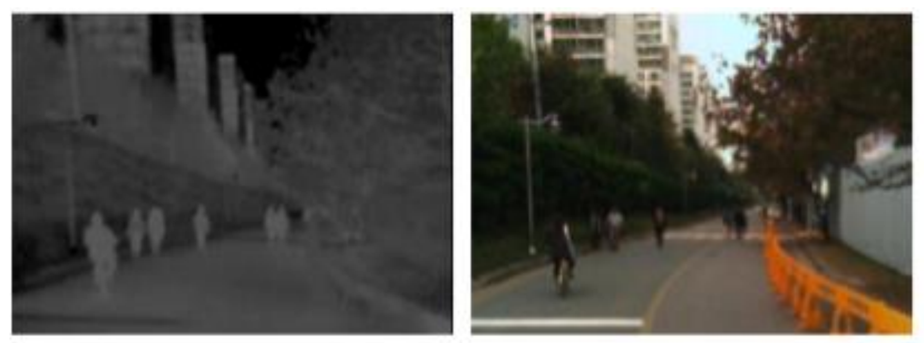
Guided Up-sampling of Low-Resolution LWIR Images with Bilateral Filtering of Aligned RGB Images

Uriel A. Rosa
SCPD Student, Stanford University

Motivation

- Pedestrian detection is crucial for the development of autonomous vehicles.
- Thermal channels are helpful in pedestrian detection under low light environments.
- Human bodies radiate in the thermal band, LWIR: long wavelength infrared (7.5 ~ 13µm).
- The 9.3 µm band is well suitable for human detection.
- However, thermal imaging has resolution limitations due to the dimensions of the camera infrared focal plane and sensor accuracy.
- Use of infrared thermal images acquired from color-thermal datasets require image post-processing enhancement for improvement of feature extractions.



The KAIST Multispectral Pedestrian Dataset consists of 95k color-thermal pairs (640x480, 20Hz) taken from a vehicle. (left image: lwir) http://multispectral.kaist.ac.kr
https://sites.google.com/site/pedestrianbenchmark/data-format

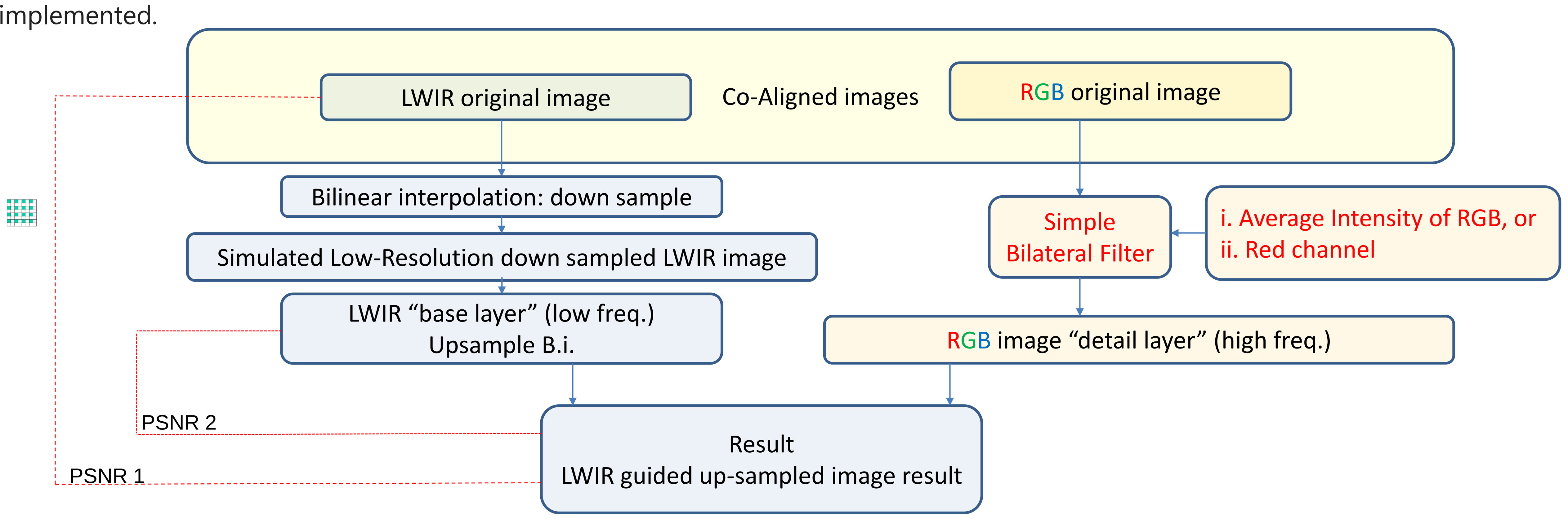
Flir Lepton uncooled low cost / low res thermal camera (160x120)
https://lepton.flir.com/application-notes/people-finding-with-a-lepton/

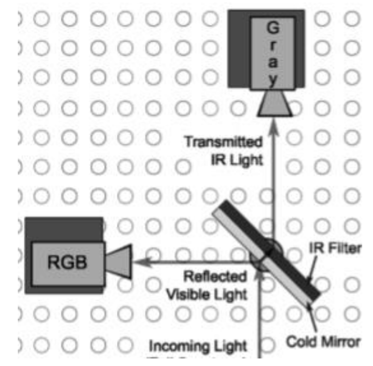
New Technique

A new technique is proposed for guided up-sampling of the low-resolution thermal images that are acquired along with co-aligned high resolution RGB images.

