2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Daejeon Convention Center
October 9-14, 2016, Daejeon, Korea

Thermal Image Enhancement using Convolutional Neural Network

Yukyung Choi*, Namil Kim*, Soonmin Hwang* and In So kweon

***Abstract*— With the advent of commodity autonomous mobiles, it is becoming increasingly prevalent to recognize under extreme conditions such as night, erratic illumination conditions. This need has caused the approaches using multi-modal sensors, which could be complementary to each other. The choice for the thermal camera provides a rich source of temperature information, less affected by changing illumination or background clutters. However, existing thermal cameras have a relatively smaller resolution than RGB cameras that has trouble for fully utilizing the information in recognition tasks.**

To mitigate this, we aim to enhance the low-resolution thermal image according to the extensive analysis of existing approaches. To this end, we introduce *Thermal Image Enhancement using Convolutional Neural Network (CNN)*, called in *TEN*, which directly learns an end-to-end mapping a single low resolution image to the desired high resolution image. In addition, we examine various image domains to find the best representative of the thermal enhancement. Overall, we propose the first thermal image enhancement method based on CNN guided on RGB data. We provide extensive experiments designed to evaluate the quality of image and the performance of several object recognition tasks such as pedestrian detection, visual odometry, and image registration.

***Index Terms*— Thermal camera, Image enhancement, Convolutional Neural Network (CNN), Object Recognition, Visibility, Multi-modal**

I. INTRODUCTION

Object recognition is a vital problem of modern robotics researches due to paramount relevance in commercial systems, spanning from self-driving cars to autonomous robots. Recent advancements in RGB sensor technologies and algorithms have encouraged the performance in reasonable scenarios. However, objects can occur in varying conditions of illumination, weather and occlusions. These variations still make object recognition based on RGB cameras a challenging problems. The current state of the art in object recognition includes methods which involve multi-modal supervision. Multi-modal supervision entails the complementary information of other spectral sensor about present objects of interest in certain conditions.

Thermal camera has been one of a promising choice, which can provide the temperature information in complex scenarios with background clutters or lack of illumination. In such scenarios, the thermal camera is more prominent in object detection and scene understanding compared to the RGB camera. Therefore, thermal sensors are an increasingly found in modern robotic systems and researches [1] [4] [5] [6]. While advances of commercial thermal devices, most thermal cameras used in previous researches have lower resolutions than RGB cameras, and this lower resolution introduce significant challenges to span a wide field of applications.

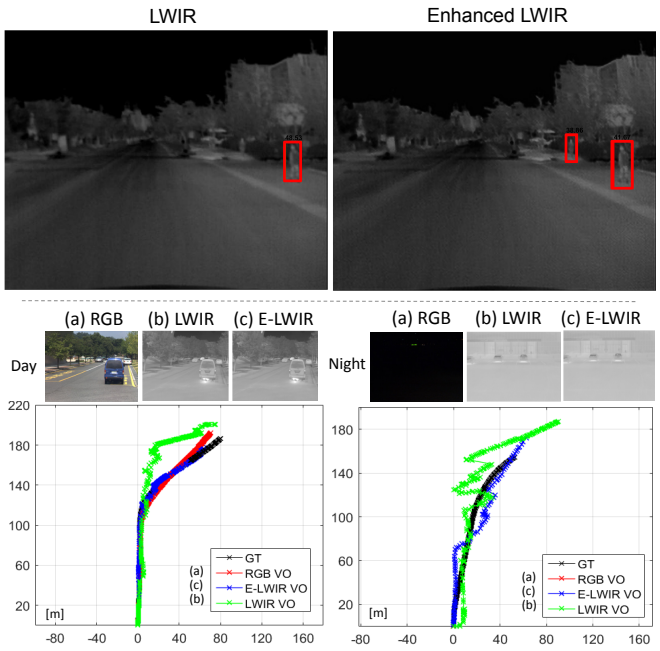


Fig. 1: This figure shows the effect of *TEN*. The results can improve not only visibility, but also the performance of various vision tasks. From our method, small person is detected and feature extraction is more reliable.

Over the decades, such problems has always been considered an important issue in RGB domain to efficient enlarge the low resolution input to the high resolution output, as known as image enhancement. Therefore, image enhancement has been widely used in applications ranging from security, surveillance, medical imaging to mobile platform, forensic science, which requires better image qualities. There are various approaches to deal with image enhancement from a conventional interpolation, Lanczos resampling [7] to internal/external similarity [8] [9], learning based method such as sparse representation [10] [11], random forest [12]. Since SRCNN [13] successfully introduced a deep learning technique into the image enhancement problem, deep learning based approaches have been used with large improvements in performance [14]. Moreover, not merely visibility, the enhancement has been proved to help the visual recognition tasks such as face recognition [15] [16], 3D modeling [17],

Y. Choi, N. Kim, S. Hwang and I.S Kweon are with Robotics and Computer Vision (RCV) Laboratory, Dep. of EE, KAIST, Daejeon, 305-701, Korea {ykchoi,nikim,smhwang}@rcv.kaist.ac.kr and iskweon@kaist.ac.kr.
*All authors contributed equally to this work.

