Fusion of Visual and Near Infrared Images in Embedded Platforms
ECE496 Senior Thesis
Progress Report 2

Gustavo Fonseca, gvf3@illinois.edu

1 Overview

The main strides in my senior thesis that I am documenting here are my initial attempts to recreate the image fusion algorithm discussed in [1]. I have created a Python code that outlines the steps I replicated and some of my initial attempts to improve the method.

2 Zhang's Algorithm Overview

To put my work in context, it is important to first provide an overview of Zhang's algorithm [1]. We can see the overall workflow for Zhang's algorithm in Fig. 1. Zhang's algorithm are performed only in the brightness layer of the visual (VIS) image and its corresponding near-infrared (NIR) image. The hue and saturation layers of the VIS image are left unchanged.

In Zhang's workflow, we can see three main aspects: The Weighted Region Mask (WRM), Transfer Contrast (TC), and Transfer Texture (TT). Each of these aspects, my reproduction attempts, and some of my initial approaches to improve are reported in the upcoming sections.

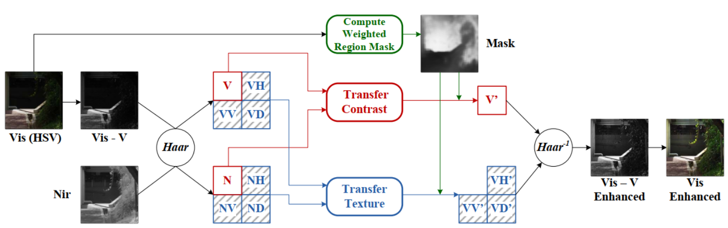

Figure 1: The 3 main steps of Zhang's [1] workflow for enhancing images: Compute Weighted Region Mask, Transfer Contrast, and Transfer Texture.

3 Weighted Region Mask

The goal of the Weigthed Region Mask (WRM) is to indicate which sections of the image need the most enhancement.

3.1 Zhang's Description

In Zhang's algorithm, the WRM is computed solely from the visual spectrum (VIS) image.

In [1], the authors reasoned that the parts of the image that lost the most detail would be the areas with the highest and lowest brightness and low saturation. To model which parts of an image should be enhanced the most, Zhang used in (1) and (2), where s and v are the saturation and brightness of each pixel respectively, and p_s and p_v are the probabilities that this specific value of s or v appear in the image.

Zhang got p_s and p_v by using a normalized histogram of the image's S (saturation) and V (brightness) components. With this, Zhang obtained the final WRM (W) using (3).

$$\begin{aligned} W_s &= 1 - e^{-p_s|s-1|}, \quad p_s \in [0,1],\, s \in [0,1], && (1) \\ W_v &= 1 - e^{-p_v|v-0.5|}, \quad p_v \in [0,1],\, v \in [0,1], && (2) \\ W &= W_s \cdot W_v. && (3) \end{aligned}$$

3.2 My Results

As seen in Fig. 2, I was able to accurately reproduce the method to create the WRM. The WRM in Fig. 2 is encouraging as we can see that the very saturated parts of the VIS image are successfully identified as regions of improvement in the WRM.

I should mention that I had to normalize the p_v and p_s between 0 and 1 because just computing the WRM from (3) was causing the values in the WRM matrix to have a very small range, with a maximum magnitude of 10^{-5}. After normalizing, these values improved significantly, giving me a maximum of the WRM matrix to be around 0.3.

Figure 2: My resulting Weighted Region Mask for test image from [2]

4 Transfer Contrast

The goal in this step is to transfer the contrast of the NIR Image to the enhanced VIS image.

4.1 Zhang's Description

Zhang proposes that, before trying to transfer the contrast, one applies a **one-level Haar wavelet decomposition** to the VIS image's brightness component (V), and to the NIR image (N) as shown in Fig. 1. Let V_{LL} and N_{LL} be the LL subband of each of the one-level Haar wavelet decomposition respectively.

TC then uses V_{LL} and N_{LL} as follows. The algorithm starts by applying a bilateral filter (bf) separately to V_{LL} and N_{LL}. The specific parameters of the filter were defined in [3]. This results in the large-scale layer for both images, $V_{LL,l}$ and $N_{LL,l}$, where the image is blurred but the edges are maintained more than a Gaussian blur.

