

Into the Twilight Zone: Depth Estimation using Joint Structure-Stereo Optimization

Aashish Sharma^[0000–0001–9211–3368] and Loong-Fah Cheong
aashish.sharma@u.nus.edu, eleclf@nus.edu.sg

Department of ECE, National University of Singapore

Abstract. We present a joint Structure-Stereo optimization model that is robust for disparity estimation under **low-light conditions**. Eschewing the traditional denoising approach – which we show to be ineffective for stereo due to its artefacts and the questionable use of the PSNR metric, we propose to instead rely on structures comprising of piecewise constant regions and principal edges in the given image, as these are the important regions for extracting disparity information. We also judiciously retain the coarser textures for stereo matching, discarding the finer textures as they are apt to be inextricably mixed with noise. This selection process in the structure-texture decomposition step is aided by the stereo matching constraint in our joint Structure-Stereo formulation. The resulting optimization problem is complex but we are able to decompose it into sub-problems that admit relatively standard solutions. Our experiments confirm that our joint model significantly outperforms the baseline methods on both synthetic and real noise datasets.

Keywords: Stereo matching, Depth estimation, Low-light vision, Structure extraction, Joint optimization

1 Introduction

Disparity estimation from stereo plays an imperative role in 3D reconstruction, which is useful for many real-world applications such as autonomous driving. In the past decade, with the development of fast and accurate methods[1][2] and especially with the advent of deep learning[3][4][5], there has been a significant improvement in the field. Despite this development, binocular depth estimation under low-light conditions still remains a relatively unexplored area. Presence of severe image noise, multiple moving light sources, varying glow and glare, unavailability of reliable low-light stereo datasets, are some of the numerous grim challenges that possibly explain the slow progress in this field. However, given its significance in autonomous driving, it becomes important to develop algorithms that can perform robust stereo matching under these conditions. Given that the challenges are manifold, we focus in this paper on the primary issue that plagues stereo matching under low-light: that images inevitably suffer from low contrast, loss of saturation, and substantial level of noise which is dense and often non-Gaussian[6]. The low signal to noise ratio (SNR) under low-light is in a sense unpreventable since the camera essentially acts like a gain-control amplifier.

While the aforementioned problem may be alleviated somewhat by using longer exposure time, this additionally causes other imperfections such as motion blur[7]. Multi-spectral imaging involving specialized hardware such as color-infrared or color-monochrome camera pair[7] can be used, but their usability is often restricted owing to high manufacturing and installation costs. Rather than relying on modifying the image acquisition process, our research interest is more that of coming to grips with the basic problems: how to recover adequate disparity information from a given pair of low-light stereo images under typical urban conditions, and to discover the crucial recipes for success.

One obvious way to **handle noise** could be to **use denoising** to clean up the images before stereo matching. However, denoising in itself either suffers from ineffectiveness in the higher noise regimes (e.g., NLM[8], ROF[9]), or creates undesirable artefacts (e.g., BM3D[10]), both of which are detrimental for stereo matching. Even some of the recent state-of-the-art deep learning solutions, such as MLP[11], SSDA[12] and DnCNN[13], only show equal or marginally better performances over BM3D[10] in terms of image **Peak Signal to Noise Ratio (PSNR)**. On the most basic level, these denoising algorithms are designed for a single image and thus may not remove noise in a manner that is consistent across the stereo pair, which is again detrimental for stereo matching. Another fundamental issue is raised by a recent paper “Dirty Pixels”[6] which demonstrated empirically that PSNR might not be a suitable criteria for evaluation of image quality if the aim is to perform high-level vision tasks such as classification, and even low PSNR images (but optimized for the vision task ahead) can outperform their high PSNR unoptimized counterparts. This debunks the general belief of a linear relationship between improving the PSNR and improving the competency of the associated vision task. We argue that the same phenomenon holds for the task of stereo matching, for which case we offer the following reasoning: unlike PSNR, in stereo matching, not all pixels are equal in terms of their impact arising from a denoising artefact. In image regions with near-uniform intensity, the energy landscape of the objective function for stereo matching is very shallow; any small artefacts caused by denoising algorithms in these regions can have a disproportionally large influence on the stereo solution. On the other hand, in textured regions, we can afford to discard some of the finer textures (thus losing out in PSNR) but yet suffer no loss in disparity accuracy, provided there are sufficient coarser textures in the same region to provide the necessary information for filling in. This latter condition is often met in outdoor images due to the well-known scale invariance properties of natural image statistics[14].

Our algorithm is founded upon the foregoing observations. **Our first key idea originates from how we humans perceive depth in low-light, which is mainly through the principal scene structures such as object boundaries and coarser textures.** The main underlying physiological explanation for the preceding is the increased spatiotemporal pooling of photoreceptor responses for increased sensitivity, under which low-light vision becomes necessarily coarser and slower. It means that for highly noisy images perturbed by randomly oriented elements, only the principal contours (i.e. lower spatial frequency contours) become salient

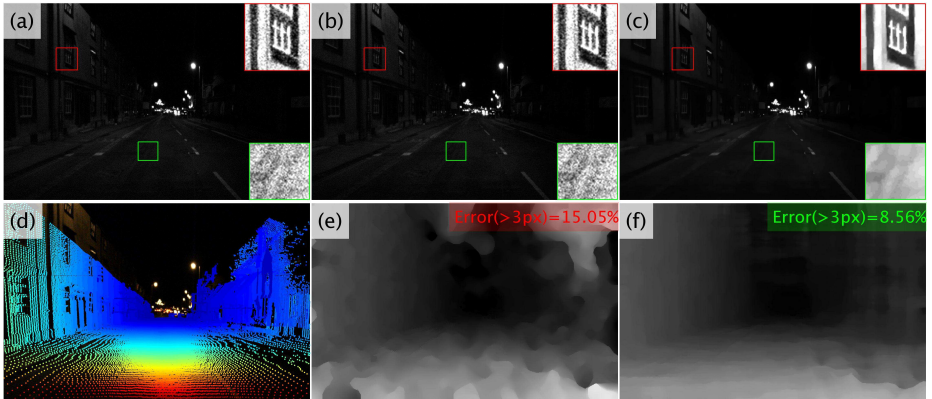


Fig. 1. (a) Sample low-light image from the Oxford dataset[15]. From the two patches (boosted with [16]), we can observe that in low-light, fine textures are barely distinguishable from dense noise, and only coarser textures and object boundaries are recoverable; (b) Denoising result from DnCNN[13] showing its ineffectiveness under low-contrast dense noise; (c) Structures from our model showing recovery of sharp object boundaries and coarse textures; (d) Image (a) with projected disparity ground truth (for visualization); (e) Disparity result from ‘DnCNN[13] + MS[17]’, (f) Disparity result from our model. Our result is more accurate, robust and has lesser artefacts, showing our model’s robustness for stereo matching under low-light conditions.

