The Effect Analysis of Thermal Infrared Colorization

Daechan Han*, Jeongmin Shin*, Hyeongjun Kim and Yukyung Choi†

***Abstract*— Due to the advantage that a thermal camera robustly works in various illumination conditions, it has become a crucial sensor in real-world applications, such as self-driving, advanced driver assistance as well as a surveillance system. However, unlike RGB images with color information, thermal images do not contain abundant information. This disadvantage makes it users difficult to recognize thermal scenes. In this paper, we aim to make pseudo-RGB that can be used at day and night by receiving thermal images as inputs and to show the necessity of pseudo-RGB research through colorization that synthesizes chromaticity of RGB in thermal images. Furthermore, we evaluate models to see whether the produced colorized image can be interpreted by perceptual models for tasks, such as pedestrian detection and depth estimation by feeding colorized images to the models that are trained with original RGB images. These experiments explicitly show the limitation and possibility of thermal colorization.**

I. INTRODUCTION

Unlike RGB cameras, thermal cameras are robust to illumination changes, so they can be used regardless of day or night. Because of this advantage, thermal camera has become a very important sensor in various systems, such as autonomous driving, advanced driver assistance [1], [2], and surveillance systems [3], [4], [5]. Besides, in various computer vision tasks such as pedestrian detection [6], [7], [8], [9], [10] and semantic segmentation [11], [12], [13], research using thermal images has been actively conducted due to the importance of the sensor.

On the other hand, since a thermal camera captures a different wavelength range that cannot be perceived in a human visual system, color information does not exist in the thermal domain. As a result, humans struggle with understanding thermal images intuitively compared to RGB images. In particular, it is not an easy task for a fast speed driver to quickly perceive the surrounding situation only using thermal images as they lack visual information. Also, thermal colorization research is quite important for vision perceptual models. Zhang [14] tested a VGG model which is pretrained on ImageNet to see the performance difference between when only using RGB images and when feeding grey images to the network. If the color information barely influences the classifier, the performance of the model when grey images are fed will be almost the same with the result when RGB images are fed. However, the performance when grey images are used as inputs is reduced by 16% compared to the RGB case. It was proved that color information has a critical role in a model inference stage. However, since thermal images which lack color information provide a silhouette of scenes, available information is limited for vision perceptual models which use thermal images as inputs.

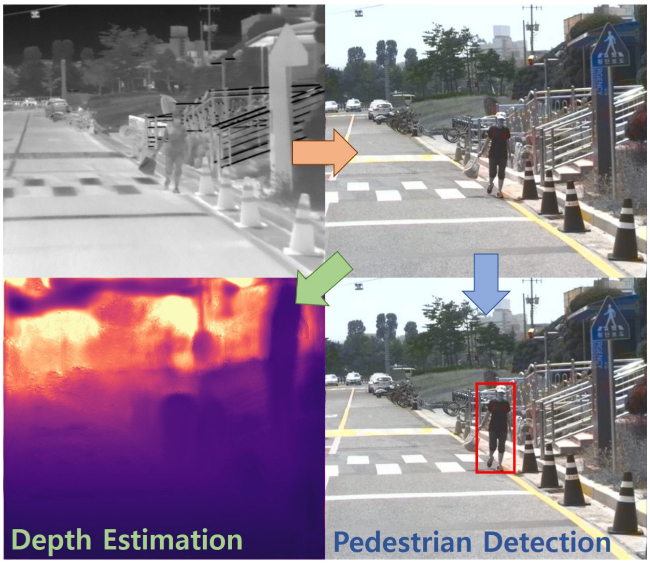

Fig. 1: **Examples of Thermal infrared Colorization for multiple vision tasks.** We depict an orange arrow to the method of creating pseudo-RGB which combining real grey images and generated color information from thermal images. The blue and green arrows indicate the evaluation method to verify the effectiveness analysis of pseudo-RGB.

In order to solve this problem, we make colorized thermal images by using instance-aware colorization method. Unlike other tasks in computer vision, colorization has difficulties because there is no specific ground truth of inputs. Nevertheless, in recent years, colorization methods using CNN have been proposed, and the methods have achieved compromising results.

