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

Uriel A. Rosa
SCPD Student
Stanford University
uarosa@stanford.edu

Abstract

An innovative method has been proposed to enhance the typical low-resolution of thermal images paired with visible images obtained for use in pedestrian detection. A low-resolution thermal image is up-sampled by using bilinear interpolation. This image is taken as a base, low frequency, image. The paired RGB image is processed by a bilateral filter. The monochromatic intensity for this image is obtained either from the red channel, or from an average of all color channels. The detail image, high frequency, is extracted and combined with the detail image to obtain the guided up-sampled thermal image. Co-aligned LWIR and RGB images were downloaded from a dataset of pedestrian images available for public use (KAIST). The results obtained from five image pairs showed modest improvements in image quality. The computed PSNR from all images covered a range from approximately 2 dB to 10 dB.

1. Introduction

Pedestrian detection is an active area of research relying on quality sensed imaging, and crucial for the development and implementation of autonomous vehicles. Historically, most available pedestrian data sets have relied on images obtained from color channels [1].

Infrared thermal images are widely used in night vision technologies, and low-light imaging. Thermal channels are helpful in pedestrian detection in low light environments [2].

Infrared cameras are generally available in two types of wavelengths, near infrared (0.75~1.3μm) and long wavelength infrared (7.5 ~ 13μm), or commonly referred as a thermal band.

Pedestrian recognition is more important in the thermal band due to the relative immunity to interference of thermal cameras to traffic lights and headlights. More importantly, human bodies radiate in the thermal band.

The thermal infrared region is referred to as passive (https://en.wikipedia.org/wiki/Infrared): “The "thermal imaging" region, in which sensors can obtain a completely passive image of objects only slightly higher in temperature than room temperature - for example, the human body - based on thermal emissions only and requiring no illumination such as the sun, moon, or infrared illuminator. This region is also called the "thermal infrared”.

Multispectral pedestrian datasets are available for public use. Paired images can be obtained in color channels, and thermal bands with wavelength of 9.3 μm. This band is well suitable for human detection. These type of data are typically used for manual image annotations for automation research.

There various datasets of thermal-color paired images available to the public, created under day and night lighting conditions [3,4,5]. These images can be processed for testing new fusion, or other, algorithms enhancing image quality of pedestrian images. Precisely captured with split prisms and co-registered color-thermal image pairs are thus available on these pedestrian databases.

LWIR and visible images from the KAIST database [3] are typically used as described in their site: “The KAIST Multispectral Pedestrian Dataset consists of 95k color-thermal pairs (640x480, 20Hz) taken from a vehicle. All the pairs are manually annotated (person, people, cyclist) for the total of 103,128 dense annotations and 1,182 unique pedestrians. The annotation includes temporal correspondence between bounding boxes like Caltech Pedestrian Dataset” [3].

However, thermal imaging has resolution limitations due to the dimensions of the camera infrared focal plane and sensor accuracy [2]. Use of infrared thermal images acquired from color-thermal datasets require image post-processing enhancement for feature extractions, and improvements.

The goal of this project is to enhance the resolution of low-resolution thermal images (i.e., enhance the thermal image quality) by extracting detail features of RGB co-aligned images via simple bilateral filtering. The LWIR image forms the base image, and the RGB detailed image is added to the up-sampled thermal image.

2. Related Work

Variances of bilateral filters have been used for image fusion of flashed and color images. Lighting is part of the

visual richness of the scenes. Subtle details are often found in low light conditions. Flash photography was developed to circumvent issues such as image blur, noise, reduced depth of field, while adjusting camera exposure times, aperture and ISO in trying to capture natural ambient illumination in low-light environments [6].

Introduction of artificial light to the scene, with addition of flash, allows for adoption of short exposure times, small apertures, and low sensor gain while capturing enough light to develop sharp images with reduced noise. Bright images have high signal to-noise ratios. Details hidden in the noise can be resolved as opposed to an image acquired under low ambient illumination.

Bilateral filters combine a classic low-pass filter with an edge stopping function. The function attenuates the filter kernel weights when the intensity difference between pixels is large.

A pair of images of low-light, an ambient image and an image with flash are used to capture image details. The flash images contain better estimates of the true high-frequency information than the ambient image. A modification of the basic bilateral filter to compute the edge-stopping functions has been created by Bae et al. [7].

Relevant implementations of related work are presented as follows. The following are partial transcriptions, through the end of this section, of mentioned literature.

