Enhanced Depth Map Estimation in Low Light Conditions for RGB Cameras

Joseph Chang and Truong Q. Nguyen
ECE Dept., UCSD
La Jolla, CA, 92093-0407, USA
jdchang@ucsd.edu, tqn001@eng.ucsd.edu

Abstract— **Most existing depth estimation methods predict depth for daytime images and do not perform well in low-light situations due to lack of clear environmental features, glare, overexposure, and noise. This is problematic for safe autonomous driving as pedestrian and guardrail detection at night is challenging and poses life-threatening situations. This paper addresses this problem by improving image quality of disparity maps obtained in low-light based on previous work. We introduce an algorithm that combines a defogging method, which enhances night images and improves luminance, with generative adversarial or fully-convolutional networks to accurately learn the correct disparity prediction. The experiments show that the proposed method outperforms state-of-the-art methods which do not use additional preprocessing.**

Keywords - depth estimation, image enhancement, low light, image signal processor, advanced driver assistance systems

I. INTRODUCTION

Depth estimation is key to enhancing safety in advanced driver assistance systems (ADAS). It is used to gain a 3D understanding of the world from 2D images. Monocular depth estimation has lower accuracy than stereo-based depth estimation, but would enable creation of many more 3D datasets in the future. Lidar estimates depth by measuring a laser's return time from its target. Although it is highly accurate, it requires expensive equipment for longer-range applications such as driving scenarios. Hence, stereo cameras are currently the best method for depth estimation as they produce accurate depth while not requiring expensive sensors. Despite the importance of depth estimation to existing applications such as virtual reality, autonomous cars, and robotic surgery, the low-light and nighttime scenarios have hardly been explored which restricts these systems to run only in daytime or under artificial light.

We formulate an algorithm to improve estimated depth map quality in low-light conditions for ADAS by adding a preprocessing method to existing unsupervised learning algorithms. Since there are nearly no 3D nighttime street view datasets with ground truth available, supervised learning is unsuitable. Oxford RobotCar provides nighttime car datasets with sparse depth from Lidar. Hence, our method utilizes these datasets to test if the proposed method accurately generates a disparity map given any nighttime RGB image. Then, the data is tested with monocular and stereo depth estimation algorithms with and without our preprocessing enhancement. The evaluation metrics show that the proposed method outperforms the current state-of-the-art methods. In future works, the enhanced depth maps from our algorithm can be applied to object detection and classification for autonomous vehicles driving at night.

II. METHOD

A. *Dataset Processing*

We use the Oxford RobotCar Dataset [1][2] which contains a mix of daytime and nighttime street view data in various weather conditions. The images are stored in a raw format, so we convert them to RGB using a lookup table (LUT). The data is then ready for preprocessing and training. However, we also need sparse ground truth disparity to calculate error later. For this, we gather the camera model and extrinsic parameters of the stereo camera, 2D Lidar data from the front bumper, Inertial Navigation System (INS) data, and timestamps from the Lidar and stereo camera for each dataset. The sparse Lidar depth maps are projected into the camera resulting in sparse depth maps which are converted to sparse disparity maps. The complete process is shown in Figure 1.

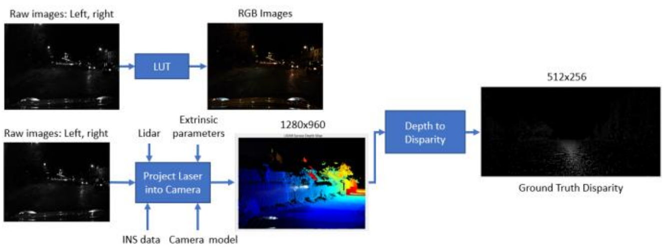


Figure 1. Dataset color conversion and sparse ground truth generation.

B. *Image Preprocessing Method*

Given a YUV nighttime image I, we invert its luma channel with 255-I as shown in Figure 2 and notice the result looks similar to a foggy daytime image. We utilize this idea to construct a preprocessing procedure that enhances nighttime images by running them through a defogging process. We hypothesize that this process will improve image clarity and depth estimation. The improved contrast should make the images appear as if they were taken during twilight where objects have clearer edges and texture. This will give the networks more features to extract while training and lead to better disparity prediction on nighttime images. To defog our dataset, we choose the joint contrast enhancement and turbulence mitigation method (CETM) [3] designed to mitigate turbulence and remove fog in images while being fast enough for real-time applications. First, we convert the nighttime images to the YUV domain and then invert the luminance channel. Then, CETM is applied to denoise and defog the images with atmospheric blur reduction. The result is smoothed by feature tracking using optical flow. To restore the image, we un-invert the luma channel with 255-I and convert the image back to RGB.

Figure 2. Inverted YUV nighttime image looks like foggy daytime image.