Zhang [14] tried to solve the uncertainty of the colorization problem by designing an objective function according to the colorization characteristics that cannot be specified with only one correct answer color. Su [15] proposed a method that model learns about the full-image level and the object level, respectively by perform-

*Equal Contribution, †Corresponding author

Daechan Han, Jeongmin Shin, Hyeongjun Kim and Yukyung Choi are with the Robotics and Computer Vision Lab, Sejong University, South Korea `{dchan,jmshin,hjkim,ykchoi}@rcv.sejong.ac.kr`

TABLE I: **Quantitative results of the colorization according to datasets**

Dataset	Test Image	PSNR ↑	SSIM ↑	LPIPS ↓
MTN[17]	$L_T + ab_T$	**27.9761**	**0.4052**	**0.5074**
	$L_G + ab_G$	35.0415	0.9692	0.0822
	$L_G + ab_T$	**34.1562**	**0.9658**	**0.1080**
	Grey	33.3379	0.9697	0.1273
Sejong	$L_T + ab_T$	**27.9214**	**0.4422**	**0.5276**
	$L_G + ab_G$	34.4943	0.9520	0.0875
	$L_G + ab_T$	**32.7555**	**0.9281**	**0.1588**
	Grey	29.2966	0.9267	0.3069

TABLE II: **Quantitative results of the colorization according to scene types.**

Dataset	Test Image	PSNR ↑	SSIM ↑	LPIPS ↓
Campus[17]	$L_G + ab_T$	**33.1988**	**0.9598**	**0.1219**
	Grey	32.5859	0.9633	0.0262
Residential[17]	$L_G + ab_T$	**35.4836**	**0.9727**	**0.0966**
	Grey	34.7123	0.9779	0.0976
Urban[17]	$L_G + ab_T$	**34.3022**	**0.9687**	**0.0943**
	Grey	32.9607	0.9714	0.0993

ing instance aware image colorization through an off-the-shelf object detector. This method greatly improves the color expression accuracy for the instance because of the reduced color confusion phenomenon between the background and the object.

We conduct thermal colorization experiments based on the network proposed in the instance-aware colorization method and evaluate models on both of the outdoor and indoor datasets. In our extensive experiments, we also show how much color information would affect the performance of models used in various computer vision tasks, such as pedestrian detection and depth estimation. By separating the images by chrominance and luminance, the reasoning ability of the model was analyzed. By splitting RGB scale images into chrominance and luminance and comparing them, we analyzed the relation between performance of the model and the way the images are described.

II. Method

We conduct thermal infrared colorization to obtain meaningful color information and design various experiments to analyze the effect of color information to models.

A. Thermal Infrared Colorization

Thermal domain is hard to represent details of instance object in comparison with RGB. Thus, if the model uses only the images without detecting objects, the colorization result cannot be compromising. Therefore, we adopt the recent work [15] proposed for meaningful instance-aware colorization as the thermal infrared colorization model.

The structure of Instance-Colorization consists of three parts: 1) full image colorization; 2) instance colorization; 3) the fusion module. These two networks are trained to perform full and instance colorization using entire images and cropped object images as inputs. Then the fusion module aims to train final colorization to estimate final color images at the object level and image level features. For employing instance-aware colorization method to thermal infrared colorization, we feed thermal images $X \in \mathbb{R}^{H \times W \times 1}$ instead of grey images as inputs and infer the outputs of these networks as color channels $Y \in \mathbb{R}^{H \times W \times 2}$ corresponding to the CIE L*a*b color space.

B. Depth Estimation

To evaluate the impact of the predicted color information on the performance of the depth estimation model, we set the base model as Monodepth [18]. This method proposed a self-supervised method of single image depth estimation and it is popular method in the task. They also suggested the left-right consistency method which focuses on to make a left disparity and a right disparity similar to each other. It can perform an accurate estimation for a depth map by preventing reflected texture on the disparity map, i.e. text copy problem.

C. Pedestrian Detection

We designed the pedestrian detection experiment to evaluate the image colorization of instance-level. Furthermore, we confirmed the correlation between detection performance and luminance and chrominance. SSD [16] was used as a base model because it is one of the well-known and fundamental methods in pedestrian detection. This model highly improves train and inference speed compared to previous works by composing the network as a one-stage structure.