because their elements are coaligned with a smooth global trajectory, as described by the Gestalt law of good continuation. In an analogous manner, we postulate that since fine details in low-light are barely irrevocable from noise (e.g., the fine textures on the building and road in the inset boxes of Fig.1a), we should instead rely on structures consisting of piecewise constant regions and principal edges (from both object boundaries and coarse textures) to obtain scene depth (see the coarse textures extracted in the inset boxes of Fig.1c)¹. For this purpose, we adopt the nonlinear $TV - L_2$ decomposition algorithm[9] to perform both denoising and extraction of the principal structures². This variational style of denoising ensures that (1) the near-uniform intensity regions will remain flat, critical for disparity accuracy, and (2) those error-prone high-frequency fine details will be suppressed, whereas the coarser textures, which

¹ Most night-time outdoor and traffic lighting scenarios in a city are amidst such a wash of artificial lights that our eyes never fully transition to scotopic vision. Instead, they stay in the mesopic range, where both the cones and rods are active (mesopic light levels range from ~ 0.001 - 3 cd/m^2). This range of luminance where some coarse textures in the interiors of objects are still visible to the human eyes will occupy our main interest, whereas extremely impoverished conditions such as a moonless scene (where even coarse textures are not discernible) will be tangential to our enquiry.

² Note that we purposely avoid calling the $TV - L_2$ decomposition as structure-texture decomposition, since for our application, the term “structure” is always understood to contain the coarser textures (such as those in the inset boxes of Fig.1c).

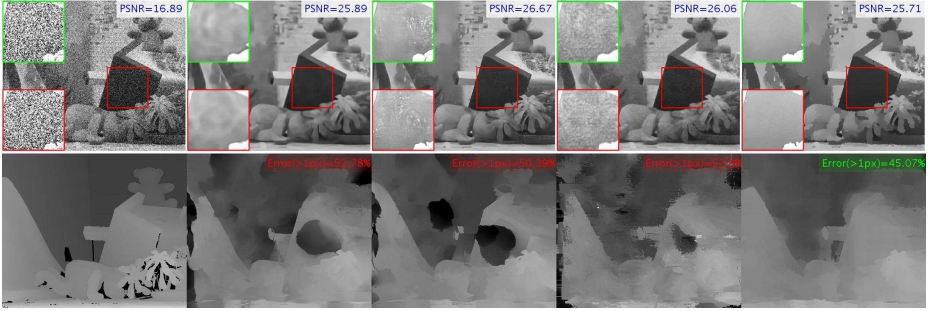


Fig. 2. Going column-wise: (i) Noisy ‘Teddy’[18] image with corresponding left-right (red-green) patches (boosted with [16]); Denoised with (ii) BM3D[10] (inconsistent artefacts across the patches); (iii) DnCNN[13] (inconsistent denoising), (iv) SS-PCA[19] (inconsistent and ineffective denoising); (v) Structures from our model (consistent and no artefacts); (vi) Disparity ground truth; Result from (vii) ‘BM3D[10]+MS[17]’, (viii) ‘DnCNN[13]+MS[17]’, (ix) SS-PCA[19], and (x) Our model. All the baseline methods show high error in the patch area, while our method produces more accurate result in there while keeping sharp edges in other parts. Also note that our structures have the lowest PSNR, but still the highest disparity performance among all the methods.

are more consistently recoverable across the images, will be retained. These attributes contribute significantly to the success of our disparity estimation (see results obtained by ‘DnCNN[13]+MS[17]’, Fig.1e and our algorithm, Fig.1f).

Our second key idea is to jointly optimize the $TV - L_2$ decomposition and the disparity estimation task. The motivation is twofold. Firstly, a careful use of $TV - L_2$ decomposition as a denoising step[9] is required since any denoising algorithm may not only remove the noise but also the useful texture information, leading to a delicate tradeoff. Indeed, without additional information, patch-based image denoising theory suggests that existing methods have practically converged to the theoretical bound of the achievable PSNR performance[20]. An additional boost in performance can be expected if we are given an alternative view and the disparity between these two images, since this allows us to take advantage of the self-similarity and redundancy of the adjacent frame. This depends on us knowing the disparity between the two images, and such dependency calls for a joint approach. In our joint formulation, the self-similarity constraint is captured by the well-known Brightness Constancy Constraint (BCC) and Gradient Constancy Constraint (GCC) terms appearing as coupling terms in the $TV - L_2$ decomposition sub-problem. The second motivation is equally important: by solving the $TV - L_2$ decomposition problem concurrently with the disparity estimation problem, we make sure that the denoising is done in a way that is consistent across the stereo pair (see Fig.2), that is, it is optimized for stereo disparity estimation rather than for some generic metric such as PSNR.

The joint formulation has significant computational ramifications. Our stereo matching cost for a pixel is aggregated over a window for increased robustness. This results in significant coupling of variables when we are solving the $TV - L_2$

decomposition sub-problem which means that the standard solutions for $TV - L_2$ are no longer applicable. We provide an alternative formulation such that the sub-problems still admit fairly standard solutions. We conduct experiments on our joint model to test our theories. We show that our model with its stereo-optimized structures, while yielding low PSNR, is still able to considerably surpass the baseline methods on both synthetic and real noise datasets. We then discuss some of the limitations of our algorithm, followed by a conclusion.

2 Related Work

As our paper is to specifically **solve the problem of stereo matching under noisy conditions**, we skip providing a comprehensive review of general stereo matching. Interested readers may refer to [21] and [22] for general stereo overview and stereo with radiometric variations respectively. Similarly, our work is not specifically concerned with denoising per se; readers may refer to [23] for a review in image denoising, and to [24] for some modern development in video denoising. Some works that target video denoising using stereo/flow correspondences include [25][26][27], but they are either limited by their requirement of large number of frames ([27]), or their dependency on pre-computed stereo/flow maps ([26]), which can be highly inaccurate for low SNR cases. [28] reviewed various structure-texture image decomposition models³, and related them to denoising.

The problem of stereo matching under low-light is non-trivial and challenging.

Despite its significance, only a few works can be found in the literature to have attempted this problem. To the best of our knowledge, there are only three related works[29][30][19] we could find till date. All the three works propose a joint framework of denoising and disparity, with some similarities and differences. They all propose to improve NLM[8] based denoising by finding more number of similar patches in the other image using disparity, and then improving disparity from the new denoised results. [29], [30] use an Euclidean based similarity metric which has been shown in [19] to be very ineffective in highly noisy conditions. Hence, the two methods perform poorly after a certain level of noise. [19] handles this problem by projecting the patches into a lower dimensional space using PCA, and also uses the same projected patches for computing the stereo matching cost.

Our work is more closely related to [19] in terms of **iterative joint optimization, but with a few key differences**. Firstly, we do not optimize PSNR to improve the stereo quality, which, as we have argued, might not have a simple relationship with PSNR. Secondly, we **rely on the coarse scale textures and object boundaries for guiding the stereo**, and not on NLM based denoising which might be ineffective in high noise. Thirdly, underpinning our joint Stereo-Structure optimization is a single global objective function that is mathematically consistent and physically well motivated, unlike the iterative denoising-disparity model proposed by [19] which has multiple components processed in sequence.

