

Revisiting Far/Near Infrared Pyramid-Based Fusion Types for Night Vision Using Matlab

Osama Mahfooz^{*}, Asadullah Shah^{**}

ABSTRACT

The night vision imaging mechanisms are developed to increase the visibility beyond normal human perception capabilities. So far, night vision methods reported in literature, such as, Morphological, Low Pass Pyramid, Contrast Pyramid, Filter Subtract Decimate and Shift Invariant methods, the Laplacian fusion method has been rated, the best method [1][2]. In this research paper four different methods of fusion of images, Gradient, Wavelet, Quincunx Lifting, including Laplacian are processed using Matlab toolbox called Matifus for night vision. For comparing the results of processed images using above methods, Mean Opinion Score (MOS) is used. MOS result of Laplacian, wavelet, gradient and quincunx methods are compared. The MOS results on the scale of 1-5 indicate a score of 4.15 for Laplacian, that means the quality of image perceived by the scorers is rated between good and excellent. By using MOS and perceptually proving that Laplacian technique is better than all others for night vision systems. However Gradient scored 3.56, Wavelet scored 3.15 and lastly 2.22 was scored by Quincunx Lifting method.

Key terms: FIR-Far Infrared, NIR-Near Infrared, Night Vision, Sensors-K-Kelvin, Micron, Optical Light, Mean Opinion Score (MOS).

1. INTRODUCTION

Night vision is the ability to see in dark surroundings. Whether by biological or technological means, night vision is made possible by a combination of two approaches: sufficient spectral range and sufficient intensity range as reported in [3].

^{*} Engr. Osama Mahfooz is Laboratory Engineer, Department of Telecommunication Engineering and Management at Institute of Business Management, Karachi, Pakistan. Email: osama_mahfooz14@hotmail.com

^{**} Dr.Asadullah Shah is Professor International Islamic University of Malaysia Kuliyah Of Information And Communication Technology Gombak , Malaysia. Email: dr_asadullahshah@hotmail.com

In order to understand night vision, it is important to understand something that the amount of energy in a light wave is related to its wavelength λ . Shorter wavelengths have higher energy. Of visible light, violet has the most energy and shorter λ , and red has the least or highest λ in visual spectrum. Just next to the visible light spectrum is the infrared spectrum [3].

Infrared light can be split into three categories:

- **Near-infrared (NIR)** - Closest to visible light, NIR has wavelengths that range from 0.7 to 1.3 **microns**, or 700 billionths to 1,300 billionths of a meter.
- **Mid-infrared (MIR)** - MIR has wavelengths ranging from 1.3 to 3 microns. Both NIR and MIR are used by a variety of electronic devices, including remote controls.
- **Thermal-infrared (TIR)** - is also known as FIR, occupying the largest part of the infrared spectrum, TIR (or FIR) has wavelengths ranging from 3 microns to 30 microns.

The key difference between TIR and the other two (NIR and MIR) is that TIR is emitted by an object instead of reflection [4][5].

Far and Near Infrared sensors provide FIR and NIR images for night vision [5]. Both have distinct features that are discussed in Section 2 along with sensors. Image fusion is a method used to combine images in such a way that all the salient information of the two images taken, is put together mostly into one image suitable for human perception or for further processing [5] and is discussed in Section 3. Results are explained in Section 4, while in Section 5 the applications of the above method are given, conclusion of work is given in Section 6 and finally in Section 7, details and directions for future work are provided.

2. NIGHT VISION SENSORS

2.1 The NIR

The NIR systems are termed active because they require illumination, that is, they operate on the principle of reflection of light signal. NIR wavelengths ranges from 0.7 μ micron to 1.1 μ micron, have been shown to provide images to support night vision. They give significant distance and contrast improvements over standard optical lighting systems.

These are the demerits of NIR as well: It has less range of visibility as compared to FIR. The sensor's ability to provide extended distance is limited by the intensity of the NIR illuminator. The range of this sensor depends on the intensity of the illuminator and the sensors ability to reduce blooming caused by approaching object [5].