"Petschnigg et al. [8] have demonstrated applications that combine the strengths of flash and no-flash photographs to synthesize new images that are of better quality than either of the originals.

The advantage of using the bilateral filter rather than a classic low-pass Gaussian filter is that they reduce haloing present in the thermal images. Petschnigg's approach allows the user to control how much detail is transferred over the entire image. Automatically adjusting the amount of local detail transferred is an area for future work. Petschnigg et al. have explored several ways of interpolating the ambient and flash images. The most effective scheme is to convert the original flash/no-flash pair into YCbCr space and then linearly interpolate them.

Petschnigg et al. suggests an exciting possibility is to use an infrared flash. While infrared illumination yields incomplete color information, it does provide high-frequency detail, and does so in a less intrusive way than a visible flash.

In multispectral image fusion, Bennett et al. [8,9] shows how to exploit infrared data in addition to standard RGB data to denoise low-light video streams. They use the dual bilateral filter, a variant of the bilateral filter with a modified range weight that accounts for both the visible spectrum (RGB) and the infrared spectrum.

Bennett et al. shows that this combination better detects edges because it is sufficient for an edge to appear in just one of the channels (RGB or infrared) to form a sharp boundary in the result. In combination with temporal filtering, they demonstrate that it is possible to obtain high-quality video streams from noisy sequences of moving objects shot in very low light.

Bilateral Filter is computationally an expensive filter. Kopf et al. [10] describes joint bilateral up-sampling, a method inspired from the bilateral filter to up-sample image data. The advantage of their approach is that it is generic and can potentially up-sample any kind of data. Given a high-resolution image and a down sampled version, one can compute the data at low resolution and then up-sample them using a weighted average. High-resolution data are produced by averaging the samples in a 5×5 window at low resolution. The weights are similar to those defined by the bilateral filter, as each neighboring pixels' influence decreases with distance and color difference. As a result, Kopf's scheme interpolates low-resolution data while respecting the discontinuities of the high-resolution input image. This filter is fast to evaluate because it only considers a small spatial footprint.

The success of the bilateral filter lies in its combination of simplicity, good results, and efficient algorithms. Although alternatives exist for each of these points, few, if any, combine all these advantages. The filter is very flexible because the range weight can be adapted to accommodate any notion of pixel value difference, including arbitrary color spaces, data from other images, or any information about the relevance of one pixel to another pixel. The original goal of the filter was denoising, in which case a small spatial kernel suffices and the residual of the filter is discarded as the noise component. In contrast, many new applications leverage the bilateral filter to create two-scale decompositions that rely on large spatial kernels and where the residual of the filter is preserved because it is much more relevant to the human visual system. The use of large spatial support has motivated a variety of accelerations schemes and the bilateral filter can now be applied in real time to large inputs.

Bennett et al. [9] 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.

Display of high-dynamic-range images reduces the contrast while preserving detail [12]. 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 [8,12]."

2.1. Bilateral Filter: Layers

Bilateral filtering was developed by Tomasi and Manduchi, in 1998 [11]. It is a non-linear filter where the output is a weighted average of the input.
The Bilateral filter, defined by an image I at pixel p, with $G\sigma s = \exp(-x^2/\sigma^2)$, a Gaussian function, denoted by BF [·], is defined by:

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

where,

$$W_{\mathbf{p}} = \sum_{\mathbf{q} \in \mathcal{S}} G_{\sigma_s}(\|\mathbf{p} - \mathbf{q}\|)\, G_{\sigma_r}(|I_{\mathbf{p}} - I_{\mathbf{q}}|) \quad (2)$$

where σ_s controls the spatial neighborhood, and σ_r the influence of the intensity difference, and Wp normalizes the weigh.

The Large-Scale Tonal Distribution method developed by Bae et al. [7], used a similar bilateral decomposition approach developed by Durand and Dorsey [12]. Since contrast is a multiplicative effect, the work was done in the logarithmic domain. They defined the base layer B and detail layer D from the input image I (where I, B and D have log values):

$$B = BF[I] \quad \text{and} \quad D = I - B \quad (3)$$

The choice of σs and σr is crucial.

Here, similar approach is followed in this work.

3. Methods

In this work, 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 associated to the thermal, up-sampled thermal image.

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 obtained after tone mapping is implemented.

