Multispectral Pedestrian Detection: Benchmark Dataset and Baseline

Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, In So Kweon
Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea.

Figure 1: Examples of proposed *multispectral pedestrian dataset*. It consists of aligned color-thermal image pairs for day and night traffic scenes. The annotations provided with the dataset such as green, yellow, and red boxes indicate no-occlusion, partial occlusion, and heavy occlusion respectively.

	Training		Testing		Properties								
	# pedestrians	# images	# pedestrians	# images	# total frames	occ. labels	color	thermal	moving cam.	video seqs.	temporal corr.	aligned channels	publication
Caltech [4]	192k	128k	155k	121k	250k	✓	✓		✓	✓	✓		'09
KITTI [1]	12k	1.6k	–	–	80k	✓	✓		✓	✓			'12
LSI [2]	10.2k	6.2k	5.9k	9.1k	15.2k			✓	✓	✓			'13
ASL-TID [5]	–	5.6k	–	1.3k	4.3k			✓		✓			'14
TIV [7]	–	–	–	–	63k			✓		✓			'14
OSU-CT [3]	–	–	–	–	17k		✓	✓		✓		✓	'07
LITIV [6]	–	–	16.1k	5.4k	4.3k		✓	✓		✓		✓	'12
Ours	41.5k	50.2k	44.7k	45.1k	95k	✓	✓	✓	✓	✓	✓	✓	'15

Table 1: Comparision of several pedestrian datasets. The proposed dataset is largest color-thermal dataset providing occlusion labels and temporal correspondences captured in a regular traffic scene.

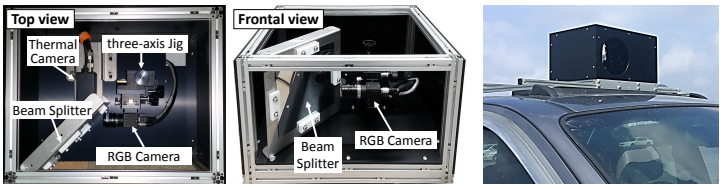

Figure 2: Our hardware capturing aligned color-thermal image pairs.

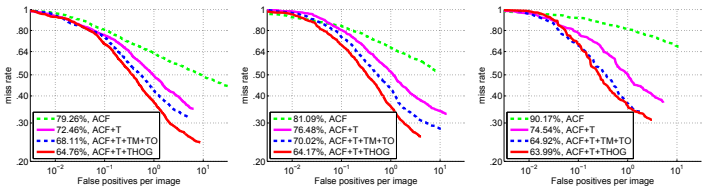

Figure 3: From left to right, three figures show pedestrian detection performance on the day&night, day, and night traffic scenes. ACF (green curve) indicates color based detection algorithm, and other curves indicate color-thermal based detection algorithms.

Pedestrian detection is active research area in the field of computer vision. Although various methods have been studied for a long time, pedestrian detection is still regarded as a challenging problem, limited by tiny and occluded appearances, cluttered backgrounds, and bad visibility at night. In particular, even though color cameras have difficulty getting useful information at night, most of the current approaches are based on color images.

To address this limitation, one possible way is to utilize additional information from another spectral band such as infrared. Among near infrared ($0.75\sim1.3\mu$m) and long-wave infrared ($7.5\sim13\mu$m, also known as the thermal band) camera, we used a long-wave infrared camera rather than near infrared cameras. Physically, living things such as human radiate heat, e.g. long-wave infrared signal. Thus, pedestrians are more visible in long-wave infrared cameras than in near infrared cameras.

Based on these facts, we introduce a *multispectral pedestrian dataset* which provides thermal image sequences of regular traffic scenes as well as color image sequences. In constrast to most previous datasets utilizing a color-thermal stereo setup, we use beam splitter-based hardware (shown in Fig. 2) to physically align the two image domains. Therefore, our dataset is free from parallax and does not require an image alignment algorithm for post processing. Examples of our dataset with annotations are shown in Fig. 1. A survey on the previous datasets are summarized in Table 1.

Our contributions are threefold: (1) We introduce the multispectral pedestrian dataset, which provides aligned color and thermal image pairs. Our dataset has number of image frames as large as widely used pedestrian datasets [1, 4]. The dataset also contains nighttime traffic sequences which are rarely provided or discussed in previous datasets. (2) We analyze the complementary relationship between the color and thermal channels, and suggest how to combine the strong points of the two channels instead of using the color or thermal channel independently. (3) We propose several baselines to handle multispectral images and analyze the performance. One of our baseline reduces the average miss rate by 15% on the proposed multispectral pedestrian dataset.

Through the experiments, we determined that the aligned multispectral images are very helpful for improving pedestrian detection performance in various conditions (shown in Fig. 3). We expect that the proposed dataset can encourage the development of better pedestrian detection methods.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage. Our multispectral pedestrian dataset is available in our project web page: http://rcv.kaist.ac.kr/multispectral-pedestrian/

[1] P.Lenz A.Geiger and R.Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[2] U. Nunes J.M. Armingol D. Olmeda, C. Premebida and A. de la Escalera. Pedestrian classification and detection in far infrared images. *Integrated Computer-Aided Engineering*, 20:347–360, 2013.

[3] J. Davis and V. Sharma. Background-subtraction using contour-based fusion of thermal and visible imagery. *Computer Vision and Image Understanding*, 106(2–3):162–182, 2007.

[4] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. 2009.

[5] M. Chli J. Portmann, S. Lynen and R. Siegwart. People detection and tracking from aerial thermal views.

[6] A. Torabi, G. MassÃl', and G.-A Bilodeau. An iterative integrated framework for thermal-visible image registration, sensor fusion, and people tracking for video surveillance applications. *Computer Vision and Image Understanding*, 116:210–221, 2012.

[7] D. Theriault Z. Wu, N. Fuller and M. Betke. A thermal infrared video benchmark for visual analysis. In *Proceeding of 10th IEEE Workshop on Perception Beyond the Visible Spectrum (PBVS)*, 2014.