III. Experimental Results

A. Datasets

1) KAIST Multispectral Dataset(MTN): We used Multispectral Transfer Network dataset [17] to evaluate colorization and depth estimation. This dataset provides a calibrated RGB stereo pair, co-aligned thermal images with left-view RGB stereo images, and 3D measurements. Therefore, it is widely used in visual perception tasks such as depth estimation, color estimation, and visual localization. KAIST multispectral dataset also focuses on real-world driving conditions, such as the campus, residential areas, urban areas, and suburbs during the day and night time. This dataset consists of 7,383 images per each camera respectively where 4,534

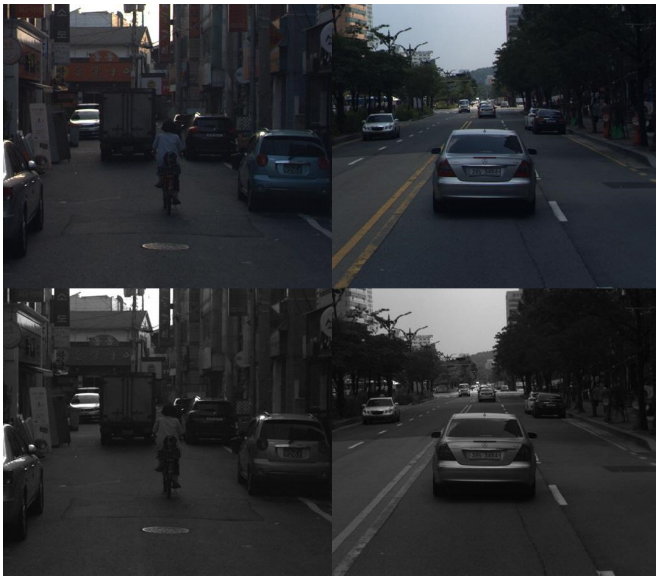

Fig. 2: **Examples of desaturated image.** First column: RGB image, second column: grey image

images are used for the train set and 1,583 images for test. We exclude suburbs that have a mis-alignment problem between thermal and RGB images caused by the vibration when hitting speed bumps. Thus, we used 3,073 images for training and 1,784 images for evaluation, respectively.

2) Sejong Multispectral Dataset: We used the Sejong Multispectral dataset to evaluate the generated color and detection performance when feeding the fake RGB images into the pedestrian detector. This dataset is designed for research on avoiding obstacles of self-driving forklifts in indoor warehouses. It provides stereo RGB and stereo thermal image pairs. Except for the images which do not include any pedestrian, we used 7,295 images for training and 7,200 images for evaluation.

B. Quantitative Results

1) Evaluation on Colorization: Table I shows the quantitative results of thermal colorization on two datasets. $L_X + ab_Y$ means the concatenation of the luminance of X images and estimated chromaticity from Y images. in addition, G and T stand for Grey and Thermal, respectively. Grey-scale images and thermal images are supplied as inputs to the model, respectively. We then create an evaluation image by concatenating each input image (L_G, L_T) and the estimated color information (ab_G, ab_T). The $L_T + ab_T$ shows extreme performance degradation on MTN and Sejong datasets compared to L_G+ab_T and grey. However, L_G+ab_T have almost the same performance compared to $L_G + ab_G$. Therefore, the performance gap between $L_T + ab_T$ and L_G+ab_T is affected by the difference of contrast. In MTN dataset, there is no significant performance difference between L_G+ab_T and grey. To analyze the above result, we evaluated the dataset by separating places such as campus, residential, urban, and report the results in Table II. In residential and urban areas, the grey images achieve higher SSIM and LPIPS [21] scores compared to campus. It indicates that the RGB images are similar to grey images due to the desaturation(Refer to Figure 2). Also, since the SSIM metric focuses on the structure of images, the metric is more affected by luminance than chrominance, and this have a small gap between grey and colorized results. However, the performance difference is observed as campus images are relatively saturated whereas LPIPS scores for grey and $L_G + ab_T$ tend to be similar to each other in urban and residential areas. This result implies that the LPIPS metric is appropriate when evaluating colorization methods based on deep features with respect to conventional distortions, such as photometric, noise and blur as well as CNN-based distortions. Besides, $L_G + ab_T$ indicates that the model is able to produce good colorization results from thermal images.