3.1. Algorithm

The algorithm is described in figure 1. The two image branches in the flow chart, LWIR and RGB, indicate the LWIR image is first downsampled for simulation purposes and bilinearly up-sampled to obtain the base layer. The RGB intensity, red channel or average color channel is bilateraly filtered to obtain the detail layer. Both, base and detail layers are combined to obtain the final image.

3.2. Image dataset

Five co-aligned image pairs were selected for the evaluation of the proposed method. The evaluation criteria were the image quality and the calculated PSNR. The PSNR1 and PSNR2 are indicated in the flow chart, figure 1. The filter parameters r = 5, σs = 3 were maintained constant. The parameter $\sigma r = \sigma i$ (in table 1) was varied from 0.1 to 1.5 as shown in table 1.

4. Results

4.1. Qualitative

Visual analysis is performed by comparing the five processed scenes, figures 2 to 6.

Figure 2. Processed images I00000 (LWIR and visible).

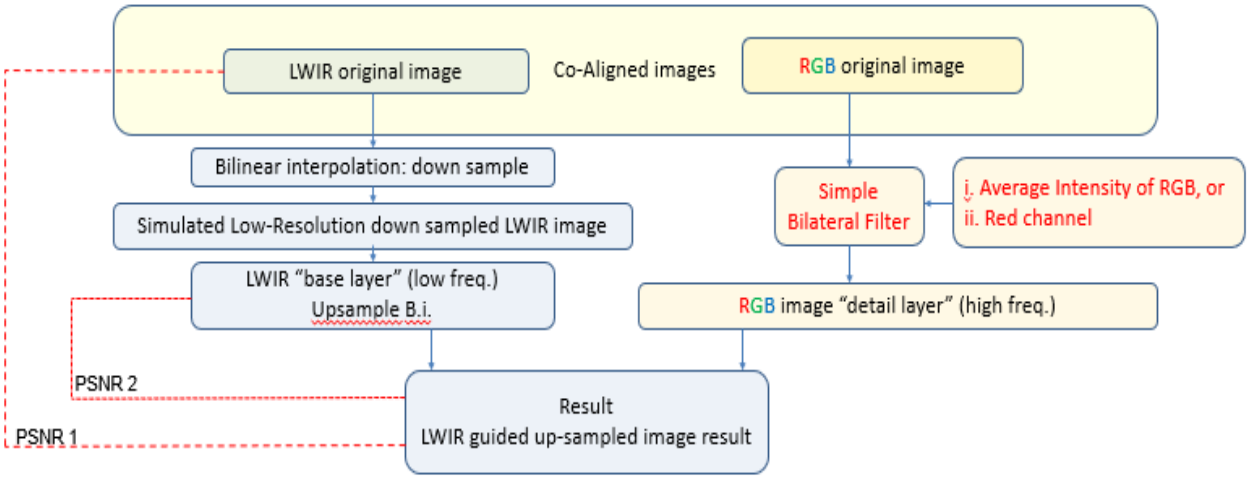

Figure 1. Block diagram for the LWIR guided up-sampled image method.

Figure 3. Processed images I00495 (LWIR and visible).

Figure 4. Processed images I01085 (LWIR and visible).

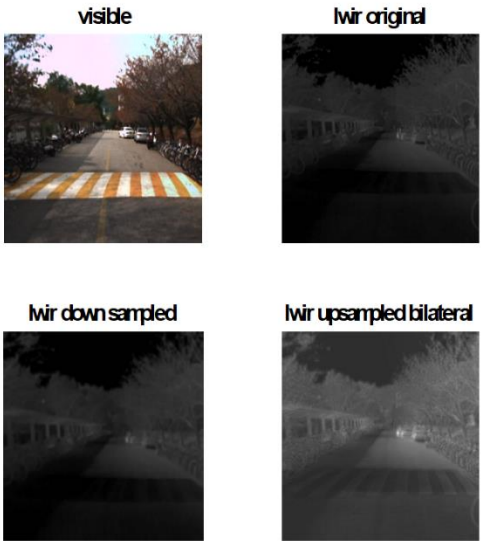

Figure 5. Processed images I01840 (LWIR and visible).

Figure 6. Processed images I01700 (LWIR and visible).

Figure 7 shows image I01700 and its base and detail layers. This image produced the most improvements in quality compared to all images.