The detail layers $V_{LL,d}$ and $N_{LL,d}$ are then obtained as shown in (4) and (5). I would like to mention that I tried to make my notation more precise than Zhang's to make the description clearer.

$$
\begin{aligned}
V_{LL,l} &= bf(V_{LL}), \quad V_{LL,d} = V_{LL}/V_{LL,l}, && (4) \\
N_{LL,l} &= bf(N_{LL}), \quad N_{LL,d} = N_{LL}/N_{LL,l}. && (5)
\end{aligned}
$$

- *Note for possible discussion:* I should mention that both [1], [3], and several other sources were very unclear about the exact computations of $V_{LL,d}$ and $N_{LL,d}$. For instance, [1] wrote $V_{LL,d} = V_{LL} - V_{LL,l}$ while writing "$V_{LL,d}$ and $N_{LL,d}$ are corresponding detail layer (after taking the logarithm)", which makes me wonder why Zhang did not write as in (4) and (5). If we later decide to pursue this TC method, this would be an aspect that needs to be investigated further.

From here, the authors of [1] propose 3 different methods to transfer the contrast, I will only focus on Method 1, as that is the only method I attempted to implement so far.

Method 1 simply matches the histogram of $V_{LL,l}$ to the histogram of $N_{LL,l}$, resulting in $V'_{LL,l}$. The final resulting LL subband of the to-be enhanced image is defined by (6):

$$
V'_{LL} = W \cdot (V'_{LL,l} \cdot V_{LL,d}) + (1 - W) \cdot V_{LL}. \tag{6}
$$

- *Note for possible discussion:* I should mention that [1] writes (6) as $V'_{LL} = W \cdot (V'_{LL,l} + V_{LL,d}) + (1 - W) \cdot V_{LL}$ and Zhang is not clear whether those images are in the log domain or not; however, if we define $V_{LL,d}$ as in (4), we must use (6).

This new V'_{LL} will serve as the new LL subband when performing the inverse wavelet transform, as illustrated in Fig. 1.

4.2 My Results

I believe that I was able to accurately reproduce the method to obtain V'_{LL}.

I have added multiple figures documenting each step of the process:

- Figs. 3 and 4 show the one-level wavelet decomposition of the VIS and NIR images from the test image of Fig. 2.
- Fig. 5 shows the bilateral transform step and the resulting large-scale ($V_{LL,l}$ and $N_{LL,l}$) and detail layers ($V_{LL,d}$ and $N_{LL,d}$), obtained from (4) and (5).
- Fig. 6 shows the resulting histogram matching using Method 1. This image shows both global and local histogram matching. Zhang proposes the use of the global histogram matching; however, as recommended by Prof. Moulin, I also tried local histogram matching.
- Fig. 7 shows the final V'_{LL} obtained from (6) using the WRM obtained in the previous step.

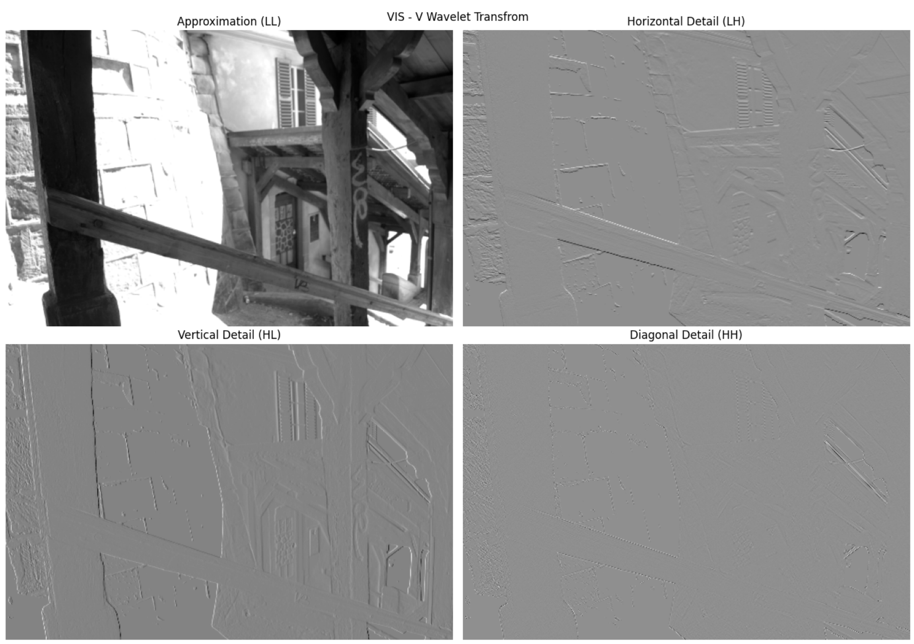

Figure 3: One-level wavelet decomposition for VIS test image from [2]