TABLE III: **Quantitative results of depth estimation**

Test image	With Color	RMSE↓	RMLSE↓
RGB	O	4.3571	0.1871
$L_G + ab_T$	O	**4.4380**	**0.1896**
$L_T + ab_T$	O	6.8368	0.2892
Grey	X	4.7158	0.2017
Thermal	X	6.8040	0.2810

TABLE IV: **Quantitative results of pedestrian detection**

Test image	With Color	MAP ↑
RGB	O	95.46
$L_G + ab_T$	O	**92.74**
$L_T + ab_T$	O	62.01
Grey	X	62.81
Thermal	X	24.61

2) Evaluation on Depth Estimation: To evaluate the reconstruction of pixel-level of ab_T, we feed $L_G + ab_T$ as inputs to depth estimation model (Monodepth [15]) trained in real RGB. If the result of depth estimation is impressive, it means that thermal colorization worked well at the pixel level. According to Table III, the RMSE decreased from 4.7158 to 4.3571 when there is no color information in the input image (RGB v.s. Grey). On the other hand, the RMSE (4.438) of using the input image with ab_T is a little different from the result of RGB (4.3571). Furthermore, performance of $L_T + ab_T$ seems to decrease significantly. We conclude the input of luminance is more critical than chromaticity for depth estimation.

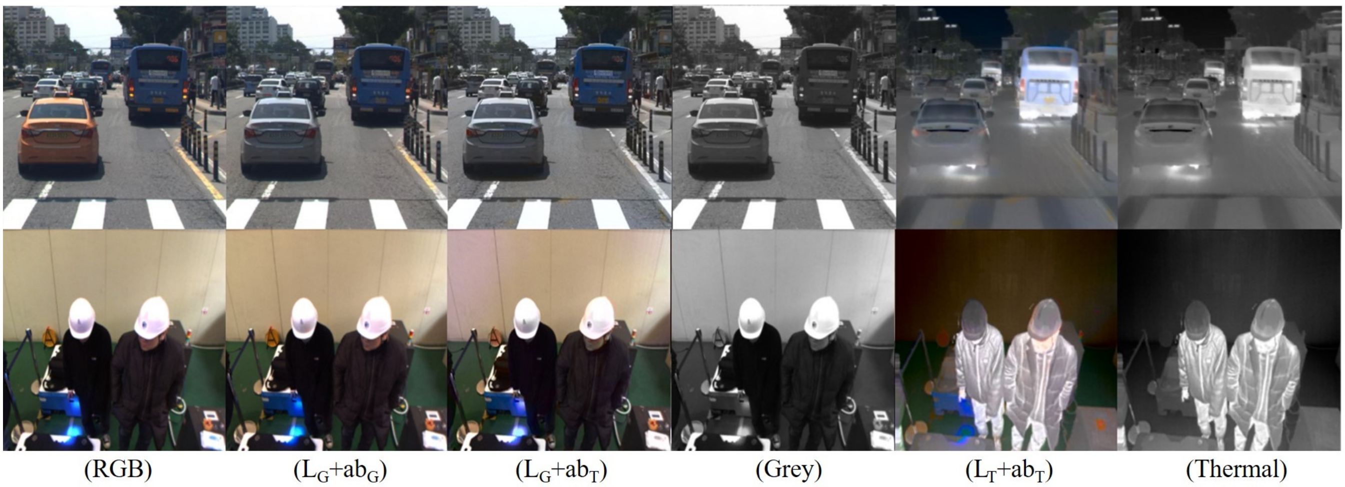

Fig. 3: **Quantitative results of thermal infrared colorization.** ab indicates the estimated color information and subscripts indicates the input source of the estimated color.

C. Evaluation on Pedestrian Detection

Likewise, evaluation on depth estimation, we feed $L_G + ab_T$ as input to pedestrian detection model (SSD [16]) trained with real RGB images. We report the quantitative results according to the input image in Table IV. We observed the extreme performance gap between RGB and thermal images. This gap occurs due to the absence of chromaticity and the difference in contrast of luminance (MAP_{RGB} 95.46 v.s. $MAP_{Thermal}$ 24.61). On the other hand, in the case of $L_T + ab_T$, there was about 40% improvement in MAP performance (24.61% → 62.01 %). Also, when the input is $L_G + ab_T$, the MAP is 92.74, which is almost similar to the result of RGB. This indicates that the effect of chromaticity is critical in the pedestrian detection model.

In addition, it suggests that colorization from the thermal image worked well.

