Pyramidal Edge-maps based Guided Thermal Super-resolution

Honey Gupta and Kaushik Mitra, Member, IEEE

Abstract—Thermal imaging is a robust sensing technique but its consumer applicability is limited by the high cost of thermal sensors. Nevertheless, low-resolution thermal cameras are relatively affordable and are also usually accompanied by a high-resolution visible-range camera. This visible-range image can be used as a guide to reconstruct a high-resolution thermal image using guided super-resolution (GSR) techniques. However, the difference in wavelength-range of the input images makes this task challenging. Improper processing can introduce artifacts such as blur and ghosting, mainly due to texture and content mismatch. To this end, we propose a novel algorithm for guided super-resolution that explicitly tackles the issue of texture-mismatch caused due to multimodality. We propose a two-stage network that combines information from a low-resolution thermal and a high-resolution visible image with the help of multi-level edge-extraction and integration. The first stage of our network extracts edge-maps from the visual image at different pyramidal levels and the second stage integrates these edge-maps into our proposed super-resolution network at appropriate layers. Extraction and integration of edges belonging to different scales simplifies the task of GSR as it provides texture to object-level information in a progressive manner. Using multi-level edges also allows us to adjust the contribution of the visual image directly at the time of testing and thus provides controllability at test-time. We perform multiple experiments and show that our method performs better than existing state-of-the-art guided super-resolution methods both quantitatively and qualitatively.

Index Terms—thermal imaging, guided super resolution, pyramidal edge-maps, deep convolutional neural networks

I. INTRODUCTION

Far-infrared or thermographic imaging has many advantages over traditional visual-range imaging as it is less affected by illumination changes. It works well in extreme imaging conditions and has found applications in varied fields such as firefighting [1], gas leakage detection [2], automation [3]–[5] and many more but the high cost of thermal sensors restricts the extensive use of these cameras. Super-resolution (SR) techniques can further increase their applicability by simulating accurate high-resolution thermal images from low-resolution images captured from inexpensive thermal cameras. Most of the lower-end thermal cameras are accompanied by high-resolution visual-range cameras, and consequently, guided super-resolution seems to be the optimal solution to super-resolve the thermal images. However, due to the difference in the wavelength ranges of these sensors, the images can have different contents and texture, which can result in erroneous reconstructions.

The existing thermal super-resolution methods either perform single image SR [6]–[10] or visual-guided SR [11]–[17]. Among single thermal image SR, Choi et al. [9] suggested a shallow three-layer convolutional neural network (CNN). Zhang et al. [10] thereafter combined compressive sensing and deep learning techniques to perform infrared super-resolution. Among the guided thermal SR methods, Lee et al. [11] used brightness information for the visible-range images. Han et al. [12] proposed a guided SR method using CNN that extracts features from infrared and visible images and combines them using convolutional layers. Interestingly, most of the existing methods for depth [18]–[28] or hyperspectral [29]–[32] guided super-resolution have a similar backbone as the thermal guided-SR methods. They both use some variation of the Siamese network [33] to simultaneously extract information from both the low-resolution and guide images and merge them to get the super-resolved image.

A crucial part of super-resolution is to predict the high-frequency details correctly and the Siamese architecture along with end-to-end training can help extract optimal high-frequency information. However, extracting optimal edges from the visual image belonging to the dissimilar domain is challenging, especially when the input image is of a very low-resolution, like the image captured from an inexpensive thermal camera. The existing guided SR methods rely on the network to extract optimal information from the visual image and this can be difficult as it is scene dependent. Fig.1 shows one such case where optimal overlapping edges involve keeping some texture information, like the structure of the car, and simultaneously discarding some, like the texts on the truck. Moreover, texture or content mismatch can introduce artifacts.
or blurred-edges in the super-resolved images, Ni et al. \cite{34} proposed a method to utilise the edge-map and perform GSR but they have a similar vulnerability as their network estimates the optimized edges, which could be a source for the artifacts.