³ Among these models, we choose $TV - L_2$ based on the recommendations given in [28](Pg.18), which advocates it when no *a-priori* knowledge of the texture/noise pattern is given at hand, which is likely to be the case for real low-light scenes.

3 Joint Structure-Stereo Model

Let $I_{n1}, I_{n2} \in \mathbb{R}^{h \times w \times c}$ be respectively the two given rectified right-left noisy stereo images each of resolution $h \times w$ with c channels. Let $I_{s1}, I_{s2} \in \mathbb{R}^{h \times w \times c}$ be the underlying structures to obtain, and $D_2 \in \mathbb{Z}_{\geq 0}^{h \times w}$ be the disparity of the left view (note that we use $D_2 = 0$ to mark invalid/unknown disparity).

Our joint model integrates the two problems of structure extraction and stereo estimation into a single unified framework and takes the energy form:

$$E_{ALL}(I_{s1}, I_{s2}, D_2) = E_{StructureData}(I_{s1}, I_{s2}) + \lambda_S \cdot E_{StructureSmooth}(I_{s1}, I_{s2}) \\ + \lambda_{SD} \cdot E_{StereoData}(I_{s1}, I_{s2}, D_2) + \lambda_{SS} \cdot E_{StereoSmooth}(D_2) \quad (1)$$

where λ_{\times} are parameters controlling strengths of the individual terms. We then decompose the overall energy form Eq.(1) into two sub-problems and solve them alternately until convergence:

$$E_{Structure}(I_{s1}, I_{s2}, D_2^*) = E_{StructureData}(I_{s1}, I_{s2}) + \lambda_S \cdot E_{StructureSmooth}(I_{s1}, I_{s2}) \\ + \lambda_{SD} \cdot E_{StereoData}(I_{s1}, I_{s2}, D_2^*) \quad (2)$$

$$E_{Stereo}(I_{s1}^*, I_{s2}^*, D_2) = \lambda_{SD} \cdot E_{StereoData}(I_{s1}^*, I_{s2}^*, D_2) \\ + \lambda_{SS} \cdot E_{StereoSmooth}(D_2) \quad (3)$$

The superscript (*) represents that the variable is treated as a constant in the given sub-problem. Let us next describe the two sub-problems in Eqs.(2)(3) in detail, and then discuss their solutions and the joint optimization procedure.

3.1 Structure Sub-problem

The first two terms of $E_{Structure}$ in Eq.(2) represent the associated data and smoothness costs for TV regularization, and are defined as

$$E_{StructureData}(I_{s1}, I_{s2}) = \sum_p \left((I_{s1}(p) - I_{n1}(p))^2 + (I_{s2}(p) - I_{n2}(p))^2 \right) \quad (4)$$

$$E_{StructureSmooth}(I_{s1}, I_{s2}) = \sum_p (\mathbf{RTV}(I_{s1}(p)) + \mathbf{RTV}(I_{s2}(p))) \quad (5)$$

where $\mathbf{RTV}(\cdot)$ or Relative Total Variation introduced in [31] is a more robust formulation of the TV penalty function $|\nabla(\cdot)|$, and is defined as $\mathbf{RTV}(\cdot) =$

$\frac{\sum_{q \in N_p} g_{\sigma}(p, q) \cdot |\nabla(\cdot)|}{|\sum_{q \in N_p} g_{\sigma}(p, q) \cdot \nabla(\cdot)| + \epsilon_s}$ where N_p is a small fixed-size window around p , $g_{\sigma}(p, q)$ is a Gaussian weighing function parametrized by σ , and ϵ_s is a small value constant to avoid numerical overflow. For noisy regions or fine textures, the denominator term in $\mathbf{RTV}(\cdot)$ summing up noisy random gradients generates small values while the numerator summing up their absolute versions generates large values, incurring a high smoothness penalty. For smooth regions or edges of both object

boundaries and coarse textures, both the terms generate similar values incurring smaller penalties. This leads to the robustness of the $\mathbf{RTV}(\cdot)$ function.

The last term of $E_{Structure}$ stems from the stereo matching constraint that provides additional information to the structure sub-problem and is defined as

$$E_{StereoData}(I_{s1}, I_{s2}, D_2^*) = \sum_p \left(\alpha \cdot \sum_{q \in W_p} \left(I_{s2}(q) - I_{s1}(q - D_2^*(q)) \right)^2 + \sum_{q \in W_p} \min \left(\left| \nabla I_{s2}(q) - \nabla I_{s1}(q - D_2^*(q)) \right|, \theta \right) \right) \quad (6)$$

where the first term represents the BCC cost with a quadratic penalty function, scaled by α and summed over a fixed-size window W_p , while the second term represents the GCC cost with a truncated L_1 penalty function (with an upper threshold parameter θ), also aggregated over W_p .

3.2 Stereo Sub-problem

The first term of E_{Stereo} in Eq.(3) represents the stereo matching cost and is essentially Eq.(6) just with a change of dependent (D_2) and constant variables (I_{s1}^*, I_{s2}^*). The second term represents the smoothness cost for disparity and is defined as

$$E_{StereoSmooth}(D_2) = \sum_p \sum_{q \in N4_p} \begin{cases} \lambda_{SS1}, & \text{if } [|D_2(p) - D_2(q)| = 1] \\ \lambda_{SS2}, & \text{if } [|D_2(p) - D_2(q)| > 1] \end{cases} \quad (7)$$

where $N4_p$ represents the 4-neighbourhood of p , $[\cdot]$ is the Iverson bracket and $\lambda_{SS2} \geq \lambda_{SS1} \geq 0$ represent the regularization parameters.

Our E_{Stereo} formulation is very similar to the classic definition of the Semi-Global Matching (SGM) objective function[1] and also closely matches with the definition proposed in SGM-Stereo[32]. However, we do not use the Hamming-Census based BCC cost used in [32] mainly to avoid additional complexities in optimizing the structure sub-problem.

4 Optimization

The overall energy E_{ALL} is a challenging optimization problem. We propose to solve the problem by first decomposing it into two sub-problems $E_{Structure}$ and E_{Stereo} as shown above, and then iteratively solve them using an alternating minimization approach. The overall method is summarized in Algorithm1.⁴

We now derive the solution for $E_{structure}$. We again decompose Eq.(2) into two sub-equations, one for each image. We have for I_{s2}

$$E_{Is2}(I_{s1}^*, I_{s2}) \simeq E_{StructureData}(I_{s2}) + \lambda_S \cdot E_{StructureSmooth}(I_{s2}) + \lambda_{SD} \cdot E_{StereoData}(I_{s1}^*, I_{s2}, D_2^*) \quad (8)$$

⁴ D_{init} is obtained using our own algorithm but with $\lambda_{SD} = 0$ (no stereo constraint).