D. Qualitative Results

1) Evaluation on colorization: In Figure 3, result of $L_T + ab_T$ shows that ab_T is strongly affected by L_T. Especially, it can be seen that pedestrian information in (3rd row, 5th column) is not distinguished due to L_T. On the other hand, we observed that color information is clearly expressed at $L_G{+}ab_T$ and $L_G{+}ab_G$. It means that the luminance information of the image is an important factor in the color image composition.

2) Evaluation on pedestrian detection: We show the qualitative results of pedestrian detection according to various inputs in Figure 4. In general, it shows that the detection performance is improved when inputs include chromaticity. In particular, in complex environments where various objects exist around pedestrians, grey (Row 1, Column 1) and thermal images (Row 1, Column 5) are difficult to detect pedestrians. Whereas, images with chromaticity RGB (Row 1,Column 2), $L_G + ab_T$(Row 1,Column 3), $L_T{+}ab_T$(Row 1,Column 4) detect pedestrians accurately. This means that chromaticity is important information for distinguishing objects in complex environments.

3) Evaluation on depth estimation: Figure 5 shows the qualitative result of depth estimation according to the input. First row and second row show the depth estimation result using RGB and $L_G{+}ab_T$. We observed the two results are very similar. Therefore, it can be seen that ab_T has been reconstructed well at pixel levels. Additionally, we prove that the depth estimation shows the result more focused on luminance as $L_T + ab_T$ is not predicted at all in the case of an object such as a car. Looking at the depth 3D visualization in the fourth column, the 3D points with chromaticity are better to recognize the depth and the current situation. This suggests that 3D visualization of thermal images, which is difficult to see, can be changed to be user-friendly.

IV. Conclusions

In this paper, we present the possibility to make a pseudo-RGB image using a thermal infrared image as input. In addition, through the results of thermal colorization studies, we show that the cognitive ability of not only the human visual system but also neural network models such as depth estimation and pedestrian detection can also be improved. These results show the importance of thermal colorization study and that chromaticity has a significant effect on the performance of human- and machine- visual perception. In this paper, the observation we found is that depth estimation is significantly affected by the precision of color estimation compared to pedestrian detection. This is analyzed as a result of reflecting the characteristic of estimating the depth based on the luminance and chromaticity of the local patch, not the global descriptor. Therefore, we found that the luminance, as well as the chromaticity, are important factors to make a perfect pseudo-RGB. In

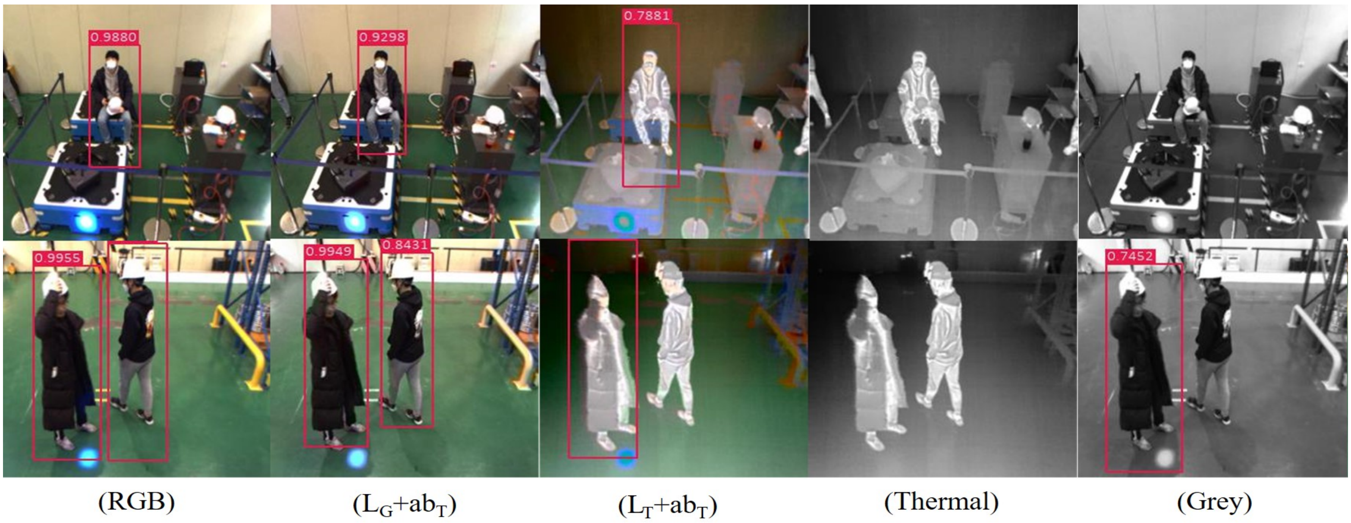

Fig. 4: **Quantitative Results of pedestrian detection.** ab indicates the estimated color information and subscripts indicates the input source of the estimated color.