In this paper, we provide a systematic and holistic solution to deal with possible edge-mismatch caused due to multimodality among the input images. Our method is motivated from the fact that layers at different depths of the network have different levels of features: going from fine texture information to object level information. Extracting edges at hierarchical perceptual scales or levels and integrating them in our tailored thermal super-resolution network can allow easier removal of the mismatched edges by the network and help reconstruct better high-frequency details. It also provides a way to adjust the guidance information provided to the network by choosing the levels of edges that are to feed to the SR network. In summary, the main contributions of this paper are:

- We propose a novel method for performing guided super-resolution by extracting pyramidal edge-maps from the visible image and integrating them into our thermal SR network in a hierarchical manner.
- We propose a way to adjust the contribution of the visual information according to factors like guide-noise, by adjusting the levels of edge-maps that are fed to the GSR network.
- We compare our model with existing state-of-the-art GSR methods and show that our method reconstructs the high-frequency details well and hence produces better result, both quantitatively and perceptually.

II. PYRAMIDAL EDGE-MAPS BASED GUIDED SUPER-RESOLUTION (PEG-SR)

The visible-range and thermal images have different contents due to differences in their wavelength ranges, as shown in Fig[1]. Instead of following the conventional approach of using a Siamese network \cite{12}, \cite{35}, we design a new network which is better suited for the task of guided super-resolution (GSR). We know that the guide image has useful high-frequency details but these have to extracted adaptively according to the input low-resolution thermal image. Non-optimality of the extracted high-frequency details from the visual image can lead to texture mismatch, which can further cause artifacts.

At the first glance, it seems that extracting the object-edges could be an ideal solution, but as one can observe from Fig[1], there are additional high-frequency details present in edge-maps at lower levels, i.e. levels 1 and 3 that are equally useful.

To resolve this conundrum, we propose a systematic approach of using edge-maps extracted at pyramidal levels from the visual image and integrating them at appropriate layers of our SR network. In this way, the network can leverage high-frequency information in a hierarchical fashion and propagate the features as per the input low-resolution thermal image.

Figure[2] shows the architecture of our proposed method. We denote the low-resolution thermal, the high-resolution visible and ground-truth thermal images as $x_I$, $g_h$ and $x_h$, respectively. Our network consists of two stages: one for thermal image super-resolution, denoted as $\Psi_x$ and one for multi-level edge extraction from the visual image, denoted as $\Psi_g$.

Many existing edge-detection methods for a single image \cite{36}, \cite{37} extract multi-level edges and merge them to obtain the desired object-edges. These methods inspired us to utilise the progressive nature of features in a convolutional neural network i.e. the relation that shallow layers have texture-like features and deeper layers have object-level features. They extract edges at different perceptual scales by taking output at different layers of the VGG \cite{38} network, as also shown in Fig[2]. Consequently, to obtain edges having visible-range information at different scales, we used the existing edge-extraction method \cite{37} as $\Psi_g$. $\Psi_g$ provides edge-maps at $n$ pyramidal levels. We denote these edge-maps as $[\nabla_{g_1}, \nabla_{g_2}, \ldots, \nabla_{g_n}]$, where $n = 5$ in case of \cite{37}.

Our super-resolution stage, $\Psi_x$, consists of two sub-networks: $\Psi_{up}$ and $\Psi_{fusion}$. The first part of the network, denoted as $\Psi_{up}$, contains convolutional and upsampling layers. For a $2^k$ super-resolution, $\Psi_{up}$ contains $k$ depth-to-space layers. The output $X_1 = \Psi_{up}(x_I)$, is fed to the fusion sub-network $\Psi_{fusion}$. For $n$ visual edge-maps, $\Psi_{fusion}$ contains $n + 1$ dense-blocks \cite{39}, denoted as $[D_1, D_2, \ldots, D_{n+1}]$. The pyramidal edges are added one-at-a-time after each dense-block in the increasing order of the receptive-field, e.g., edge-map with fine-textures is added after the first dense-block and similarly for the rest of edge-maps:
After multiple experiments, we found the optimal values for $\gamma_1$, $\gamma_{\text{perc}}$ and $\gamma_{\nabla}$ to be 10, $1 \times 10^{-4}$ and 1, respectively.

### III. Experiments and results

To compare our method with the existing GSR methods, we performed qualitative and quantitative comparisons on two datasets: FLIR-ADAS [45] and CATS [46]. FLIR-ADAS contains unrectified thermal and visible image pairs having a resolution of $512 \times 640$ and $1600 \times 1800$, respectively. Since the dataset does not contain any calibration images, we rectified one image pair manually and used the estimated relative transformation to rectify the rest of the images in the dataset. The CATS dataset, on the other hand, contains rectified thermal and visible images, both of dimensions $480 \times 640$.