Algorithm 1 Optimize E_{ALL} **Initialize:** $I_{s1} = I_{n1}$; $I_{s2} = I_{n2}$; $D_2 = D_{init}$ **repeat**

Solve the structure sub-problem:

Fix $D_2^* = D_2$, optimize $E_{structure}$ w.r.t (I_{s1}, I_{s2}) using Algorithm2

Solve the stereo sub-problem:

Fix $(I_{s1}^*, I_{s2}^*) = (I_{s1}, I_{s2})$, optimize E_{stereo} w.r.t D_2 using SGM[1]**until** converged**Post-Processing** D_2 : Left-Right consistency[1] + Weighted Median Filtering[1]

and similarly, $E_{Is1}(I_{s1}, I_{s2}^*)$ for I_{s1} . We can observe that the stereo constraint now acts as a coupling term between the two sub-equations, thus bringing to bear the redundancy from the adjacent frame and help extract more stereo-consistent structures. Now, for solving Eq.(8), we first substitute for the individual terms, write it as a combination of two parts $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ containing the convex and non-convex parts respectively, and then solve it via the alternating direction method of multipliers(ADMM). Specifically, $E_{Is2}(I_{s1}^*, I_{s2}) = \mathbf{f}(I_{s2}) + \mathbf{g}(I_{s2})$, where

$$\begin{aligned} \mathbf{f}(I_{s2}) &= \sum_p \left((I_{s2}(p) - I_{n2}(p))^2 + \lambda_S \cdot \mathbf{RTV}(I_{s2}(p)) + \right. \\ &\quad \left. \lambda_{SD} \cdot \alpha \cdot \sum_{q \in W_p} (I_{s2}(q) - I_{s1}^*(q - D_2^*(q)))^2 \right) \\ \mathbf{g}(I_{s2}) &= \sum_p \left(\lambda_{SD} \cdot \sum_{q \in W_p} \min \left(\left| \nabla I_{s2}(q) - \nabla I_{s1}^*(q - D_2^*(q)) \right|, \theta \right) \right) \end{aligned} \quad (9)$$

where we use the approximated convex quadratic formulation of the $\mathbf{RTV}(\cdot)$ function from [31] to include it in $\mathbf{f}(\cdot)$. Now, representing $\widetilde{I}_{s1}^* = \mathbf{W}_{D_2^*}(I_{s1}^*)$ where $\mathbf{W}_{D_2^*}(\cdot)$ represents our warping function parametrized by D_2^* , and with some algebraic manipulations of $\mathbf{f}(\cdot)$, it can be defined in vector form ($\vec{\cdot}$) as

$$\begin{aligned} \mathbf{f}(\vec{I}_{s2}) &= (\vec{I}_{s2} - \vec{I}_{n2})^T (\vec{I}_{s2} - \vec{I}_{n2}) + \lambda_S \cdot \vec{I}_{s2}^T \mathbb{L}_{I_{s2}} \vec{I}_{s2} \\ &\quad + \lambda_{SD} \cdot \alpha \cdot \left((\vec{I}_{s2} - \vec{\widetilde{I}}_{s1}^*)^T \Lambda (\vec{I}_{s2} - \vec{\widetilde{I}}_{s1}^*) \right) \end{aligned} \quad (10)$$

where $\mathbb{L}_{I_{s2}}$ and Λ are some matrix operators defined later. From Eq.(10), we can see that $\mathbf{f}(\cdot)$ is a simple quadratic function and is easy to optimize. Now, for $\mathbf{g}(\cdot)$, the complication is more severe because of the windowed operation combined with a complicated penalty function, thereby coupling different columns of I_{s2} together, which means that the proximal solution for $\mathbf{g}(\cdot)$ is no longer given by iterative shrinkage and thresholding (or more exactly, its generalized version for truncated L_1 [33]). To resolve this, we swap the order of summations, obtaining

$$\mathbf{g}(I_{s2}) = \sum_{i=[-|W_p|/2, -|W_p|/2]}^{[+|W_p|/2, +|W_p|/2]} \lambda_{SD} \sum_p \min \left(\left| \nabla \mathbf{S}_i(I_{s2}(p)) - \nabla \mathbf{S}_i(\widetilde{I}_{s1}^*) \right|, \theta \right) \quad (11)$$

where $\mathbf{S}_i(\cdot)$ represents our shift function such that $\mathbf{S}_{[dx, dy]}(\cdot)$ shifts the variable by dx and dy in the x -axis and y -axis respectively. Next, if we represent $\nabla \mathbf{S}_i(\cdot)$ by a function say $\mathbf{A}_i(\cdot)$, and $-\nabla \mathbf{S}_i(\bar{I}_{s1}^*)$ by a variable say B_i , we can show that

$$\begin{aligned} \min_{I_{s2}} E_{I_{s2}}(I_{s1}^*, I_{s2}) &= \min_{I_{s2}} \mathbf{f}(I_{s2}) + \sum_i \mathbf{g}_s(\mathbf{A}_i(I_{s2}) + B_i) \\ &= \min_{I_{s2}} \mathbf{f}(I_{s2}) + \sum_i \mathbf{g}_s(Z_i) \quad \text{s.t.} \quad Z_i = \mathbf{A}_i(I_{s2}) + B_i \end{aligned} \quad (12)$$

where $\mathbf{g}_s(\cdot)$ represents $\lambda_{SD} \cdot \sum_p \mathbf{min}(|\cdot|, \theta)$ penalty function, for which we have a closed form solution[33]. Next, since $\nabla(\cdot)$, $\mathbf{S}_i(\cdot)$ $\mathbf{W}_{D_2^*}(\cdot)$ are all linear functions representable by matrix operations, we can define Eq.(12) in vector form ($\vec{\cdot}$) as

$$\min_{\vec{I}_{s2}} \mathbf{f}(\vec{I}_{s2}) + \sum_i \mathbf{g}_s(\vec{Z}_i) \quad \text{s.t.} \quad \vec{Z}_i = A_i \vec{I}_{s2} + \vec{B}_i \quad (13)$$

where A_i and \vec{B}_i are operators/variables independent of \vec{I}_{s2} , also defined later. We see that Eq.(8) reduces to a constrained minimization problem Eq.(13). The new equation is similar to the ADMM variant discussed in (Sec 4.4.2, [34]) (of the form $\mathbf{f}(\vec{I}_{s2}) + \mathbf{g}_s(A\vec{I}_{s2})$) except that our second term comprises of a summation of multiple $\mathbf{g}_s(\vec{Z}_i)$ over i rather than a single $\mathbf{g}_s(\vec{Z})$, with dependency among the various \vec{Z}_i caused by $\vec{Z}_i = A_i \vec{I}_{s2} + \vec{B}_i$. Each of these ‘‘local variables’’ \vec{Z}_i should be equal to the common global variable \vec{I}_{s2} ; this is an instance of Global Variable Consensus Optimization (Sec 7.1.1, [35]). Hence, following [34][35], we write Eq.(13) first in its Augmented Lagrangian form defined as