III. Experimental Results

To demonstrate our analysis, we utilize the Oxford RobotCar Dataset [1][2]. Our test set contains randomly selected nighttime street view images from 2014-11-14 with data covering a variety of scenarios including glare, motion blur, noise, and over exposed lights. Ground truth is obtained as detailed in Section 2A. All images are preprocessed (enhanced) with CETM defogging according to Section 2B. Then, we test our non-enhanced and enhanced images on depth estimation algorithms to evaluate if our method outperforms the original method which has no preprocessing.

A. Depth Estimation

For monocular depth estimation from single images, we choose monodepth2 [4], a prominent algorithm which uses fully convolutional networks. We run single images from the left stereo camera through the mono+stereo_1024x320 model pretrained on daytime images. The output depth is converted to disparity. Some sample results are shown in Figure 3.

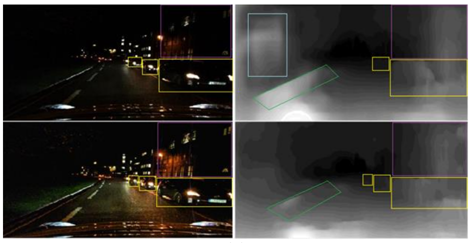
For stereo depth estimation, we choose Stereo-Consistent Cyclic Translations [5], an algorithm created specifically for nighttime images. It is trained using a generative adversarial network (GAN). We run 6,001 stereo images from the left and right camera through the pretrained model and achieve disparity results shown in Figure 3.

B. Error Evaluation Metric

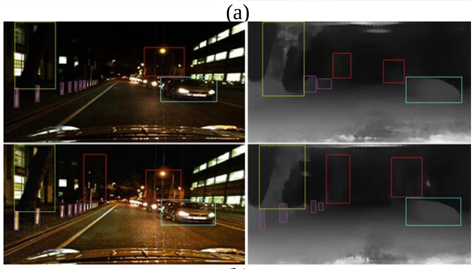
We calculate disparity error as the percentage of pixels that are $k=0\text{-}20$ away from the true disparity where $k=0$ indicates a correct disparity estimate and $k=20$ indicates large error. Since our ground truth is sparse, we only test non-zero ground truth pixels. The lower 10% of the ground truth and estimated disparity maps are cropped as they are often covered by a car hood and return noisy results. Furthermore, the estimated disparity's resolution is resized to match the ground truth using nearest neighbor.

A good disparity estimation algorithm should yield lower percentage error and higher percentage accuracy. We use an accumulated pixel percentage accuracy (APPA) metric for this experiment. Let $P(k)$ be the percentage pixel accuracy for a given k, then the accumulation function is defined as

$$Y(k) = Y(k-1)+P(k) \tag{1}$$



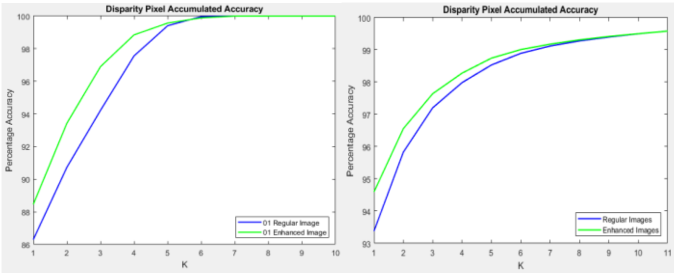
(a)



(b)

Figure 3. Estimated disparity for normal nighttime images and preprocessed nighttime images (a) monocular (b) stereo cameras.

APPA results for this experiment are shown in Figure 4. Images enhanced with our preprocessing have greater accuracy for monodepth2 and Stereo-Consistent Cyclic Translations. We conclude that nighttime images enhanced with the proposed preprocessing achieve greater disparity accuracy than the original state-of-the-art algorithms.



(a) (b)

Figure 4. APPA for depth estimation on pretrained models with the Oxford dataset run on (a) single monocular images (b) stereo dataset

IV. Conclusion

We propose a preprocessing algorithm to improve depth estimation for low-light and nighttime images. The preprocessing algorithm consists of inverting the low-light image and applies a defogging algorithm. The proposed method yields more accurate depth estimation for both monocular and stereo-based depth estimation algorithms.

References

[1] W. Maddern, G. Pascoe, *et. al.*, "1 year, 1000km: The oxford robotcar dataset," in Int. Journal of Robotics Research (IJRR), 2016.

[2] Dan Barnes, Matthew Gad, *et. al.*, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in IEEE International Conference on Robotics and Automation (ICRA) 2020.

[3] K. Gibson, *et. al.*, "An analysis and method for contrast enhancement turbulence mitigation methods," in IEEE Transac. on Image Proc. 2014.

[4] Clement Godard, *et. al.*, "Digging into self-supervised monocular depth estimation," in IEEE Conf. on Comput. Vis. (ICCV), Oct. 2019.

[5] L. H. Aashish Sharma, *et. al.*, "Nighttime stereo depth estimation using joint translation-stereo learning," in Int. Conf. on 3D Vis.. (3DV), 2020.