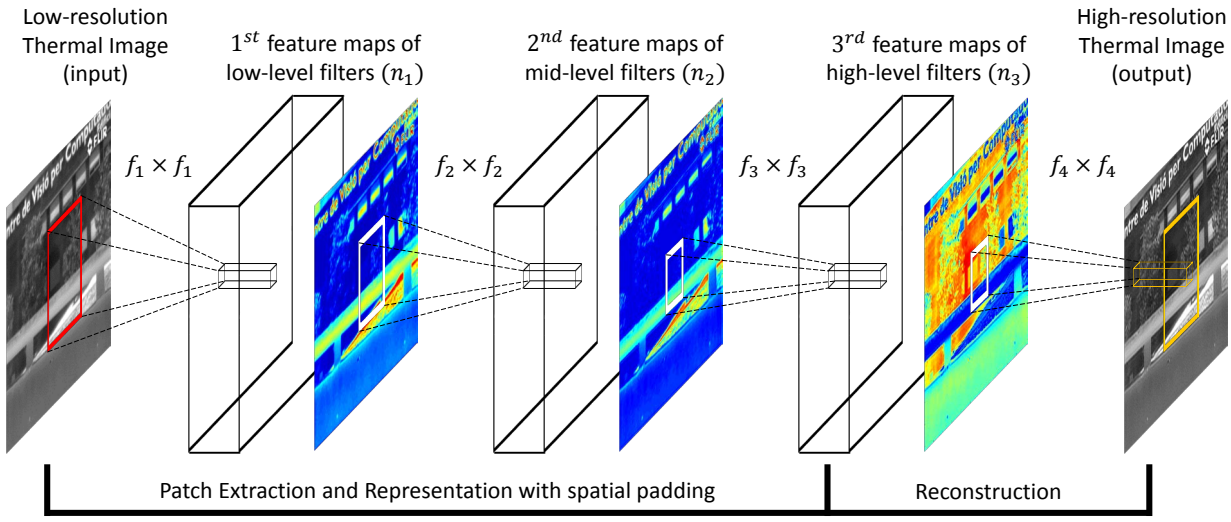


Fig. 2: Given a low-resolution thermal image, the first to third convolutional layers extract a set of feature maps. The last layer combines the predictions within a spatial neighborhood to reconstruct a high resolution image. In every layer, we pad the feature maps to preserve the size of original input image. Color activation indicates a max-pooled feature map in each layer. The model architecture is as follows: $C_1(64,7,3) - ReLU_1 - C_2(32,5,2) - ReLU_2 - C_3(32,3,1) - ReLU_3 - C_4(1,3,1)$, where $C_i(n,k,p)$ is a n filters of spatial size f x f applied with padding p.

SLAM [18], and robot navigation [19] by applying image enhancement algorithms.

Unlike the success of RGB-based approaches, the understanding of the thermal image[1] enhancement is still relatively unexplored in computer vision and robotics societies. There are only few techniques to heuristically enhance the low-resolution thermal image such as the manually tuned various camera parameters [21], the basic image processing techniques (histogram equalization) [1].

In this paper, we consider the thermal image enhancement problem. This is primarily inspired by recent advances in RGB-based approaches, which use convolutional neural networks. Firstly, we examine the domain selection as the best representative to enhance the thermal image. Note that the training data domain can be crucially affected in the quality of the enhanced results. After the selection, we compactly design the networks to yield the high quality of output, and to be used as practical usage in lower computational environment such as CPU. Lastly, we provide the extensive experiments on public thermal benchmarks dataset [1] [20] with various object recognition tasks: pedestrian detection, visual odometry, and image stitching with feature correspondence to prove the visibility and the usefulness of our enhanced results. To our best knowledge, we firstly adjust a concept of thermal image enhancement to deep neural network framework. We named the proposed network as the *Thermal Enhance Network (TEN)*[2].

II. Thermal Enhancement Network

A. Proposed Network

For thermal image enhancement, we designed a relatively shallow convolutional neural network inspired by [13]. While the deeper neural network shows the better performance for image enhancement, it requires much memory capacity on a high specification. Our network is relatively lightweight structure enough to operate in CPU environments as a practical usage.

The configuration of our proposed network is outlined in Fig.2. Given a single low resolution image, we first crop the entire image into the desired size with a uniform stride and then upscale using bi-cubic interpolation. Denote the interpolated image as X. Our goal is to recover from X a reconstructed image Y which is as similar as possible to the original high quality ground truths G. Our network has two main components: *Patch Extraction / Representation* and *Reconstruction*.

The input X is convolved by a set of filters. This operation allows to comprise a set of feature maps, which are originated from extracted overlapping patches from X. Formally, these layers are expressed as an operation below:

$$F(X) = W_3 * (W_2 * (W_1 * X + B_1) + B_2) + B_3 \qquad (1)$$

, where W_i and B_i indicate the trainable filters and biases respectively, and $*$ means the convolution operation. Here, we use total 3 convolutional layers for this purpose. The first convolutional layer has a 64 (n_1) filter of the size 7×7 (f_1), where 64 feature maps are generated by 7×7 spatial convolution. The second and third layers have 32 (n_2) filter of the 5×5 (f_2) spatial region and 32 (n_3) filter of the 3×3 (f_3) spatial region respectively. We apply the Rectified Linear Unit (ReLU) on the filtered output of each convolutional layer.

[1] Note that the thermal image we targeted is LWIR (Long-Wavelength InfraRed; $8-14um$)

[2] The dataset, demo code and high-resolution paper can be downloaded in the webpage: **https://sites.google.com/site/ykchoicv/ten**

Similar to the [14], we pad zeros before convolutions to keep the sizes of all feature maps the same as that of input. Generally, when convolution operations are applied, the size of the feature map becomes smaller in every time. For example, when an input of size $N \times N$ is applied to a network with $K \times K$ spatial convolution filter, stride S, and padding size P, the output size M is calculated below:

$$M = \frac{(N + 2 \times P - K)}{S} + 1. \quad (2)$$

This reduction can cause the bias to concentrate on recovery of center regions, while surrounding region cannot correctly be restored. Therefore, in contrast with [13], we assign the padding size P according to the size of filters K to keep the same dimension between a low quality input and a high quality output. Here, P_1, P_2, and P_3 are 3, 2, 1, respectively. With this strategy, our network can effectively reconstruct a high quality image through overall regions.