$$\begin{aligned} \min_{\vec{I}_{s2}, \vec{Z}_i, \vec{U}_i} \mathcal{L}(\vec{I}_{s2}, \vec{Z}_i, \vec{U}_i) &= \min_{\vec{I}_{s2}, \vec{Z}_i, \vec{Y}_i} \mathbf{f}(\vec{I}_{s2}) + \sum_i \mathbf{g}_s(\vec{Z}_i) \\ &\quad + \rho \cdot \sum_i \vec{U}_i^T (A_i \vec{I}_{s2} + \vec{B}_i - \vec{Z}_i) + \frac{\rho}{2} \cdot \sum_i \left\| A_i \vec{I}_{s2} + \vec{B}_i - \vec{Z}_i \right\|_2^2 \end{aligned} \quad (14)$$

where \vec{U}_i represent the scaled dual variables and $\rho > 0$ is the penalty parameter. Now substituting for the individual terms and minimizing Eq.(14) over the three variables, we can get the following update rules

$$\begin{aligned} \vec{I}_{s2}^{k+1} &:= \left((2\mathbb{1} + 2\lambda_S \mathbb{L}_{I_{s2}} + \lambda_{SD} \alpha (\mathbb{1} - W_2)^T (\Lambda + \Lambda^T) (\mathbb{1} - W_2)) + \rho \sum_i A_i^T A_i \right)^{-1} \\ &\quad \left((2\vec{I}_{n2} + \lambda_{SD} \alpha (\mathbb{1} - W_2)^T (\Lambda + \Lambda^T) W_1 \vec{I}_{s1}^*) - \underbrace{\rho \sum_i A_i^T (\vec{B}_i - \vec{Z}_i^k + \vec{U}_i^k)}_{\text{consensus}} \right) \\ \vec{Z}_i^{k+1} &:= \mathbf{prox}_{\frac{1}{\rho} \mathbf{g}_s} (A_i \vec{I}_{s2}^{k+1} + \vec{B}_i + \vec{U}_i^k) \\ \vec{U}_i^{k+1} &:= \vec{U}_i^k + A_i \vec{I}_{s2}^{k+1} + \vec{B}_i - \vec{Z}_i^{k+1} \end{aligned} \quad (15)$$

The update rules have an intuitive meaning. The local variables \vec{Z}_i, \vec{U}_i are updated using the global variable \vec{I}_{s2} , which then seeks consensus among all the

Algorithm 2 Optimize $E_{Structure}$

Obtain warping operators W_1, W_2 from D_2^* using Eq.(17); let $G_{xy} = G_x + G_y$

repeat

Solve $E_{Is_2}(I_{s1}^*, I_{s2})$: Obtain $\mathbb{L}_{I_{s2}}$ from Eq.(16)

1. For each i : compute S_i , $A_i = G_{xy}S_i(\mathbb{1} - W_2)$, and $\vec{B}_i = -G_{xy}S_iW_1\vec{I}_{s1}^*$

2. Solve for I_{s2} using the update rules in Eq.(15), and assign it to I_{s2}^*

Solve $E_{Is1}(I_{s1}, I_{s2}^*)$: Obtain $\mathbb{L}_{I_{s1}}$ from Eq.(16)

1. For each i : compute S_i , $A_i = -G_{xy}S_iW_1$, and $\vec{B}_i = G_{xy}S_i(\mathbb{1} - W_2)\vec{I}_{s2}^*$

2. Solve for I_{s1} using the update rules in Eq.(18), and assign it to I_{s1}^*

until converged

local variables until they have stopped changing. Now, let's define the individual terms. In Eq.(15), $\mathbb{1}$ is an identity matrix; $\mathbb{L}_{I_{s \times}} = G_x^T U_x V_x G_x + G_y^T U_y V_y G_y$ is a weight matrix[31] such that G_x, G_y are Toeplitz matrices containing the discrete gradient operators, and $U(\cdot), V(\cdot)$ are diagonal matrices given by

$$U(\cdot)(q, q) = \sum_{q \in N_p} \frac{g_\sigma(p, q)}{|\sum_{q \in N_p} g_\sigma(p, q) \cdot \partial(\cdot) I_{s \times}^k(q)| + \epsilon_s}, \quad V(\cdot)(q, q) = \frac{1}{|\partial(\cdot) I_{s \times}^k(q)|} \quad (16)$$

W_1, W_2 are warping operators such that $\vec{I}_{s1}^* = W_1 \vec{I}_{s1} + W_2 \vec{I}_{s2}$, and are given by

$$W_1(p, q) = \begin{cases} 1, & \text{if } q = p - (h \cdot \vec{D}_2^*(p)) \\ 0, & \text{if } \vec{D}_2^*(p) = 0 \end{cases}, \quad W_2(p, p) = \begin{cases} 1, & \text{if } \vec{D}_2^*(p) = 0 \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

Thus, W_1 warps \vec{I}_{s1}^* towards \vec{I}_{s2} for all the points except where $\vec{D}_2^*(p) = 0$ (invalid/unknown disparity), where we simply use the diagonal W_2 to fill-up data from \vec{I}_{s2} and avoid using our stereo constraint. Then we have $\Lambda = \sum_i S_i^T S_i$,

where S_i represents our shift operator (analogous to the definition of $\mathbf{S}_i(\cdot)$ above) defined as $S_{[dx, dy]}(p, q) = 1$, if $q = (p - dy - (h \cdot dx)) \forall p \notin V(dx, dy)$, and 0 otherwise; $V(dx, dy)$ is a set containing border pixels present in the first or last $|dx|^{\text{th}}$ column ($1 \leq |dx| \leq w$) and $|dy|^{\text{th}}$ row ($1 \leq |dy| \leq h$) depending upon whether $dx, dy > 0$ or $dx, dy < 0$, $A_i = (G_x + G_y)S_i(\mathbb{1} - W_2)$ and lastly $\vec{B}_i = -(G_x + G_y)S_iW_1\vec{I}_{s1}^*$.