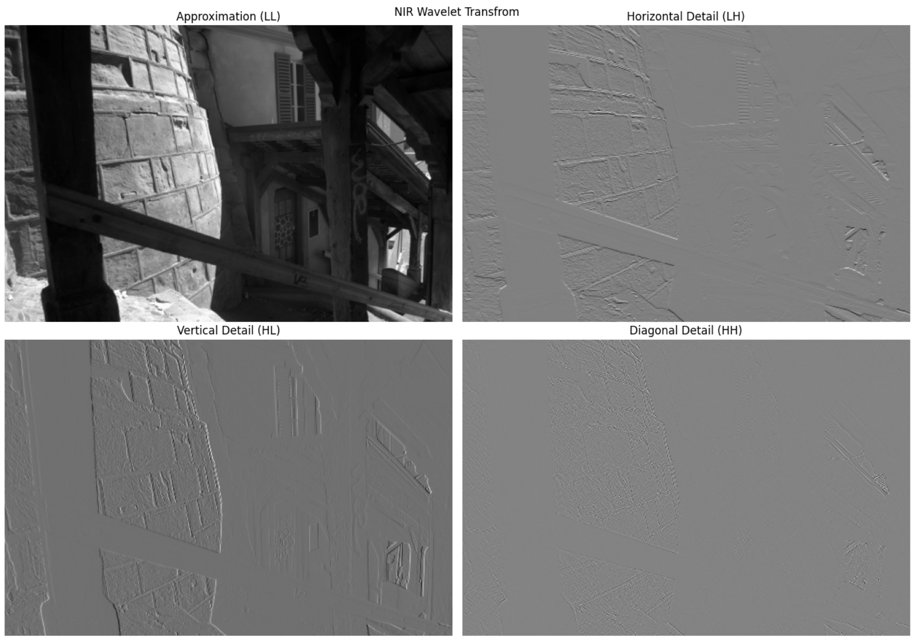

Figure 4: One-level wavelet decomposition for NIR test image from [2]

5 Transfer Texture

The goal in this step is to transfer the texture of the NIR image to the enhanced VIS image.

5.1 Zhang's Description

The TT step in Zhang's approach is straightforward: use the WRM for alpha-blending of the LH, HL, and HH subbands of NIR and VIS images; and, together with the LL subband given by V'_{LL}, obtain the enhanced image through the inverse wavelet decomposition, as indicated in Fig 1.

More precisely, the LH, HL, and HH subbands are obtained with:

$$V'_{LH} = W \cdot N_{LH} + (1 - W) \cdot V_{LH} \tag{7}$$
$$V'_{HL} = W \cdot N_{HL} + (1 - W) \cdot V_{HL} \tag{8}$$
$$V'_{HH} = W \cdot N_{HH} + (1 - W) \cdot V_{HH} \tag{9}$$
$$\tag{10}$$

and the brightness layer of the enhanced image is obtained by the inverse wavelet transform of the coefficients in V'_{LL}, V'_{LH}, V'_{HL}, and V'_{HH}.