Next, the stacked high level features produce a final high quality image in last convolutional layer. In the view of traditional methods, this layer has a role as the weighted averaging pooling of a set of high-level feature maps.

$$Y = W_4 * F(X) + B_4 \quad (3)$$

B. Training

To reconstruct an enhanced image in end-to-end manners, we should properly choose the objective function to find optimal parameters $\theta = [W_j, B_j]$. Given a training pair $[X^i, G^i]$, our goal is to minimize the pixel-level distance between an estimate and a ground truth G. We use a Mean Square Error (MSE) as the object function below:

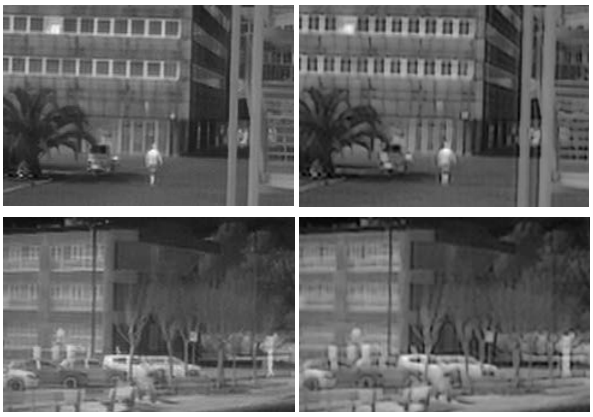
$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} ||Y^i - G^i||^2. \quad (4)$$

The optimization process is done by stochastic gradient descents with the general back-propagation scheme [22].

When the training is proceeded, we simultaneously measure the PSNR (Peak Signal-to-Noise Ratio), which is a well-known metric for quantitatively evaluating image restoration quality and the perceptual quality to check the progress of learned reconstruction results. Even though, the PSNR measurement is a very suitable metric toward what we aim to learn, it is not easy to be used as the objective function, because of differentiable matters [13].

C. Domain Selection

For our purpose, we require pairs of low- and high-resolution thermal image to train the model, because CNN-based method currently rely heavily on the availability of resources. However, there is only few thermal dataset, and even many of them is made up low resolution thermal images, not in pairs, because the high quality of thermal camera is still expensive to be generally used as a research purpose. Also the thermal measurement from creatures can be easily changed by occlusions and surroundings and is not sharp or evident rather than other domain images. This variations can be disturbed to reconstruct the desired output.



(a) RGB (b) MWIR

Fig. 3: Qualitative results on RGB- and MWIR- based model.

MDS benchmark [20] (scale factor x2)			
PSNR	Bi-cubic	RGB-*TEN*	MWIR-*TEN*
(dB)	39.2000	<u>40.8257</u>	33.3964

TABLE I: Qualitative results on RGB- and MWIR- based model.

Dataset	Algorithm			
	scale factor x2		scale factor x3	
	Bi-cubic	*TEN*	Bi-cubic	*TEN*
TSD [23]	39.2000	40.8257	35.4164	37.7097
KAIST [3]	42.9740	44.2124	39.6617	40.4814
MSD [20]	40.9116	42.7314	36.3489	37.7351

TABLE II: Evaluation results of PSNR (dB) on several dataset.

Naturally, other spectrum domain is a strong candidate for an alternatives to train model. Fortunately, there are many pairs of dataset in other spectrum domains to be designed as image enhancement tasks. Moreover, we assume that lower-spectrum domain can be more helpful to enhance the sharpness and contrast of thermal images, because of their characteristics of spectrum according to the wavelength. Therefore, we comprise two types of domains, RGB (*RGB 91*) [11] and MWIR (*ThermalStereo*) [23], which have relatively lower wavelength than our target domain, having sharper gradient information. For the fair comparison, we trained the proposed model with same parameters and conditions, and we tested *MDS* benchmark [20], including 100 thermal images. The bi-cubic interpolation is used as the baseline and we measured the PSNR averaged over all testing images.

As shown as quantitative evaluation in Table I, it turns out that the RGB-based model surprisingly well in thermal domain enhancement, showing the better performance than baseline method, whereas the MWIR-based model does not properly recover the desired output and even it could not