We used the blur-downscale degradation model [47] to create the low-resolution images. For the training set, we down-sampled images with blur kernels $\in [0, 4]$ at a step of 0.5. For the FLIR-ADAS dataset, our training set contains 43830 downsampled image pairs and the test-set contains 1257 pairs. Whereas, in the CATS dataset, our training set contains 944 image-pairs and the test-set contains 50 pairs. Since CATS training-set is quite small, we used to it fine-tune the models pre-trained on FLIR-ADAS training-set and then tested them on CATS test-set. We perform experiments for $\times 4$ and $\times 8$ upsampling scales. For both cases, the input thermal-resolution is close to the resolution of many low-cost thermal cameras, like FLIR AX8 and FLIR One. For FLIR-ADAS, the low-resolution dimensions are $64 \times 80$ and for CATS, it is $60 \times 80$.

#### A. Quantitative comparison.

We compare our method with 9 existing guided super-resolution methods: TGV2-L2 [23], FBS [40], Joint-BU [22],...
TABLE I

| Method            | FLIR-ADAS dataset | CATS dataset |
|-------------------|-------------------|--------------|
|                   | ×4                | ×8           | ×4                | ×8           |
|                   | PSNR  | SSIM  | LPIPS | PSNR  | SSIM  | LPIPS | PSNR  | SSIM  | LPIPS |
| TGV2-L [23]       | 28.77 | 0.892 | 0.422 | 26.42 | 0.821 | 0.399 | 32.17 | 0.938 | 0.225 |
| FBS [40]          | 25.48 | 0.787 | 0.387 | 25.03 | 0.770 | 0.476 | 29.12 | 0.825 | 0.450 |
| Joint-BU [22]     | 27.77 | 0.874 | 0.286 | 25.61 | 0.803 | 0.406 | 31.23 | 0.953 | 0.233 |
| Infrared SR [12]  | 28.21 | 0.889 | 0.405 | 26.03 | 0.817 | 0.521 | 28.27 | 0.901 | 0.348 |
| SDF [41]          | 28.70 | 0.875 | 0.321 | 26.72 | 0.819 | 0.363 | 32.56 | 0.941 | 0.246 |
| MSF-SR [35]       | 29.21 | 0.901 | 0.200 | 27.92 | 0.835 | 0.249 | 29.37 | 0.830 | 0.415 |
| MSG-Net [42]      | 29.46 | 0.897 | 0.184 | 27.29 | 0.827 | 0.296 | 31.56 | 0.964 | 0.177 |
| PixTransform [43] | 24.84 | 0.787 | 0.329 | 23.31 | 0.836 | 0.371 | 28.48 | 0.792 | 0.442 |
| Deep-ISTA [44]    | 25.86 | 0.828 | 0.529 | 25.56 | 0.778 | 0.598 | 33.72 | 0.956 | 0.178 |
| PEG-SR(Ours)      | 29.63 | 0.910 | 0.151 | 27.95 | 0.837 | 0.213 | 35.50 | 0.971 | 0.139 |

Infrared SR [12], SDF [41], MSF-SR [35], MSG-Net [42], Pix-Transform [43] and Deep-ISTA [44]. We use three metrics to quantitatively assess the methods’ outputs: PSNR, SSIM and Perceptual distance(LPIPS) [48]. Among these, PSNR and SSIM are distortion-based metrics and can be biased towards blurred images. Hence, we also used LPIPS that computes the perceptual distance between the reconstructed and the ground-truth images. Higher PSNR, SSIM and lower LPIPS are better.

Table I shows the results for ×4 and ×8 upsampling scales. Most of the existing methods perform quite well in terms of distortion metrics but poorly in terms of the perceptual metric. However, our method outperforms the existing methods in terms of both metrics. Among the existing methods, MSF-STI, MSG-Net and Joint-BU results are considerably good, yet our method reconstructs high-frequency details more faithfully and has much sharper edges. It is evident perceptually as well as from the metrics that our results are closest to the ground-truth.

B. Qualitative comparison.

We show the qualitative comparison of our method for ×4 and ×8 SR on FLIR-ADAS dataset in Fig 3. Most of the existing methods have blurred edges, especially in the case of ×8 SR, which could be caused either due to very low-resolution of the input thermal images(60 × 80) or due to multimodality. Among the existing methods, MSF-STI, MSG-Net and Joint-BU results are considerably good, yet our method reconstructs high-frequency details more faithfully and has much sharper edges. It is evident perceptually as well as from the metrics that our results are closest to the ground-truth.