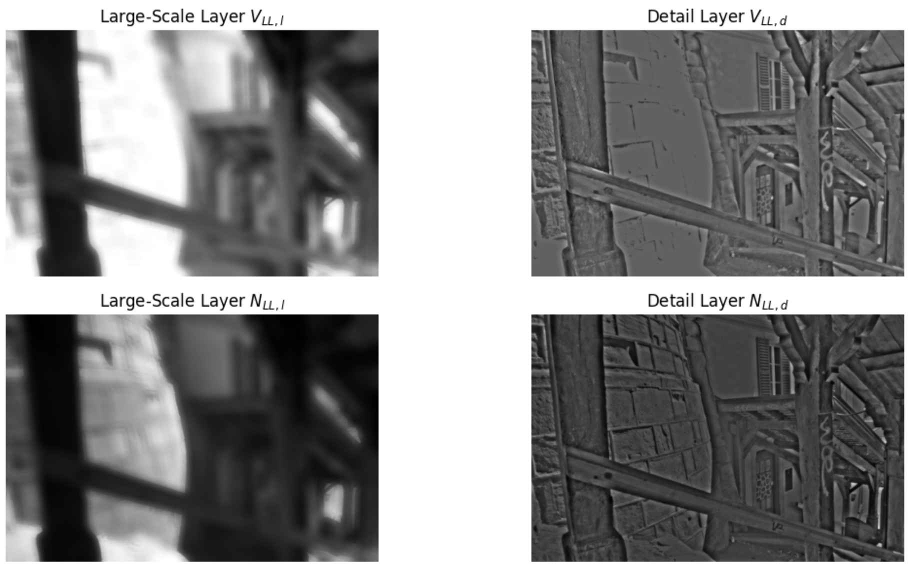

Figure 5: My resulting Bilateral Filtering for VIS and NIR test images from [2]

Lastly, the hue and saturation layers from the original VIS image were recombined with the resulting brightness layer to produce the final enhanced VIS color image.

5.2 My Results

I believe that I was able to accurately reproduce the method to obtain the various subbands and the enhanced image.

- Fig. 8 shows all the subbands before the inverse wavelet transform. For simplicity, Fig. 8 displays only the subbands using the local histogram matching to produce V'_{LL}.
- Fig. 9 shows the enhanced image obtained with local histogram matching. For comparison, the figure also displays the original image and the enhanced image obtained with global histogram matching.

From Fig. 9, it is possible to observe that the method did not have a significant impact on the test image.

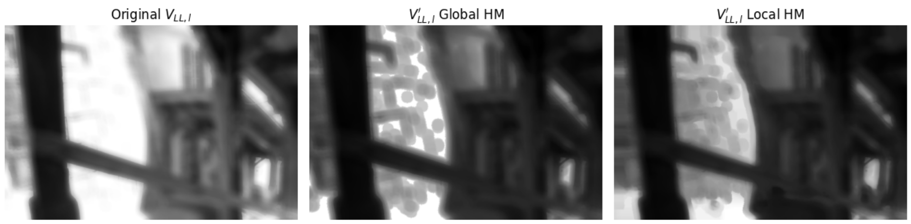

Figure 6: My resulting GLobal and Local Histogram Matching (HM) for test image from [2]. The local HM had window size of 32.

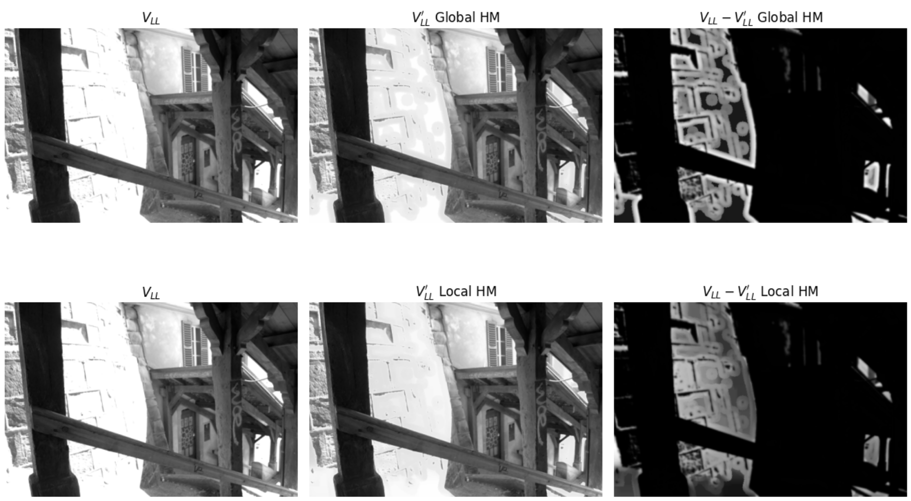

Figure 7: My resulting V'_{LL} considering Global and Local Histogram Matching (HM) for test image from [2]. Difference $V_{LL} - V'_{LL}$ also shown for both.