2.2 The FIR

The FIR consists of radiation emitted in the range 8 – 12 μ micron band. In this band black bodies at a temperature of 300K have a maximum in energy emission. Therefore outdoor objects at ambient temperatures emit strongly in this band making it attractive for imaging road scenes. Bodies with internal heat sources such as pedestrians and vehicles in use appear very clearly. There is a minimum in atmospheric absorption for this band allowing, in suitable weather conditions, vision to the horizon, and sometimes improved visibility for hot objects in fog when compared to the eye. FIR has a dynamic range than NIR.

There are number of disadvantages in the use of FIR radiation for example in automotive applications. Thermal images have lower spatial resolution than images formed with NIR radiation, reducing sensitivity to textures. Thermal images generally have lower grey level. The energy emitted by the objects in the FIR band depends on their temperature. In an outdoor scene there are a number of factors that affect temperature and hence thermal contrast. These are the diurnal cycle, cloud cover, humidity, precipitation and wind. All of these effects alter and constantly change thermal contrast making road scenes viewed in the FIR difficult to interpret. Therefore the FIR image alone cannot be relied upon in night visions systems such as used for the automotive environment [5].

3. RESEARCH METHODOLOGY

Both NIR and FIR sensors have different qualities and characteristics regarding the night vision. The NIR image has better contrast, spatial resolution and a reliable image form but a shorter image range [5]. FIR sensor, on the other hand, can provide long range vision and enhances objects with internal heat sources; such as cars and pedestrians, but is susceptible to non-intuitive atmospheric and diurnal effects [5]. The combination of both sensors and their supportive features for the night vision provides significant enhancements for night vision imaging.

Matlab V7.0 with its toolbox Matifus is used in this research for fusion of night vision images from the above mentioned two sensor types using the available fusion methods.

The images of FIR and NIR sensors processed for this research are provided in the toolbox (Matifus) and the rest of the images given in this paper are the results generated by Matlab through fusion processes available in the software.

Experimental set up:

The objective of this research is to compare the results generated by Matlab and judge through the MOS results.

Image Quality Testing:

We used quality testing method called MOS to determine which method is more suitable for night vision imaging, judging subjectively the quality of images and determining the ratings of viewers. The MOS provides a numerical indication of the perceived quality of received media after processing. The MOS is expressed as a single number in the range 1 to 5, where 1 is lowest perceived image quality, and 5 is the highest perceived quality measurement.

There are three types of experimental setups. For first experiment set “A” the images with a resolution of **360 x 270** was used. For experiment set “B” the resolution of **505 x 510** and for experiment set “C”, the image resolution of **512 x 512** were considered. All three sets of images were different in nature.

Following methods were used to carry out image processing:

a. Steerable Pyramid

The Steerable Pyramid provides a linear multi-scale, multi-orientation image decomposition.

In the toolbox we allowed for a maximum of six orientations, implying the same number of bands. The transform is of over complete type [6][7].

The steerable pyramid transform introduced by Freeman et al. [8] is a linear multi-scale, multi orientation image decomposition that provides a useful front-end for image-processing and computer vision applications. The “steerable filter” refers to a class of filters, in which a filter of arbitrary orientation can be synthesized as a linear combination of a set of “basis filters” [9]

For any function

$$f(x, y), f^{\theta}(x, y) \text{ is } f(x, y)$$

rotated through an angle Θ about the origin. We call $f(x, y)$ is steerable if it satisfies the following equation:

$$f^{\theta}(x, y) = \sum_{j=1}^M k_j(\theta) f^{\theta_j}(x, y), \quad (3.1)$$

where $k(\Theta)$ are the interpolation functions ($j=1, \dots, M$). More details about steerable pyramid can be found in reference [8].

In our algorithm, four orientation band-pass components and two-level decomposition are adopted in the steerable pyramid algorithm. The interpolation functions $k(\Theta)$ can be expressed as follows based on Theorem 1 in [8]:

$$k_j(\theta) = \frac{1}{4}[2\cos(\theta - \theta_j) + 2\cos(3(\theta - \theta_j))], \quad (3.2)$$

b. Quincunx Lifting Scheme (LISQ)

The toolbox LISQ has been built on the lifting scheme and uses quincunx down sampling of rectangular two-dimensional grids. The lifting scheme has been invented by Sweldens [6]. Lifting scheme was used with quincunx down sampling to develop non-separable wavelets on a rectangular grid. The prediction filters include linear filters (Neville orders of increasing order) as well as a few nonlinear filters which preserve local maxima and / or minima[6][7][10].