Now following a similar procedure for the other image I_{s1} , we can derive the following update rules

$$\begin{aligned} \vec{I}_{s1}^{k+1} &:= \left((2\mathbb{1} + 2\lambda_S \mathbb{L}_{I_{s1}} + \lambda_{SD} \alpha (-W_1)^T (\Lambda + \Lambda^T) (-W_1)) + \rho \sum_i A_i^T A_i \right)^{-1} \\ &\quad \left((2\vec{I}_{n1} + \lambda_{SD} \alpha W_1^T (\Lambda + \Lambda^T) (\mathbb{1} - W_2) \vec{I}_{s2}^*) - \rho \sum_i A_i^T (\vec{B}_i - \vec{Z}_i^k + \vec{U}_i^k) \right) \\ \vec{Z}_i^{k+1} &:= \text{prox}_{\frac{1}{\rho} \mathbf{g}_s} (A_i \vec{I}_{s1}^{k+1} + \vec{B}_i + \vec{U}_i^k) \\ \vec{U}_i^{k+1} &:= \vec{U}_i^k + A_i \vec{I}_{s1}^{k+1} + \vec{B}_i - \vec{Z}_i^{k+1} \end{aligned} \quad (18)$$

with $A_i = -(G_x + G_y)S_iW_1$ and $\vec{B}_i = (G_x + G_y)S_i(\mathbb{1} - W_2)\vec{I}_{s2}^*$. Finally, we have the definition of $\text{prox}_{\frac{1}{\rho}\mathbf{g}_s}(\cdot)$ given by $\text{prox}_{\frac{1}{\rho}\mathbf{g}_s}(v) = \begin{cases} x_1, & \text{if } \mathbf{h}(x_1) \leq \mathbf{h}(x_2) \\ x_2, & \text{otherwise} \end{cases}$ where $x_1 = \text{sign}(v) \max(|(v|, \theta)|)$, $x_2 = \text{sign}(v) \min(\max(|(v| - (\lambda_{SD}/\rho), 0), \theta)$, and $\mathbf{h}(x) = 0.5(x - v)^2 + (\lambda_{SD}/\rho) \min(|x|, \theta)$. This completes our solution for $E_{\text{Structure}}$, also summarized in Algorithm2. The detailed derivations for Eq.(10), Eq.(11) and Eq.(15) are provided in the supplementary paper for reference.

5 Experiments

In this section, we evaluate our algorithm through a series of experiments. Since there are not many competing algorithms, we begin with creating our own baseline methods first. We select the two best performing denoising algorithms, BM3D[10] and DnCNN[13] till date, to perform denoising as a pre-processing step, and then use MeshStereo[17], a recent high performance stereo algorithm, to generate the disparity maps. The codes are downloaded from the authors' websites. We refer to these two baseline methods as 'BM3D+MS' and 'DnCNN+MS' respectively. Our third baseline method is a recently proposed joint denoising-disparity algorithm[19], which we refer to as 'SS-PCA'. Due to unavailability of the code, this method is based on our own implementation.

For our first experiment, we test our algorithm against the baseline methods on the Middlebury(Ver3) dataset[18] corrupted with Gaussian noise at levels: 25, 50, 55 and 60, i.e. we consider one low and three high noise cases, the latter resulting in low SNR similar to those encountered in night scenes. To ensure a fair comparison, we select three images 'Playroom', 'Recycle' and 'Teddy', from the dataset and tune the parameters of BM3D and SS-PCA to generate the best possible PSNR results for every noise level, while for DnCNN, we pick its blind model trained on a large range of noise levels. Furthermore, we keep the same disparity post-processing steps for all the algorithms including ours to ensure fairness. Our stereo evaluation metric is based on the percentage of bad pixels, i.e. percentage (%) of pixels with disparity error above a fixed threshold δ . For our algorithm, we set the parameters $\{\lambda_S, \epsilon_s, \lambda_{SD}, \alpha, \theta, \rho, \lambda_{SS}, \lambda_{SS1}, \lambda_{SS2}\} = \{650.25, 5, 1, 0.003, 15, 0.04, 1, 100, 1600\}$, $|W_p| = 25 (= 5 \times 5)$, and use $\sigma = 1.0, 2.0, 2.5$ and 3.0 for the four noise levels respectively. The number of outermost iteration is fixed to 5 while all the inner iterations follow $(\Delta E_{\times}^{k+1}/E_{\times}^k) < 10^{-4}$ for convergence. Our evaluation results are summarized in Tables 1 & 2.

For our second experiment, we perform our evaluation on the real outdoor Oxford RobotCar[15] dataset, specifically those clips in the 'night' category. These clips contain a large amount of autonomous driving data collected under typical urban and suburban lighting in the night, with a wide range of illumination variations. It comes with rectified stereo images and their corresponding raw sparse depth ground truth. We create two sets of data, 'Set1' containing 10 poorly-lit images (such as in Fig.1a), and 'Set2' containing 20 well-lit images (selection criteria is to maximize variance in the two sets in terms of scene content therefore no consecutive/repetitive frames; scenes with moving objects are

Table 1. Image-wise evaluation on the Middlebury dataset with added Gaussian noise at levels: [25, 50, 55, 60]. Error threshold $\delta = 1\text{px}$. Bold font indicates lowest error.

Image	BM3D+MS				DnCNN+MS				SS-PCA				Ours			
	25	50	55	60	25	50	55	60	25	50	55	60	25	50	55	60
‘Adirondack’	37.57	52.95	56.98	62.02	35.80	47.99	51.37	56.13	60.01	66.40	80.57	84.67	38.76	44.85	49.00	50.74
‘Jadeplant’	66.17	79.52	76.84	80.42	68.49	77.43	76.45	78.90	64.42	75.78	78.30	81.75	72.29	78.92	77.76	80.40
‘Motorcycle’	40.75	50.86	51.66	52.80	37.63	50.46	50.61	49.62	41.74	47.81	50.63	54.16	40.44	45.17	43.21	44.17
‘Pipes’	41.35	58.08	60.47	63.07	37.07	47.62	53.28	53.20	39.52	50.73	56.97	61.31	45.82	54.48	55.90	60.56
‘Playroom’	46.82	55.35	57.23	55.72	41.46	49.21	54.77	57.64	57.82	62.96	71.65	75.56	43.87	48.87	50.36	52.74
‘Recycle’	48.65	61.28	62.91	63.43	44.20	57.72	60.52	60.22	51.64	64.45	66.04	69.20	50.42	57.72	57.38	54.83
‘Shelves’	60.18	69.24	71.44	70.56	55.82	66.05	64.68	66.64	63.28	68.03	74.96	73.99	58.89	62.58	63.07	63.93
‘Teddy’	30.15	49.20	52.78	58.79	27.01	44.05	50.39	49.46	32.65	44.14	52.89	52.75	31.39	40.86	45.07	45.71

Table 2. Overall evaluation on the Middlebury dataset with added Gaussian noise at levels: [25, 50, 55, 60] for error threshold δ . Bold font indicates lowest error.