the future, we plan to further improve the completeness of pseudo-RGB by studying not only color estimation of thermal images but also luminance estimation.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(NRF-2020M3F6A1109603,NRF-2020R1F1A1076987).

REFERENCES

[1] Hwang Soonmin, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. "Multispectral pedestrian detection: Benchmark dataset and baseline." In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2015.

[2] Liu, Qiong, Jiajun Zhuang, and Jun Ma. "Robust and fast pedestrian detection method for far-infrared automotive driving assistance systems." Infrared Physics Technology, vol. 60, pp. 288-299, 2013.

[3] Zheng Wu, Nathan Fuller, Diane Theriault, Margrit Betke, "A thermal infrared video benchmark for visual analysis." in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)-Workshops, 2014.

[4] Davis, James W., and Vinay Sharma. "Fusion-based background-subtraction using contour saliency." in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)-Workshops, 2005.

[5] Torabi, Atousa, Guillaume Massé, and Guillaume-Alexandre Bilodeau. "An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications." Computer Vision and Image Understanding, vol. 116, no, 2, pp. 210-221. 2012.

[6] González Alejandro *et al.*., "Pedestrian detection at day/night time with visible and FIR cameras: A comparison." Sensors, vol. 16, no. 6, p. 820, 2016.

[7] Liu, Jingjing, Shaoting Zhang, Shu Wang, and Dimitris N. Metaxas. "Multispectral deep neural networks for pedestrian detection." arXiv preprint arXiv:1611.02644, 2016.

[8] Konig Daniel, Michael Adam, Christian Jarvers, Georg Layher, Heiko Neumann, and Michael Teutsch. "Fully convolutional region proposal networks for multispectral person detection." in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)-Workshops, 2017.

[9] Li, Chengyang, Dan Song, Ruofeng Tong, and Min Tang. "Illumination-aware faster R-CNN for robust multispectral pedestrian detection." Pattern Recognition, vol. 85, pp. 161-171, 2019.

[10] Wolpert, Alexander, Michael Teutsch, M. Saquib Sarfraz, and Rainer Stiefelhagen. "Anchor-free Small-scale Multispectral Pedestrian Detection." arXiv preprint arXiv:2008.08418. 2020.

[11] Qiao, Yulong, Ziwei Wei, and Yan Zhao. "Thermal infrared pedestrian image segmentation using level set method." Sensors, vol. 17, no. 8, p. 1811, 2017.

[12] D. Feng *et al.*., "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges." in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 3, pp. 1341-1360, 2020

[13] Y. Sun, W. Zuo, P. Yun, H. Wang and M. Liu, "FuseSeg: Semantic Segmentation of Urban Scenes Based on RGB and Thermal Data Fusion." in IEEE Transactions on Automation Science and Engineering, 2020.

[14] Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." in Proceeding of European conference on computer vision (ECCV), 2016.

[15] Su, Jheng-Wei, Hung-Kuo Chu, and Jia-Bin Huang. "Instance-aware image colorization." in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.

[16] Yuxin Wu and Alexander Kirillov and Francisco Massa and Wan-Yen Lo and Ross Girshick. "Detectron2" https://github.com/facebookresearch/detectron2. 2019.

[17] Zhang, Richard, *et al.*. "Real-time user-guided image colorization with learned deep priors." arXiv preprint arXiv:1705.02999, 2017.

[18] Godard, Clément, Oisin Mac Aodha, and Gabriel J. Brostow. "Unsupervised monocular depth estimation with left-right consistency." in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.

[19] Liu, Wei, *et al.*"Ssd: Single shot multibox detector." in Proceeding of European conference on computer vision (ECCV), 2016.

[20] Kim, Namil, Yukyung Choi, Soonmin Hwang, and In So Kweon. "Multispectral transfer network: Unsupervised depth estimation for all-day vision." in Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2018.