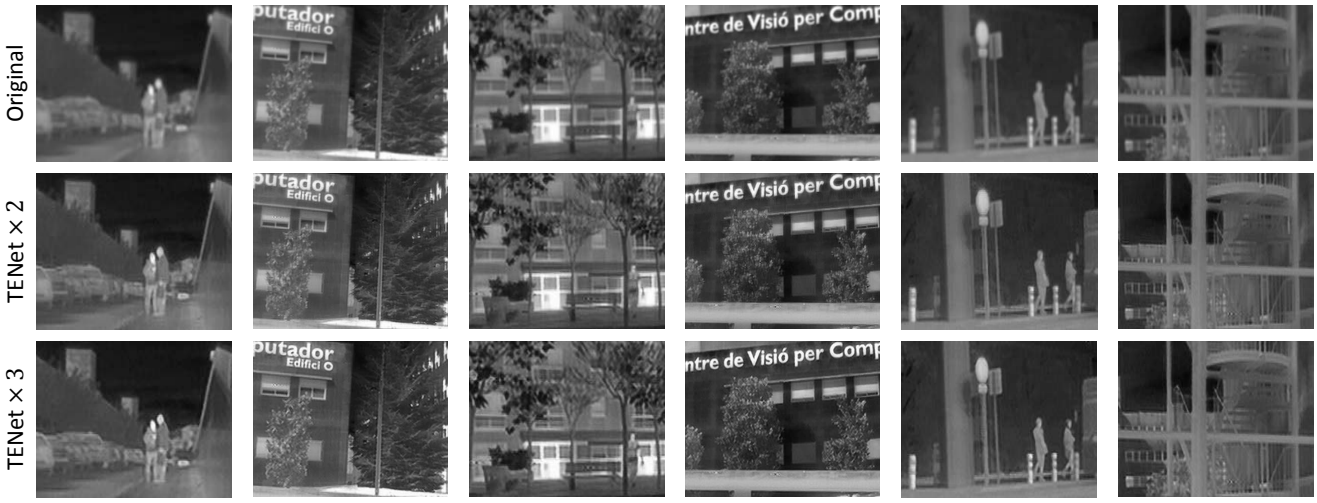


Fig. 4: Thermal image enhancement results of TEN with scale factors $\times 2$ and $\times 3$.

reach the PSNR of the baseline. We can show the same situation in qualitative results. Fig. 3 depicts the output of each domain models to restore the same input data. The RGB outputs seems to clean and sharp, while the results of MWIR are not visually pleasing. Moreover, contexts of MWIR are distorted as dilation, bending, and it seem to spread out to the surrounding region of the object. From this observation, we decide to train the model using RGB images, which is the basic concept of our RGB-guided image enhancement.

D. Analysis

In this section, we evaluate the enhancement performance of our proposed approach on several benchmarks. We first briefly describe dataset used for training and testing our method. Next, we explain parameter settings for training to be reproduced. Lastly, we analyze the experiment results.

1) Training dataset: As above mentioned, we use the RGB training dataset, consisting of *RGB 91* images from Yang et al. [11]. In this experiment, we did not augment original dataset such as flipping, rotation for the general purpose. We train two types of scale network as factor 2 and 3. In the training phase, the ground truth is prepared as 36×36 patches, which are allowed in 6 pixel overlapping of adjacent patches. To synthesize the low resolution patches, we intentionally blur a patch by a proper Gaussian kernel according to the scale factor, and upscale it by the same factor through bi-cubic interpolation. In contrast with previous works [13] [14] which consider the luminance channel in YCbCr color space, we only used the gray-channel in our experiments because the thermal image cannot be converted to the YCbCr color space as the RGB. Note that different channels can be affected in quantitative results [13], however, we do not compare the these different in this paper.

2) Test dataset: For benchmark, we use three datasets: ThermalStereo (TSD) [23], KAIST-Multispectral [3], MultimodalStereo (MSD) [20]. All dataset are explained into details in following sections. Similar to the training dataset, we prepared as 36×36 patches with 6 stride and two scale factors. The validation of training and test dataset is measured by the PSNR averaged over the target dataset.

3) Parameters: We provide parameters used to train and reproduce our network model. We use a network of depth 4 and the size of batch as 128. Momentum and weight decay parameters are set to 0.9 and 0.0005, respectively. For weight initialization, we use the method described in [24]. We train all experiments over 100 epoch. Learning rate started to 0.001 and decreased by a factor 10 every 30 epochs until 60 epochs. We implement our model using the customized MatConvNet library [25].

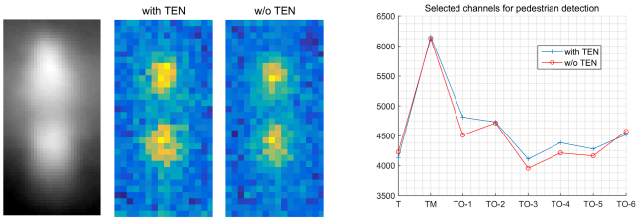
4) Results: As shown in Table II, the proposed *TEN* yields the better average PSNR than baseline in all experiments. Considering that PSNR measurement has log-scale unit, our method achieve the high quality output more than numerical value increased (see Fig. 4). The blurry boundaries of object in original thermal image almost enhance the results of *TEN* in all scale factors. In particular, our network model only sharpens the object boundaries and details with minimizing pixel noises, smoothing, intensity artifacts such as spot, spread, which are usually occurred in general image processing techniques such as histogram equalization, de-noise.

III. EXPERIMENTS

In this section, we verify the effectiveness of our enhancement method in several computer vision tasks. We select three representative tasks: pedestrian detection for high-level recognition task, visual odometry for mid-level task, and image registration for low-level task.

A. Pedestrian Detection

First of all, we present experiments for pedestrian detection on thermal image. The experiments are conducted



(a) Selected spatial locations (b) Selected channels

Fig. 5: Analyses for trained models. (a) Average of pedestrian patches and drawings for learned discriminative features. Proposed method helps detector focus on valid regions. (b) Learned discriminative channels. Since our method enhance the gradients, more gradient parts are selected.

miss rate	Near	Medium	Far
None-occ	45.1% (-1.9%)	71.9% (-3.3%)	96.7% (-0.6%)
Partial-occ	58.4% (+3.6%)	84.8% (-2.2%)	95.1% (-0.7%)

TABLE III: Evaluation on various subsets[3]. The *TEN(x2)* is applied both at training and testing phases. Numbers in parentheses represent performance gain from the baseline [1] Negative value is preferred since the measure is miss rate.

on KAIST Multispectral Pedestrian Benchmark [1] which is a most recent benchmark for thermal domain. We analyze that how the improvement of low-level information (enhancement of visual quality) affects the performance of high-level task (pedestrian detection performance).