δ	BM3D+MS				DnCNN+MS				SS-PCA				Ours			
	25	50	55	60	25	50	55	60	25	50	55	60	25	50	55	60
1px	46.45	59.55	61.29	63.35	43.43	55.06	57.76	58.97	51.39	60.04	66.48	69.17	47.74	54.19	55.22	56.59
3px	22.68	30.57	33.72	34.63	22.04	29.62	32.67	32.68	30.41	35.32	42.02	43.67	25.12	29.00	29.45	30.48
5px	16.22	22.01	24.17	25.07	16.82	21.53	24.36	23.94	23.14	26.07	31.48	32.93	18.21	20.94	20.60	21.81

also discarded due to unreliability of ground truth); together they span a range of conditions such as varying exposure, sodium vs LED lightings, amount of textures, image saturation, and error sources such as specularities (specific details in supplementary). We set the parameters $\{\lambda_S, \lambda_{SD}, \lambda_{SS}\} = \{50.25, 0.1, 0.1\}$ while keeping other parameters exactly the same as before for both the sets, and compare our algorithm only against ‘DnCNN+MS’ since there are no corresponding noise-free images available to tune the other baseline algorithms for maximizing their PSNR performance. Our evaluation results are summarized in Table 3 (‘Set2 (f.t)’ denotes evaluation with parameters further fine tuned on ‘Set2’).

From the experimental results, we can see that for all the highly noisy (or low SNR) cases, our algorithm consistently outperforms the baseline methods quite significantly with improvements as high as 5-10% in terms of bad pixels percentage. Our joint formulation generates stereo-consistent structures (unlike denoising, see Fig.2) which results in more accurate and robust stereo matching under highly noisy conditions. The overall superiority of our method is also quite conspicuous qualitatively (see Fig.3). We achieve a somewhat poorer recovery for ‘Jadeplant’ and ‘Pipes’, the root problem being the sheer amount of spurious corners in the scenes which is further aggravated by the loss of interior texture in our method. For low noise levels, there is sufficient signal (with finer textures) recovery by the baseline denoising algorithms, thus yielding better disparity solutions than our structures which inevitably give away the fine details. Thus, our algorithm really comes to the forth for the high noise (or low SNR) regimes. For the real data, our algorithm again emerges as the clear winner (see Table 3 and middle block of Fig.3). First and foremost, we should note that the parameters used for ‘Set1’ and ‘Set2’ are based on those tuned on two sequences in ‘Set1’. The fact these values are transferable to a different dataset (‘Set2’) with rather different lighting conditions showed that the parameter setting works quite well

Table 3. Comparison with the baseline methods on the Oxford RobotCar dataset. Error threshold is specified by δ . Bold font indicates lowest error.

	DnCNN+MS					Ours				
	$\delta = 1\text{px}$	$\delta = 2\text{px}$	$\delta = 3\text{px}$	$\delta = 4\text{px}$	$\delta = 5\text{px}$	$\delta = 1\text{px}$	$\delta = 2\text{px}$	$\delta = 3\text{px}$	$\delta = 4\text{px}$	$\delta = 5\text{px}$
Set1	63.86	41.66	30.96	24.40	19.66	58.76	33.75	23.03	16.99	12.31
Set2	58.96	28.82	16.71	10.73	7.35	57.76	28.80	16.10	10.29	6.82
Set2 (f.t)	58.96	28.82	16.71	10.73	7.35	56.45	26.43	14.54	9.20	6.08

under a wide range of lighting conditions (depicted in the middle block of Fig.3). Qualitatively, the proficiency of our algorithm in picking up 3D structures in the very dark areas, some even not perceivable to human eyes, is very pleasing (see red boxes in the middle block of Fig.3, row 1: wall on the left, rows 2 and 3: tree and fence). It is also generally able to delineate relatively crisp structures and discern depth differences (e.g. the depth discontinuities between the two adjoining walls in row 4), in contrast to the patchwork quality of the disparity returned by ‘DnCNN+MS’. Finally, our algorithm also seems to be rather robust against various error sources such as glow from light sources, under-to-over exposures. Clearly, there will be cases of extreme darkness and such paucity of information, against which we cannot prevail (bottom block of Fig.3, top-right: a scene with sole distant street lamp). Other cases of failures are also depicted in the bottom block of this figure, namely, lens flare and high glare in the scene.

6 Discussion and Conclusion

We have showed that under mesopic viewing condition, despite the presence of numerous challenges, disparity information can still be recovered with adequate accuracy. We have also argued that for denoising, PSNR is not meaningful; instead there should be a close coupling with the disparity estimation task to yield stereo-consistent denoising. For this purpose, we have proposed a unified energy objective that jointly removes noise and estimates disparity. With careful design, we transform the complex objective function into a form that admits fairly standard solutions. We have showed that our algorithm has substantially better performance over both synthetic and real data, and is also stable under a wide range of low-light conditions.

The above results were obtained based on the assumptions that effects of glare/glow could be ignored. Whilst there has been some stereo works that deal with radiometric variations (varying exposure and lighting conditions), the compounding effect of glare/glow on low-light stereo matching has not been adequately investigated. This shall form the basis of our future work.

Acknowledgement

The authors are thankful to Robby T. Tan, Yale-NUS College, for all the useful discussions. This work is supported by the DIRP Grant R-263-000-C46-232.

