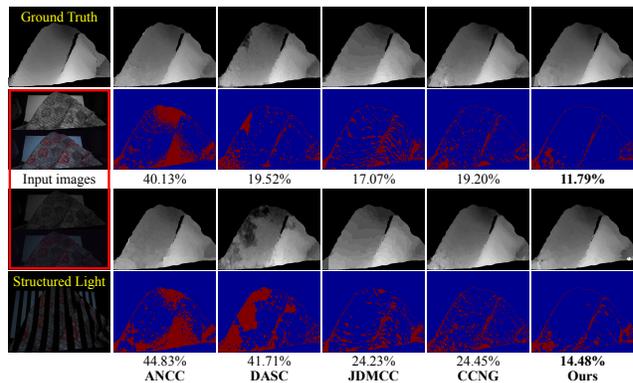


Figure 11. Quantitative evaluation using a controlled indoor scene captured from our prototype system. The odd rows show the results with image pairs captured under bright and dark illumination conditions, respectively. The even rows are corresponding bad pixel error maps.

among the competing methods; however, it exhibited large quantization errors, as shown in Fig. 8. We guess that the absence of color information leads to failure of color equalization of the JDMCC, and that this causes large errors in the JDMCC results.

We also evaluated the effectiveness of our colorization process. For the quantitative evaluation, we measure the PSNR and SSIM of three different colorization methods: initial color mapping with an estimated disparity map, conventional colorization [21], and our colorization with chrominance consistency weight. Fig. 10 shows the evaluation result, and that our method outperforms the competing methods and consistently improves colorization quality in terms of both PSNR and SSIM. This is because the newly proposed method corrects the color bleeding errors in out-of-plane regions.

### 7.2. Experiment with our Prototype System

We implemented our prototype system using two Point-Grey Flea3 cameras, one color and one monochrome camera, with baseline of 5cm, and maximum disparity of about 80 pixels. The stereo system was pre-calibrated and

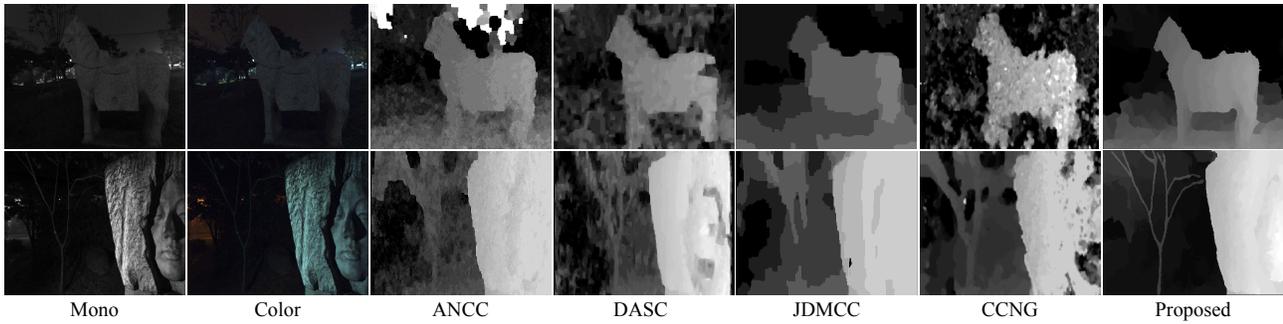


Figure 12. Comparison of disparity estimation results on outdoor scenes captured from our prototype system.



(a) Recovered image (b) Refocusing (c) Stylization  
 Figure 13. Applications to depth-aware image processing using the accurate disparity map and high-quality color image of our method. Note that the input images were captured in challenging low-light conditions and we amplify the brightness of the results for the visualization purpose.

images from the cameras were rectified using the MATLAB built-in camera calibration toolbox.

First, we investigated performance degradation with respect to illumination conditions. As shown in Fig. 11, we captured the same scene under two different illuminations in a controlled laboratory environment. For quantitative evaluation, we also estimated a ground-truth disparity map using a structured-light 3D scanner. Fig. 11 shows the comparison of the estimated disparity maps with their bad pixel rates. Compared to the competing methods, our method achieved the best results regardless of illumination conditions, with the least degradation of performance.

Fig. 12 shows the results of outdoor datasets captured at night (i.e., low-light conditions). All the state-of-the-art methods produced reasonable results; however, our method achieved the most accurate disparity map among them while the other methods suffered from holes and errors in dark regions. Note that our method reconstructs both depth discontinuities and fine structure, such as the horse stone statue in the 1<sup>st</sup> row, and the branches in the 2<sup>nd</sup> row of Fig. 12.

### 7.3. Applications to Depth-aware Image Processing

An accurate disparity map can facilitate many applications. As examples, we show photographic editing applica-

tions such as digital refocusing and image stylization.