1) Dataset and Library: The KAIST Benchmark is the largest and most challenging dataset for thermal domain. It consists of 95k frames on urban traffic environment and dense annotations for 1,182 unique pedestrians. Our baseline detector is from the public code of [1], but as our scope in this paper is limited to thermal image we only use thermal information for channel features. So, our features have totally 8-channels: normalized thermal intensity (1ch), normalized gradient magnitude (1ch), and histograms of oriented gradients (6ch). Note that this condition is different from the final results of [1], which fully utilize the RGB and thermal information together. We train the detectors on Set00-Set05 and the trained detectors are evaluated on Set06-Set11 with sampled images at 20Hz.

2) Effect of the proposed method for pedestrian detection: For object detectors, sharpness of object boundary is an important evidence for existence of object. As shown in Fig. 6-(Left), proposed method enhances the shape of objects. This enhancement could be useful for detectors to classify pedestrians accurately. Unfortunately there are several failure cases (Fig. 6), for which the shape information is not magnified properly.

[3]Near: height $\geq 115pixel$, Medium: $115pixel \geq height \geq 45pixel$, Far: $45pixel \geq height$, Partial-occ: $50\% \geq occlusion$ excluding none-occlusion cases, Heavy-occ: $occlusion \geq 50\%$

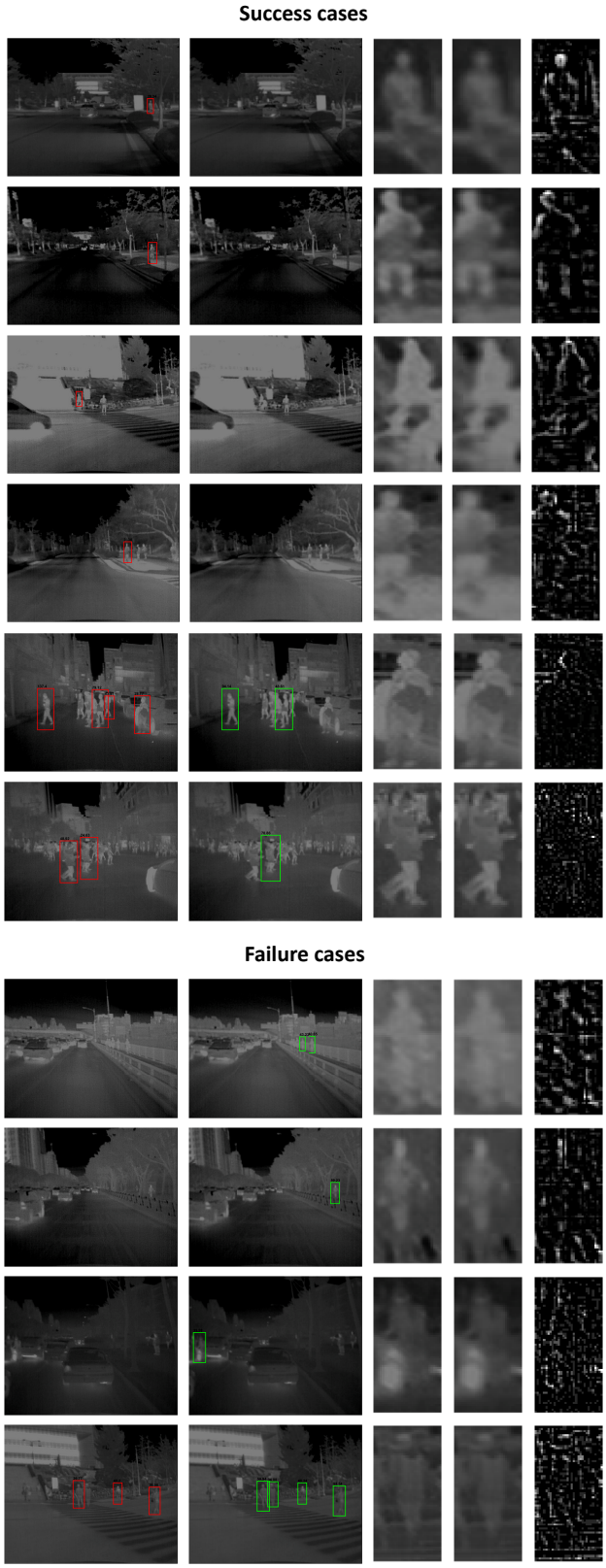


Fig. 6: Examples of detection results. (from left to right) Detect with the TEN(x2), Detect w/o the TEN, Enhanced patch, original patch, difference between enhanced and original patches.