C. Variation with respect to noise and hierarchy of edge-maps

We performed experiments to analyze the contribution of different levels of edge-maps. Moreover, we also analyze the variation in performance with respect to the hierarchy of edge-maps when the input is a noisy visible image. To do this, we considered 3 variations of PEG-SR that contain the same backbone SR network but separate sets of edge-maps as input. These models were trained on clean visual images but we tested them on edges extracted from noisy visual images containing added Gaussian noise with mean 0 and variance σ = 0.01, 0.03. Table II summarises the results from the performed study. PEGSR(0) is our model Ψ E without any edge supervision; PEGSR(5) is Ψ E network having only Level-5 edge-maps. Similarly, PEGSR(3−5) contains edges from Levels 3, 4 and 5; and PEGSR(1−5) is our proposed PEGSR network containing edges from Levels 1 to 5.

We observe that when the visual image contains no noise, the performance follows the order PEGSR(1−5) > PEGSR(3−5) > PEGSR(5) > PEGSR(0), which shows the effectiveness of using edge-maps at different levels. In case of noisy visual images, we observe that PEGSR(1−5) performs lower, mostly because the noise majorly affects lower level edge-maps i.e. 1 and 2. PEGSR(3−5) performs the best, followed by PEGSR(5) and PEGSR(0), which is consistent with the amount of guide information provided to the models. This shows that our model can be used to adjust the contribution of guide image and tune the network according to factors like noise. However, training with noisy visual images makes our proposed PEGSR network learn to adapt to the noise and hence, it performs much better, as shown in Table II(c).

IV. CONCLUSION

We proposed a hierarchical edge-based guided super-resolution algorithm (PEG-SR) that tackles multimodality and edge-mismatch between the input low-resolution thermal and high-resolution visible-range image in a systematic and holistic manner. Our method robustly combines multi-level edge information extracted from the visual image into our tailored thermal super-resolution network and consequently produces better high-frequency details. We showed that our results are better both perceptually and quantitatively than the existing state-of-the-art guided super-resolution methods.

TABLE II

| Model(Edge-levels) | No noise | σ = 0.01 | σ = 0.03 |
|--------------------|----------|-----------|-----------|
| (a) PEGSR(0) (Single image) | 29.06 | 0.907 | 0.1755 |
| (b) PEGSR(5) | 29.19 | 0.907 | 0.1733 |
| (c) PEGSR(3−5) | 29.28 | 0.904 | 0.1793 |
| (d) PEGSR(1−5) (Proposed) | 29.64 | 0.910 | 0.1508 |
| (e) PEGSR(1−5) (Retrained with noise) | 29.33 | 0.912 | 0.1589 |
REFERENCES

[1] B. C. Arrue, A. Oller, and J. M. De Dios, “An intelligent system for false alarm reduction in infrared forest-fire detection,” *IEEE Intelligent Systems and Their Applications*, vol. 15, no. 3, pp. 64–73, 2000.

[2] A. Prata and C. Bernardo, “Retrieval of volcanic ash particle size, mass and optical depth from a ground-based thermal infrared camera,” *Journal of Volcanology and Geothermal Research*, vol. 186, no. 1-2, pp. 91–107, 2009.

[3] S. Khattak, C. Papachristos, and K. Alexis, “Marker based thermal-inertial localization for aerial robots in obscure filled environments,” in *International Symposium on Visual Computing*. Springer, 2018, pp. 565–575.

[4] S. Khattak, C. Papachristos, and K. Alexis, “Keyframe-based direct thermal-inertial odometry,” in *arXiv preprint arXiv:1903.00798*, 2019.

[5] P. V. K. Borges and S. Vidas, “Practical infrared visual odometry,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2205–2213, 2016.

[6] A. Panagiotopoulos and V. Anastassopoulos, “Super-resolution reconstruction of thermal infrared images,” in *Proceedings of the 4th WSEAS International Conference on REMOTE SENSING*, 2008.

[7] H. Jones and X. Smaila, “Scaling of thermal images at different spatial resolution: the mixed pixel problem,” *Agronomy*, vol. 4, no. 3, pp. 380–396, 2014.

[8] F. Almasri and O. Debeir, “Multimodal sensor fusion in single thermal image super-resolution,” in *Asian Conference on Computer Vision*. Springer, 2018, pp. 418–433.