[21] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2018.

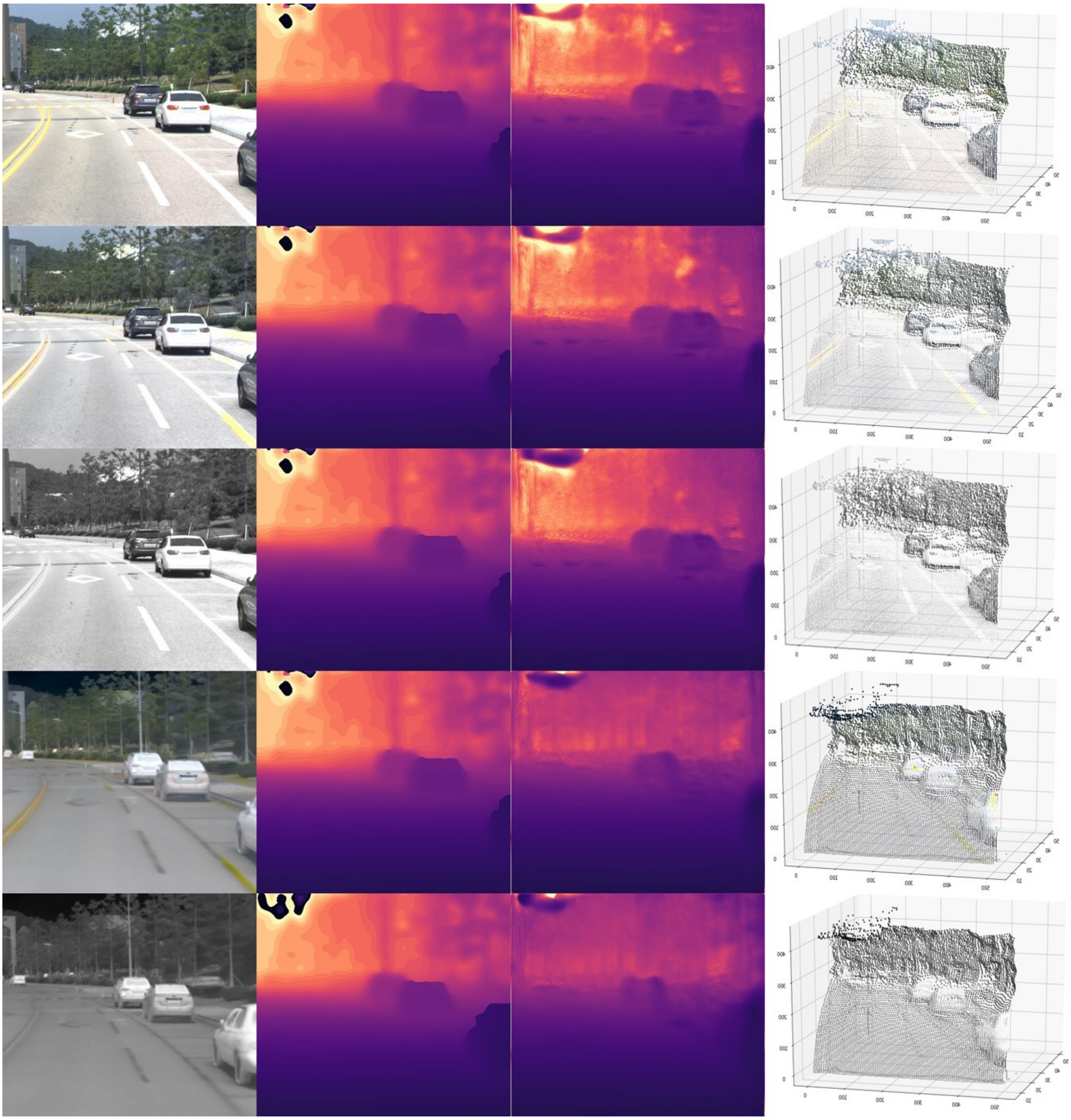

Fig. 5: **Quantitative results of depth estimation.** (From left to right: input image, ground truth, estimated depth, 3D visualization of estimated depth), (From up to down: RGB image, $L_G + ab_T$ image, Grey image, $L_T + ab_T$ image, Thermal image)