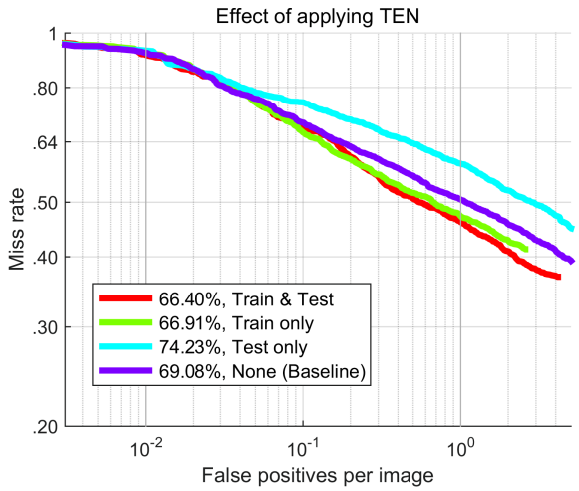


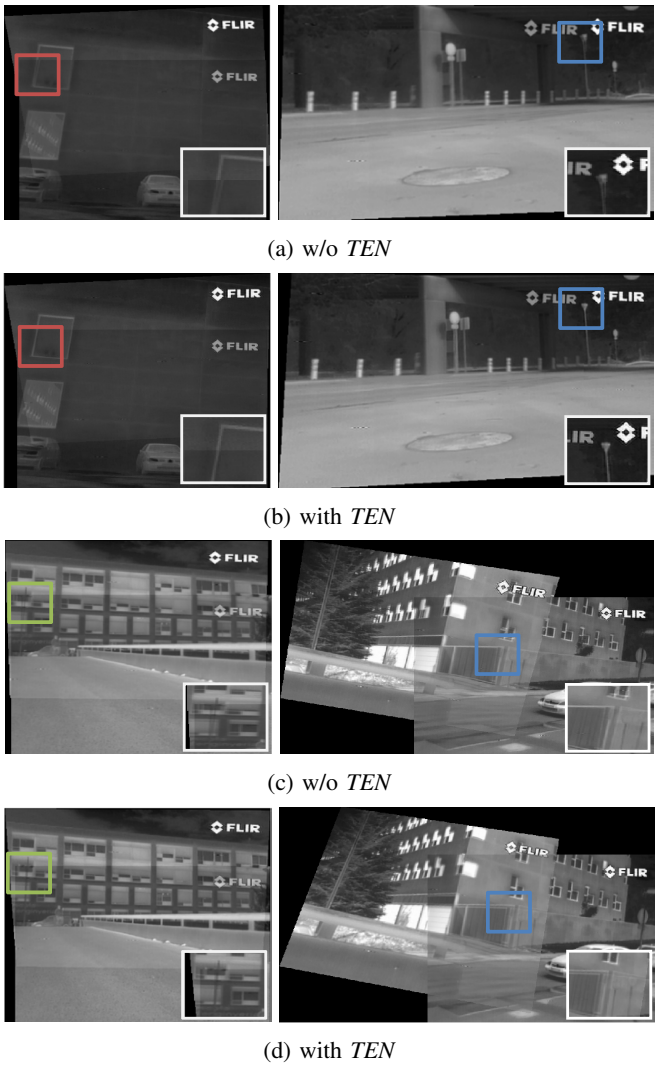
Fig. 7: Performance evaluation on KAIST-Multispectral Pedestrian Dataset [1], reasonable subset[4]. The numbers in the graph are log-averaged miss rate on FPPI $\in [10^{-2}, 10^{0}]$. The proposed method has an effect on the performance in training phase.

3) Performance improvement: For evaluation criterion, we follow the standard measure for pedestrian detection called log-average miss rate on FPPI $\in [10^{-2}, 10^{0}]$ [27]. As shown in Fig. 7, the miss rates are decreased with our enhancement in most cases. An interesting thing is that when we apply our method only at the testing phase, it degrades on the performance. In addition, if we apply proposed method at training phase, it shows good performance. These phenomenon shows that the enhancement of visual quality is more effective at training phase to alleviate ambiguous regions. Furthermore, proposed enhancement method helps to detect pedestrians in most cases of various scenarios with respect to occlusions and distance from the camera (Table. III).

4) Analyses for trained models: To check the validity of the trained detectors, we draw a voting map for the spatial locations of selected features [1] [26]. As we consider spatial occupancy of general pedestrians in bounding boxes (Fig. 5a-(Left)), our enhancement helps to learn more reasonable regions, i.e. on the human part, as discriminative part. This characteristic is an evidence of well-trained detector. Furthermore, we check what kinds of channels are preferred to the detectors. In Fig. 5b, we can see that edge-related information is selected more after applying proposed enhancement method. From these analyses, we show that the proposed approach has a potential to boost the performance of detection algorithms by training more discriminative parts of objects.

B. Image Registration

First of all, we test a qualitative analysis for blob-feature properties and show stitching results as a quantitative evaluations to prove the effectiveness of our method.



(a) w/o *TEN*

(b) with *TEN*

(c) w/o *TEN*

(d) with *TEN*

Fig. 8: Qualitative evaluation with TEN(x2). Results show some examples for image stitching. Though cropped regions from (a,c) are slightly mis-aligned, ours from (d,b) show good results.

1) Dataset and Library: Multimodel Stereo Dataset (MSD)[5] consists of 100 pairs of outdoor VS-LWIR images collected from various urban environments by a stable stereo rig. In this experiment, we only used 8 bits quantized thermal images, captured by FLIR PathFindIR with 19mm focal lens. Since MSD does not provide geometric relationship between image pair, we carefully select 47 pairs for experiments. We use a popular libraries as *vlfeat* library [28] for SIFT features and the parameters are set as follows: *octable = 1, Levels = 30, PeakThresh = 0, EdgeThres = 20*. These settings are intended for extracting features as many as possible.