The LWIR image is threatened as the base image, where the low frequency components are maintained.

The use of a simple bilateral filter is proposed for simple feature extraction of the RGB image, instead of using it for typical image denoising. The RGB detailed layer, high frequency, is extracted via bilateral filter and combined with the base layer image, i.e., the LWIR image. The resulted image is shown after tone mapping is implemented.





Related Work

- The goal of this project is to enhance the resolution of low resolution thermal images by extracting detail features of RGB co-aligned images via simple bilateral filtering. The LWIR image forms the base image, and the RGB detail is added.
- Bilateral filtering was developed by Tomasi and Manduchi in 1998. It is a non-linear filter where the output is a weighted average of the input,[1].
- The Bilateral filter, denoted by BF[·], is defined by:

$$BF[I]_{\mathbf{p}} = \frac{1}{W_{\mathbf{p}}} \sum_{\mathbf{q}\in\mathcal{S}} G_{\sigma_s}(\|\mathbf{p}-\mathbf{q}\|)\, G_{\sigma_r}(|I_{\mathbf{p}} - I_{\mathbf{q}}|)\, I_{\mathbf{q}}, \quad \text{where,} \quad W_{\mathbf{p}} = \sum_{\mathbf{q}\in\mathcal{S}} G_{\sigma_s}(\|\mathbf{p}-\mathbf{q}\|)\, G_{\sigma_r}(|I_{\mathbf{p}} - I_{\mathbf{q}}|)$$

- Bennett et al. [2] employed noise reduction in the visible-light video to improve the quality of large-scale feature fusion.
- They acquired detail features from the less-noisy IR video. In addition to noise reduction, the bilateral filter is used because it decomposes images into two components which have meaningful perceptual analogs.
- The bilateral's filtered image has large areas of low frequencies separated by sharp edges, "large-scale features.
- Kopf et al. [3] describe joint bilateral upsampling, a method inspired from the bilateral filter to upsample image data.
- Data at low resolution can be computed and then upsampled using a weighted average. High-resolution data are produced by averaging the samples in a 5x5 window at low resolution. The weights are similar to those defined by the bilateral filter.
- Display of high-dynamic-range images reduces the contrast while preserving detail [4]. It is based on a two-scale decomposition of the image into a base layer, encoding large-scale variations, and a detail layer. Only the base layer has its contrast reduced, thus preserving detail. The base layer is obtained using an edge-preserving filter, the bilateral filter.

References

[1] S. Paris, P. Kornprobst, J. Tumblin and F. Durand. *Bilateral Filtering: Theory and Applications*. Foundations and Trends R in Computer Graphics and Vision Vol. 4, No. 1 (2008) 1–73. 2009.
[2] E. P. Bennett, J. L. Mason, and L. McMillan. *Multispectral bilateral video Fusion*. IEEE Transactions on Image Processing, vol. 16, no. 5, pp. 1185–1194, May 2007.
[3] J. Kopf, M. Uyttendaele, O. Deussen, and M. Cohen. *Capturing and viewing gigapixel images*. ACM Transactions on Graphics, vol. 26, no. 3, p. 93, Proceedings of the ACM SIGGRAPH conference, 2007.
[4] Fredo Durand and Julie Dorsey. Fast Bilateral Filtering for the Display of High-Dynamic-Range Images. 2002.

Experimental Results

KAIST Images LWIR / RGB	I00000		I00495		I01085		I01700		I01840	
B.F. (r=5, σs=3) σi	PSNR1	PSNR2	PSNR1	PSNR2	PSNR1	PSNR2	PSNR1	PSNR2	PSNR1	PSNR2
Intensity ch: Av. RGB										
0.2	9.0285	8.4820	2.4499	2.1544	-		-		-	
0.4	9.0221	8.4763								
0.6	9.0168	8.4717								
1.5	9.0078	8.4635								
Intensity ch: Red										
0.1	9.1186	8.5721								
0.2	9.1154	8.5692	2.3438	2.0480	9.9379	8.8865	4.4171	3.9121	8.3969	6.5522
0.4	9.1097	8.5643								
0.6	9.1049	8.5601								

Method evaluation:
- The guided upsampling of the five co-aligned images shown above indicated the method produced image resolution improvements, but these were modest.
- The analyze of the method by varying parameters and measured PSNR, indicated the range of obtained PSNR varied with the image from 2~10.
- The use of the red channel vs av. RGB channels produced similar PSNRs.
- There are no results available from related work in the literature for direct comparison with this approach.