[9] Y. Choi, N. Kim, S. Hwang, and I. S. Kweon, “Thermal image super-resolution,” in *Infrared Physics & Technology*, vol. 9, no. 4, p. 316, 2017.

[10] K. Hayat, “Multimedia super-resolution via deep learning: A survey,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 307–316.

[11] M. Shoeibi, A. Robles-Kelly, R. Timofte, R. Zhou, F. Lahoud, S. Suistrunk, Z. Xiong, Z. Shi, C. Chen, D. Liu et al., “Pirn2018 challenge on spectral image super-resolution: methods and results,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[12] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, “Fully-convolutional siamese networks for object tracking,” in *European conference on computer vision*. Springer, 2016, pp. 850–865.

[13] M. Ni, J. Lei, R. Cong, K. Zheng, B. Peng, and X. Fan, “Color-guided depth map super-resolution using convolutional neural network,” *IEEE Access*, vol. 5, pp. 26 666–26 672, 2017.

[14] F. Almasri and O. Debeir, “Multimodal sensor fusion in single thermal image super-resolution,” in *arXiv preprint arXiv:1812.09276*, 2018.

[15] X. Zhang, C. Li, Q. Meng, S. Liu, Y. Zhang, and J. Wang, “Infrared image super-resolution by combining compressive sensing and deep learning,” in *Sensors*, vol. 18, no. 8, p. 2587, 2018.

[16] K. Lee, J. Lee, J. Lee, S. Hwang, and S. Lee, “Brightness-based convolutional neural network for thermal image enhancement,” *IEEE Access*, vol. 5, pp. 26 867–26 879, 2017.

[17] T. Y. Han, Y. J. Kim, and B. C. Song, “Convolutional neural network-based infrared image super-resolution under low light environment,” in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 803–807.

[18] F. Liu, P. Han, Y. Wang, X. Li, L. Bai, and X. Shao, “Super-resolution reconstruction of infrared images based on classified dictionary learning,” in *Infrared Physics & Technology*, vol. 90, pp. 146–155, 2018.

[19] C. Sun, J. Lv, J. Li, and R. Qiu, “A rapid and accurate infrared image super-resolution method based on zoom mechanism,” *Infrared Physics & Technology*, vol. 88, pp. 228–238, 2018.

[20] N. Yokoya, “Texture-guided multisensor superresolution for remotely sensed images,” *Remote Sensing*, vol. 9, no. 4, p. 316, 2017.

[21] P. Song, X. Deng, J. F. Mota, N. Deligiannis, P. L. Dragotti, and M. R. Rodrigues, “Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries,” in *arXiv preprint arXiv:1709.08680*, 2017.

[22] X. Chen, G. Zhai, J. Wang, C. Hu, and Y. Chen, “Color guided thermal image super-resolution,” in *2016 Visual Communications and Image Processing (VCIP)*. IEEE, 2016, pp. 1–4.

[23] G. Riegler, D. Ferstl, M. Rüther, and H. Bischof, “A deep primal-dual network for guided depth super-resolution,” *arXiv preprint arXiv:1607.08569*, 2016.

[24] W. Zhou, X. Li, and D. Reynolds, “Guided deep network for depth map super-resolution: How much can color help?” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 1457–1461.

[25] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, “Hierarchical features driven residual learning for depth map super-resolution,” *IEEE Transactions on Image Processing*, 2018.

[26] J. Xie, R. Feis, and M.-T. Sun, “Edge-guided single depth image super-resolution,” vol. 25, no. 1, pp. 428–438, Jan 2016.

[27] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, “Joint bilateral upsampling,” in *ACM Transactions on Graphics (ToG)*, vol. 26, no. 3. ACM, 2007, p. 96.

[28] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof, “Image guided depth upsampling using anisotropic total generalized variation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 993–1000.

[29] Z. Shi, C. Chen, Z. Xiong, D. Liu, Z.-J. Zha, and F. Wu, “Deep residual attention network for spectral image super-resolution,” in *European Conference on Computer Vision*. Springer, 2018, pp. 35–50.

[30] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[32] J. B. Barron and B. Poole, “The fast bilateral solver,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016, pp. 850–865.

[33] B. Phelan, K. Sherbondy, and C. Kambhamettu, “Cats: A color and texture-guided deep network for guided depth super-resolution,” in *2017 IEEE International Conference on Computer Vision*, 2017, pp. 426–434.

[34] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *CVPR*, 2018.