2) Effect of the proposed method for image stitching: We can show that our proposed method can make original

[4]Reasonable subset: up to partial occlusion, $height \geq 55 pixel$

[5]http://adas.cvc.uab.es/projects/simeve/index539a.html?q=node/2

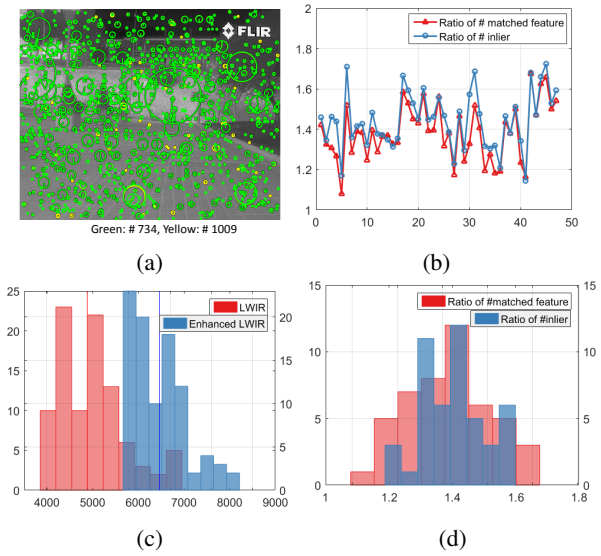


Fig. 9: Quantitative evaluation with TEN(x2). Results show that extracted features are reliable and good features to match. (a) shows extracted features in LWIR (original image) and E-LWIR (enhanced image). Circles indicate keypoints and the size of circle represents scales of keypoints. (b) shows a ratio of # matched feature to # inlier in each test pairs. (c) shows a histogram of # extracted feature in LWIR and E-LWIR. (d) shows a histogram of the ratio of # feature to # inlier.

images to enhance the extracting more reliable blobs. Fig. 9-(a) is an example of extracted blobs on the enhanced thermal image. As shown in Fig. 9-(c), the number of extracted features are increasing applying *TEN*. Moreover, the ratio of the number of matched feature and the number of inlier have a similar distribution in the most of image sequences in Fig. 9-(b). It means that the enhanced thermal image can increase not only the number of matched features, but also the number of inlier. Note that if the number of unreliable features are increasing, the inlier's ratio does not have a similar tendency of the matched feature's ratio. Fig. 9-(d) shows that the histogram, containing the ratio between the number of features which are passed in matching ratio test and the number of features which are passed in RANSAC.

3) Performance improvement: To prove the usefulness of the enhancement to image stitching tasks, we conduct the procedure of the conventional mosaic method. As shown in Fig. 8, enhanced image is accurately matched due to the estimated homography from reliable blobs. However, there is still problem in texture-less regions, where our method can rarely be covered.

C. Visual Odometry

In this paragraph, we prove that the enhanced thermal image can be helpful for visual odometry task. Experiments are conducted on the subset of KAIST-Multispectral Dataset [3], and the environment is as following. We first extracted

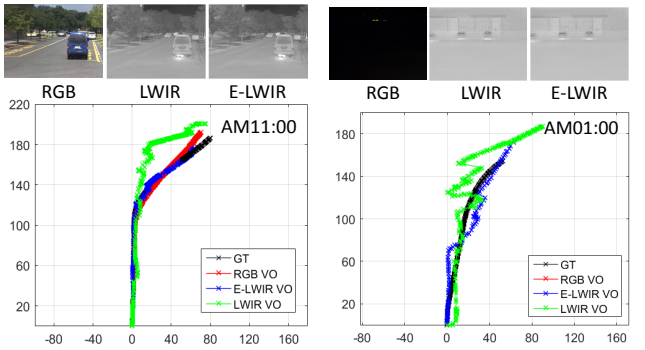


Fig. 10: Trajectories estimated using visual odometry [29] with TEN(x2). Unit of each-axis is meter and zero coordiante (0,0) indicates a starting point in this test. The result obtained by E-LWIR shows the right path at day and night.

corner points to estimate the relative camera pose within successive frames in every sequence, and evaluate the estimated parameters according to the criteria [29].

1) Dataset: The KAIST-Multispectral benchmark [3] is the extension of [1] [2], which is made up a large-scale and high-quality of multi-modal data with RGB, thermal, 3D Lidar, and inertial sensor (GPS/IMU) in all-day conditions. For the fair comparison, we collected the results of visual odometry using each spectrum domains. Since thermal images are provided in 16 bits format, we converted to 8 bits level to test and visualize the data. Note that we use a clipped linear mapping approaches for the contrast enhancement of thermal image as a pre-processing.

2) Effect of the proposed method for visual odometry: As a result of image registration, the enhancement can increase the number of reliable corner points in each frames. Therefore, the estimated camera pose is more accurate than estimate of raw-thermal images, and the trajectory can be stably built in any conditions, as similar to the ground truth.

3) Performance improvement: We denote RGB based Visual Odometry as RGB VO, thermal based Visual Odometry as LWIR VO, enhanced thermal based Visual Odometry as E-LWIR VO, and ground truth as GT. In Fig. 10, RGB VO shows the better results in day scenario, whereas at night it could be failed to estimate the right camera pose due to illumination deficiency. LWIR VO estimates the better parameters than RGB VO at night, but it is still an inexact trajectory, because the contextual information such as edge, boundary is not observable to extract the reliable corner points in raw-thermal images. Unlike other results, E-LWIR VO can follow the right path at day and night, because of the growing number of reliable feature points.