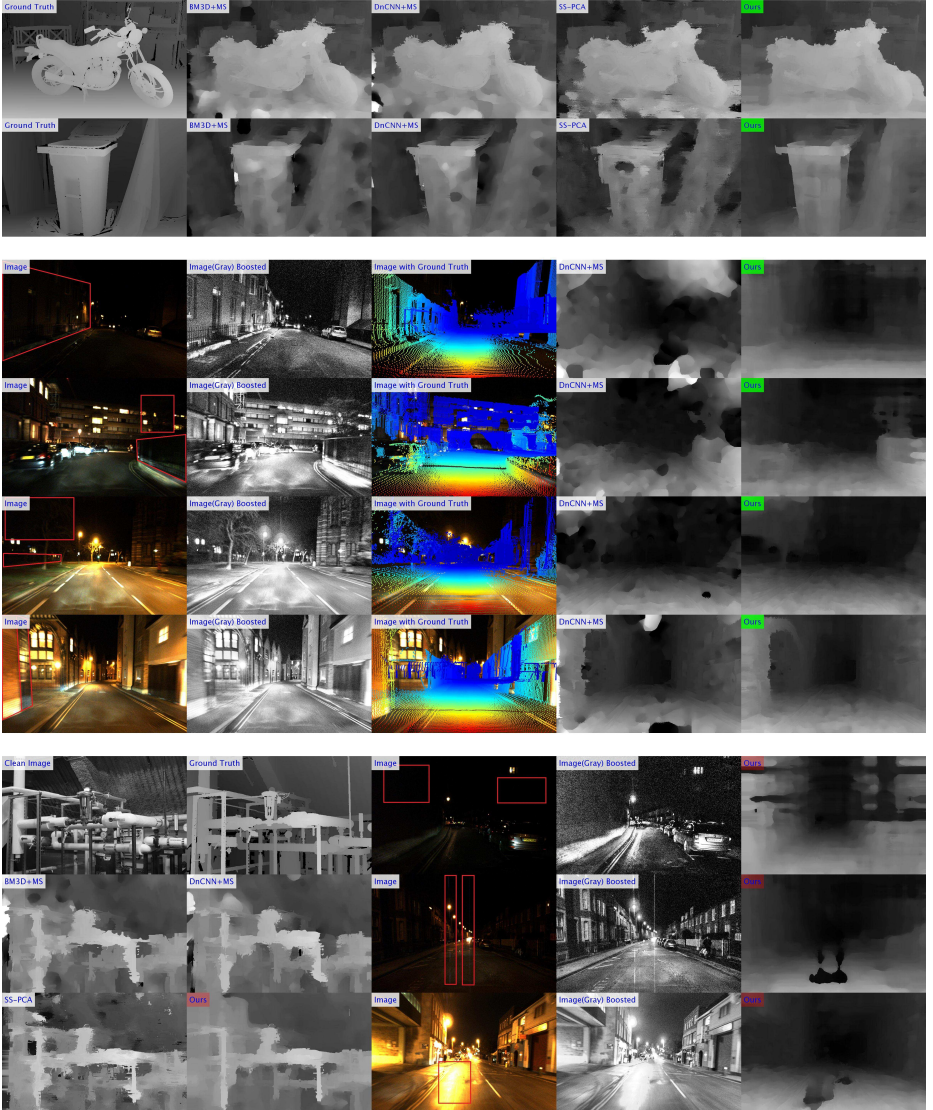


Fig. 3. Qualitative analysis of our algorithm against the baseline methods. For Middlebury (first two rows), we observe more accurate results with sharper boundaries (see ‘Recycle’ image, second row). For the Oxford dataset (middle four rows), our algorithm generates superior results and is quite robust under varying illumination and exposure conditions, and can even pick up barely visible objects like fence or trees (see areas corresponding to red boxes in middle second and third row). Our algorithm also has certain limitations in extremely dim light information-less conditions (see red boxes, third last row) or in the presence of lens flare or high glow/glare in the scene (bottom two rows), generating high errors in disparity estimation.

References

1. Hirschmuller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 2., IEEE (2005) 807–814
2. Bleyer, M., Rhemann, C., Rother, C.: Patchmatch stereo-stereo matching with slanted support windows. In: Bmvc. Volume 11. (2011) 1–11
3. Zbontar, J., LeCun, Y.: Computing the stereo matching cost with a convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 1592–1599
4. Luo, W., Schwing, A.G., Urtasun, R.: Efficient deep learning for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 5695–5703
5. Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., Bry, A.: End-to-end learning of geometry and context for deep stereo regression. CoRR, vol. abs/1703.04309 (2017)
6. Diamond, S., Sitzmann, V., Boyd, S., Wetzstein, G., Heide, F.: Dirty pixels: Optimizing image classification architectures for raw sensor data. arXiv preprint arXiv:1701.06487 (2017)
7. Jeon, H.G., Lee, J.Y., Im, S., Ha, H., So Kweon, I.: Stereo matching with color and monochrome cameras in low-light conditions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 4086–4094
8. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 2., IEEE (2005) 60–65
9. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena* **60**(1-4) (1992) 259–268
10. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing* **16**(8) (2007) 2080–2095
11. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with bm3d? In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 2392–2399
12. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: Advances in neural information processing systems. (2012) 341–349
13. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing* **26**(7) (2017) 3142–3155
14. Ruderman, D.L., Bialek, W.: Statistics of natural images: Scaling in the woods. In: Advances in neural information processing systems. (1994) 551–558
15. Maddern, W., Pascoe, G., Linegar, C., Newman, P.: 1 Year, 1000km: The Oxford RobotCar Dataset. *The International Journal of Robotics Research (IJRR)* **36**(1) (2017) 3–15
16. Guo, X.: Lime: A method for low-light image enhancement. In: Proceedings of the 2016 ACM on Multimedia Conference, ACM (2016) 87–91
17. Zhang, C., Li, Z., Cheng, Y., Cai, R., Chao, H., Rui, Y.: Meshstereo: A global stereo model with mesh alignment regularization for view interpolation. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 2057–2065

18. Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P.: High-resolution stereo datasets with subpixel-accurate ground truth. In: German Conference on Pattern Recognition, Springer (2014) 31–42
19. Jiao, J., Yang, Q., He, S., Gu, S., Zhang, L., Lau, R.W.: Joint image denoising and disparity estimation via stereo structure pca and noise-tolerant cost. *International Journal of Computer Vision* **124**(2) (2017) 204–222
20. Levin, A., Nadler, B., Durand, F., Freeman, W.T.: Patch complexity, finite pixel correlations and optimal denoising. In: European Conference on Computer Vision, Springer (2012) 73–86
21. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision* **47**(1-3) (2002) 7–42
22. Hirschmuller, H., Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. *IEEE transactions on pattern analysis and machine intelligence* **31**(9) (2009) 1582–1599
23. Buades, A., Coll, B., Morel, J.M.: Image denoising methods. a new nonlocal principle. *SIAM review* **52**(1) (2010) 113–147
24. Wen, B., Li, Y., Pfister, L., Bresler, Y.: Joint adaptive sparsity and low-rankness on the fly: an online tensor reconstruction scheme for video denoising. In: IEEE International Conference on Computer Vision (ICCV). (2017)
25. Li, N., Li, J.S.J., Randhawa, S.: 3d image denoising using stereo correspondences. In: TENCON 2015-2015 IEEE Region 10 Conference, IEEE (2015) 1–4
26. Liu, C., Freeman, W.T.: A high-quality video denoising algorithm based on reliable motion estimation. In: European Conference on Computer Vision, Springer (2010) 706–719
27. Zhang, L., Vaddadi, S., Jin, H., Nayar, S.K.: Multiple view image denoising. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 1542–1549
28. Aujol, J.F., Gilboa, G., Chan, T., Osher, S.: Structure-texture image decompositionmodeling, algorithms, and parameter selection. *International journal of computer vision* **67**(1) (2006) 111–136
29. Xu, Y., Long, Q., Mita, S., Tehrani, H., Ishimaru, K., Shirai, N.: Real-time stereo vision system at nighttime with noise reduction using simplified non-local matching cost. In: Intelligent Vehicles Symposium (IV), 2016 IEEE, IEEE (2016) 998–1003
30. Heo, Y.S., Lee, K.M., Lee, S.U.: Simultaneous depth reconstruction and restoration of noisy stereo images using non-local pixel distribution. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE (2007) 1–8
31. Xu, L., Yan, Q., Xia, Y., Jia, J.: Structure extraction from texture via relative total variation. *ACM Transactions on Graphics (TOG)* **31**(6) (2012) 139
32. Yamaguchi, K., McAllester, D., Urtasun, R.: Efficient joint segmentation, occlusion labeling, stereo and flow estimation. In: European Conference on Computer Vision, Springer (2014) 756–771
33. Gong, P., Zhang, C., Lu, Z., Huang, J., Ye, J.: A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: International Conference on Machine Learning. (2013) 37–45
34. Parikh, N., Boyd, S., et al.: Proximal algorithms. *Foundations and Trends® in Optimization* **1**(3) (2014) 127–239
35. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* **3**(1) (2011) 1–122