Digital refocusing that shifts the in-focus region after taking a photo [2, 4] is one of the most popular depth-aware processing techniques. An accurate disparity map is necessary to create a realistic refocused image. In Fig. 13(b), we added synthetic blurs to the images using our disparity estimates and produced a shallow depth of field image.

Another emerging application is image stylization, which changes the photographic look of an image. When a disparity map is given, we can easily change the color of a certain depth range and produce visually pleasing photographic looks as shown in Fig. 13(c). Our application results show distinctively realistic photographic effects even in night scenes.

## 8. Conclusions

We have proposed a new stereo framework for high-quality depth and color image acquisition under low-light conditions. We achieved this by utilizing a fundamental trade-off of the advantages of color and monochrome cameras and validated the effectiveness of the proposed framework through extensive quantitative and qualitative evaluation. We expect that the proposed framework could become popular as a robust stereo system for mobile phones in the near future.

In this study, we found some challenges that should be overcome, and which will be considered in future work. First, the performance of our method is not guaranteed for datasets with low-texture and refractive media. This is considered a fundamental issue of stereo matching. Second, we need to account for a narrower baseline system than our prototype for practical utility in mobile devices. Last, the current computational burden is another problem to be solved.

**Acknowledgements.** This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No.2010- 0028680). Hae-Gon Jeon was partially supported by Global PH.D Fellowship Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2015H1A2A1034617).

## References

- [1] Flea3 gige imaging performance specification. <http://www.ptgrey.com/support/downloads/10109/>.
- [2] HTC One (m8). <http://www.htc.com/us/smartphones/htc-one-m8/>.
- [3] Huawei p8. <http://consumer.huawei.com/minisite/worldwide/p8/>.
- [4] Venue 8 7000 series. <http://www.dell.com/en-us/shop/productdetails/dell-venue-8-7840-tablet/>.
- [5] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Multiplexing for optimal lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(8):1339–1354, 2007.
- [6] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 34(11):2274–2282, 2012.
- [7] A. Chakrabarti, W. T. Freeman, and T. Zickler. Rethinking color cameras. In *Proceedings of IEEE International Conference on Computational Photography (ICCP)*, 2014.
- [8] P. Chatterjee, N. Joshi, S. B. Kang, and Y. Matsushita. Noise suppression in low-light images through joint denoising and demosaicing. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [10] E. S. Gastal and M. M. Oliveira. Domain transform for edge-aware image and video processing. In *ACM Transactions on Graphics (TOG)*, volume 30, page 69, 2011.
- [11] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(6):1397–1409, 2013.
- [12] Y. S. Heo, K. M. Lee, and S. U. Lee. Robust stereo matching using adaptive normalized cross-correlation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(4):807–822, 2011.
- [13] Y. S. Heo, K. M. Lee, and S. U. Lee. Joint depth map and color consistency estimation for stereo images with different illuminations and cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(5):1094–1106, 2013.
- [14] H. Hirschmüller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(9):1582–1599, 2009.
- [15] J. Holloway, K. Mitra, S. J. Koppal, and A. N. Veeraraghavan. Generalized assorted camera arrays: Robust cross-channel registration and applications. *IEEE Transactions on Image Processing (TIP)*, 24(3):823–835, 2015.
- [16] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [17] R. Irony, D. Cohen-Or, and D. Lischinski. Colorization by example. In *Eurographics Symp. on Rendering*, volume 2, 2005.
- [18] S. Kim, D. Min, B. Ham, S. Ryu, M. N. Do, and K. Sohn. Dasc: Dense adaptive self-correlation descriptor for multi-modal and multi-spectral correspondence. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [19] D. Krishnan and R. Fergus. Dark flash photography. In *ACM Transactions on Graphics (TOG)*, volume 28, page 96, 2009.
- [20] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [21] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 689–694, 2004.
- [22] C. Lu, L. Xu, and J. Jia. Contrast preserving decolorization with perception-based quality metrics. *International Journal of Computer Vision (IJCV)*, 110(2):222–239, 2014.
- [23] P. Pinggera, T. Breckon, and H. Bischof. On cross-spectral stereo matching using dense gradient features. In *Proceedings of British Machine Vision Conference (BMVC)*, 2012.
- [24] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [25] Y. Song, L. Bao, X. Xu, and Q. Yang. Decolorization: is rgb2gray () out? In *SIGGRAPH Asia Technical Briefs*, page 15, 2013.
- [26] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar. Picam: an ultra-thin high performance monolithic camera array. *ACM Transactions on Graphics (TOG)*, 32(6):166, 2013.
- [27] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.
- [28] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *Proceedings of European Conference on Computer Vision (ECCV)*. 2010.
- [29] Q. Yang. Stereo matching using tree filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37(4):834–846, 2015.
- [30] L. Yatziv and G. Sapiro. Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing (TIP)*, 15(5):1120–1129, 2006.