6 Current State of Project, Observations, Discussions, and Next Steps

Although the method did not show a significant impact on these initial tests, it is too early to judge the effectiveness of the method for two reasons: (1) these initial tests were performed only on a single test image; and (2) we need an objective metric to judge the method's effectiveness.

Although the paper was not clear on certain aspects of the transfer contrast step, I believe I was able to accurately recreate the main steps from what I could gather from [1]. I intend to discuss with Professor Moulin my interpretation of how to obtain the Detail Layer in the TC step of section 4.

Looking subjectively at the global and local histogram matching results, I

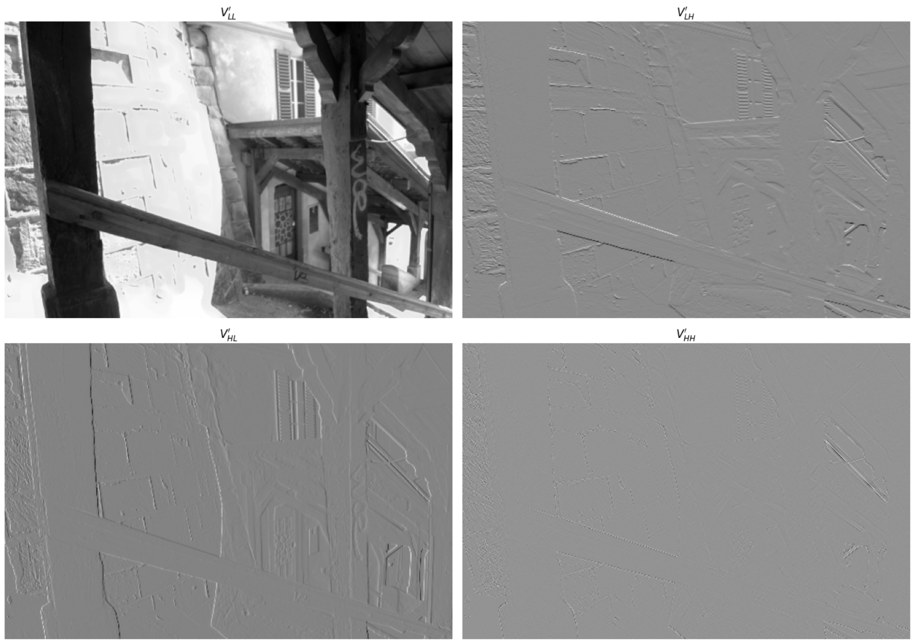

Figure 8: My resulting V'_{LL}, V'_{LH}, V'_{HL}, and V'_{HH} subbands considering Local Histogram Matching (HM) for test image from [2].

believe that the local histogram matching made a significant improvement as the edges of the areas where the image was "enhanced" were more accurate to the original colors and made the image look more faithful to the original image given the NIR details. Having said that, neither histogram matching schemes seem to have given a significant amount of details to the areas I was looking for.

For my next steps in the project, I intend to explore the following ideas:

- Explore how the this algorithm works with other images from [2] and images captured with my camera system.
- Implement and use objective measures to enable proper evaluation. I intend to implement some of the measures discussed in [4] or [5]
- Explore the idea of using more than a single-level wavelet decomposition for transferring contrast and texture.

References

[1] X. Zhang, T. Sim, and X. Miao, "Enhancing photographs with near infra-red images," in *2008 IEEE conference on computer vision and pattern recognition*. IEEE, 2008, pp. 1–8.

[2] E. IVRL, "RGB-NIR Scene Dataset," https://www.epfl.ch/labs/ivrl/research/downloads/rgb-nir-scene-dataset/, accessed: 2024-10-10.

[3] S. Bae, S. Paris, and F. Durand, "Two-scale tone management for photographic look," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 637–645, 2006.

[4] M. A. Herrera-Arellano, H. Peregrina-Barreto, and I. Terol-Villalobos, "Color outdoor image enhancement by v-nir fusion and weighted luminance," in *2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*. IEEE, 2019, pp. 1–6.

[5] A. V. Vanmali, S. G. Kelkar, and V. M. Gadre, "A novel approach for image dehazing combining visible-nir images," in *2015 Fifth national conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG)*. IEEE, 2015, pp. 1–4.

Figure 9: Final enhanced images for both Global and Local Histogram Matching (HM) for test image from [2]. Original image also shown for comparison.