IV. CONCLUSIONS

In this work, we have addressed the task of the image enhancement on general thermal data with a low resolution. We have analyzed that other domain datasets can be useful to enhance the thermal image. The analysis motivates us to propose a novel thermal image enhancement method using convolutional neural network with the concept of RGB-guidance. We have demonstrated that our method has the powerful capacities to enhance visibility and to improve the thermal-based robotics applications such as pedestrian detection, visual odometry and image registration. We believe that our approach is readily applicable to other applications such as tracking, stereo matching and 3D reconstruction. In the future we are interested in the integrating multiple scale networks into single network and training an aligned RGB and thermal dataset to improve the performance.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by Korea government (MSIP) (No.2010-0028680), and partially supported by the Development of Autonomous Emergency Braking System for Pedestrian Protection project funded by the Ministry of Trade, Industry and Energy of Korea (MOTIE) (No.10044775).

REFERENCES

[1] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, Multispectral Pedestrian Detection: Benchmark Dataset and Baseline, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[2] Y. Choi, N. Kim, K. Park, S. Hwang, J. S. Yoon, I. S. Kweon, All-Day Visual Place Recognition: Benchmark Dataset and Baseline, IEEE Conference on Computer Vision and Pattern Recognition Workshop on Visual Place Recognition in Changing Environments (CVPRW-VPRICE), 2015.

[3] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, I. S. Kweon, KAIST Multi-spectral Recognition Dataset in day and night, The International Journal of Robotics Research, *Under review*.

[4] W. Maddern, S. Vidas, Towards Robust Night and Day Place Recognition using Visible and Thermal Imaging, Robotics Science and Systems (RSS), 2012.

[5] S. Y. Cheng, S. Park, M. M. Trivedi, Multiperspective Thermal IR and Video Arrays for 3D Body Tracking and Driver Activity Analysis, IEEE Conference on Computer Vision and Pattern Recognition Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum (CVPRW-OTCBVS), 2005.

[6] T. Mouats, N. Aouf, A. D. Sappa, C. Aguilera, R. Toledo, Multispectral Stereo Odometry, IEEE Transactions on Intelligent Transportation Systems (TITS), 2015.

[7] C. E. Duchon, Lanczos Filtering in One and Two Dimension, Journal of Applied Meteorology, 1979.

[8] G. Freedman, R. Fattal, Image and Video Upscaling from Local Self-Examples, ACM Transactions on Graphics (TOG), 2011.

[9] J. Yang, Z. Lin, S. Cohen, Fast Image Super-resolution Based on In-place Example Regression, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[10] J. Yang, J. Wright, Y. Ma, T. Huang, Image Super-Resolution as Sparse Representation of Raw Image Patches, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[11] J. Yang, J. Wright, T. S. Huang. Y. Ma, Image Super-Resolution Via Sparse Representation, IEEE Transactions on Image Processing (TIP), 2010.

[12] S.Samuel, C.Leistnerand, H.Bischof, Fast and Accurate Image Upscaling With Super-Resolution Forests, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[13] C. Dong, C. C. Loy, K. He, X. Tang, Image Super-Resolution Using Deep Convolutional Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2015.

[14] J. Kim, J. K. Lee, K. M. Lee, Accurate Image Super-Resolution Using Very Deep Convolutional Networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[15] L. Tao, M. Seow, V. K. Asari, Nonlinear Image Enhancement to Improve Face Detection in Complex Lighting Environment, In Proc. SPIE, Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning, 2006.

[16] E. Bilgazyev, U. Kurkure, S. K. Shah, I. A. Kakadiaris, ASIE: Application Specific Image Enhancement for Face Recognition, In Proc. SPIE, Biometric and Surveillance Technology for Human and Activity Identification X, 2013.

[17] R. Maier, J. Stuckler, D. Cremers, Super-Resolution Keyframe Fusion for 3D Modeling with High-Quality Textures, IEEE International Conference on 3D Vision (3DV), 2015.

[18] M. Meilland, A. I. Comport, Super-Resolution 3D Tracking and Mapping, IEEE International Conference on Robotics and Automation (ICRA), 2013.

[19] K. Okarma, M. Teclaw, P. Lech, Application of Super-Resolution Algorithms for the Navigation of Autonomous Mobile Robots, Image Processing and Communications Challenges, 2015.

[20] F. Barrera, F. Lumbreras, A. Sappa, Multispectral Piecewise Planar Stereo using Manhattan-World Assumption, Pattern Recognition Letters (PRL), 2013.

[21] T. Mouats, N. Aouf, L. Chermak, M. A. Richardson, Thermal Stereo Odometry for UAVs, IEEE Sensors Journal, 2015.

[22] Y.Lecun, L.Bottou, Y.Bengio, P.Haffner, Gradient-based Learning applied to Document Recognition, Proceedings of the IEEE, 1998.

[23] M. S. Kristoffersen, J. V. Dueholm, R. Gade, T. B. Moeslund, Pedestrian Counting with Occlusion Handling Using Stereo Thermal Camera, Sensors, 2016.

[24] K.He, X.Zhang, S.Ren, J.Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, IEEE International Conference on Computer Vision (ICCV), 2015.

[25] A. Vedaldi and K.Lenc, MatConvNet – Convolutional Neural Networks for MATLAB, Proceedings of the 25th annual ACM international conference on Multimedia (ICMM), 2015.

[26] S. Zhang, C. Bauckhage, A. B. Cremers, Informed Haar-like Features Improve Pedestrian Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[27] P. Dollar, C. Wojek, B. Schiele, P. Perona, Pedestrian Detection: An Evaluation of the State of the Art, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2012.

[28] A. Vedaldi, B. Fulkerson, VLFeat: An Open and Portable Library of Computer Vision Algorithms, http://www.vlfeat.org/, 2008.

[29] A. Geiger, J. Ziegler, C. Stiller, StereoScan: Dense 3D Reconstruction in Real-time, Intelligent Vehicles Symposium (IV), 2011.