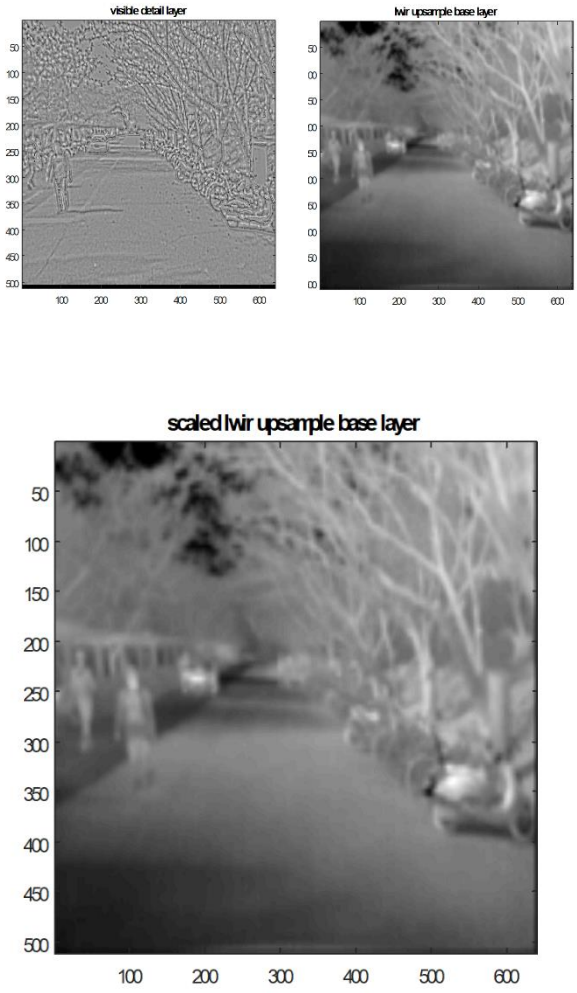

Figure 7. The detail and base layers are shown for the I01700. This image produced the most quality improvements among all processed images.

The processing time for each co-aligned image pair was about 2.5 minutes, using the lab computer.

4.2. Quantitative

The PSNR calculated for all images are shown in table 1.

KAIST Images LWIR / RGB	I00000		I00495		I01085		I01700		I01840	
B.F. (r=5, σ_s=3) σ_i	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								

Table 1. PSNR obtained for all images used in this study.

The image I01700 showed most improvements, with a PSNR around 10dB. The range of PSNR was from 2dB to 10 dB, approximately.

5. Discussion

The guided up-sampling evaluation for the five co-aligned images performed in this study has indicated the method produced image resolution improvements, but these improvements 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.

5.1. Future Work

A parametric analysis can be performed by exploring extended parameter ranges.

It is suggested the implementation of fast bilateral filter to reduce processing time. In this case each mage pair required about 2.5 minutes of run time using the lab computer.

The use of other non-visible images could also be investigated for use as the detail layer.

6. Acknowledgements

I would like to thank Prof. Gordon Wetzstein for his suggestions and guidance on this project.

References

[1] St-Laurent L., Maldague X., Prévost D. Combination of colour and thermal sensors for enhanced object detection; Proceedings of the 2007 10th International Conference on Information Fusion; Quebec City, QC, Canada. 9–12 July 2007; pp. 1–8.

[2] A Single-Image Super-Resolution Algorithm for Infrared Thermal Images. Kiran Y, Shrinidhi V, W Jino Hans and N Venkateswaran. IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.10, October 2017.

[3] KAIST. Dataset Info. - KAIST Multispectral Pedestrian Benchmark. https://sites.google.com/site/pedestrianbenchmark/data-format

[4] CVPR2015.https://soonminhwang.github.io/rgbt-ped-detection/ 2015.

[5] CVC-14: Visible-FIR Day-Night Pedestrian Sequence Dataset.http://adas.cvc.uab.es/elektra/enigma-portfolio/cvc-14-visible-fir-day-night-pedestrian-sequence-dataset/

[6] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. F. Cohen, and K. Toyama, “Digital photography with flash and no-flash pairs,” ACM Trans. Graph., vol. 23, no. 3, pp. 661–669, 2004.

[7] Soonmin Bae, Sylvain Paris, Frédo Durand: Two-scale tone management for photographic look. ACM Trans. Graph. 25(3): 637-645 (2006).

[8] 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 DOI: 10.1561/0600000020..

[9] 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.

[10] 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.

[11] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in Proc. Int. Conf. Computer Vision, 1998, pp. 836–846.

[12] Fredo Durand and Julie Dorsey. Fast Bilateral Filtering for the Display of High-Dynamic-Range Images. 2002. MIT LCS · SIGGRAPH 2002.