Let $x_c(t)$, where $t = [t_1 \ t_2]^T$, be a two-dimensional function in the space domain, and $X_c(w)$, where $w = [w_1 \ w_2]^T$, be its representation in the frequency plane. The discretization operation of this function consists of obtaining samples of its values at fixed intervals of space. Given two linearly independent vectors d_1 and d_2 which define two directions along which samples are taken, we define $D = [d_1 \ d_2]^T$ as the sampling matrix, which generates the sampling lattice of the discrete version of $x_c(t)$. We can express this operation with the equation:

$$x[n] = x_c(Dn) \quad (3.3)$$

where $n = [n_1 \ n_2]^T$ is a vector of integer numbers. The most common sampling lattice is the separable one, which is described by the vectors $d_1 = [1 \ 0]^T$ and $d_2 = [0 \ 1]^T$. The sampling sublattice is determined by D as all the vectors Dn . The coset of a lattice is the set of points

obtained by shifting the entire lattice by an integer vector k . For a given lattice, there will be always $N = |\det D|$ cosets. In the case of sub band coding, N will also determine the number of subbands in each decomposition. It is possible to relate frequency representation of the

discrete function $x[n]$ with the representation in frequency Equation 3.3, we find

$$X(w) = \frac{1}{N} \sum_1 X_c(D^{-T}w - 2\pi D^{-T}l) \quad (3.4)$$

where D^{-T} denotes the inverse transpose of D . It is important to observe that Equation 3.4 represents the replication of the spectrum of x_c along the points represented by $2\pi D^{-T}l$, which shows that aliasing may occur if the signal is not band-limited. It is possible to show that aliasing will occur if the signal has components outside the region defined by

$$\{\mathbf{D}^{-T}\boldsymbol{\omega} : \boldsymbol{\omega} \in [-\pi, \pi] \times [-\pi, \pi]\} \quad (3.5)$$

The down sampling operation of a discrete signal on an arbitrary lattice is represented by

$$y[\mathbf{n}] = x[\mathbf{D}\mathbf{n}] \quad (3.6)$$

Following the same reasoning used in Equations 3.5, we can relate the spectrum of $y[\mathbf{n}]$ with that of $x[\mathbf{n}]$. We will find that

$$Y(\boldsymbol{\omega}) = \frac{1}{N} \sum_{i=0}^{N-1} X(\mathbf{D}^{-T}\boldsymbol{\omega} + 2\pi\mathbf{D}^{-T}\mathbf{k}_i) \quad (3.7)$$

where \mathbf{k}_i is the integer vector that represents the i -th coset of the sampling lattice.

The upsampling operation is defined as the inverse Operation

$$y[\mathbf{n}] = x[\mathbf{D}^{-T}\mathbf{n}] \quad (3.8)$$

with the constraint that this operation is only valid where $\mathbf{D}^{-T}\mathbf{n}$ maps to valid points of $x[\mathbf{n}]$, being defined as 0 when this doesn't occur. Its frequency representation is given by

$$Y(\boldsymbol{\omega}) = X(\mathbf{D}^T\boldsymbol{\omega}) \quad (3.9)$$

We are interested in quincunx sampling, which is described by vectors $\mathbf{q}_1 = [1 \ 1]^T$ and $\mathbf{q}_2 = [1 \ -1]^T$. We define the sampling matrix $\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2]^T$, below, as the quincunx sampling matrix

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (3.10)$$

c. Laplacian Pyramid

The Laplacian Pyramid is the classic hard-to-beat approach in the analysis and synthesis of images. It leads to an over complete representation. The Laplace operator is a second order differential operator in the n -dimensional Euclidean space, defined as the divergence ($\nabla \cdot$) of the gradient (∇f) [6][7].

Laplacian pyramid breaks the image in several levels. It samples the image through the high pass filters. The top level of the image contains highest spatial frequency the edges, the intermediate level contains the transition from high spatial frequency to low spatial frequency, and the low level contains the smooth part of the image. Laplacian pyramid technique can be improved by doing spatial differentiation. This technique has to be applied on both the images. The 1st order differentiation is

$$\frac{\partial f}{\partial x} = f(x+1) - f(x) \quad \frac{\partial f}{\partial x} = f(x+1) - f(x) \quad (3.11)$$

It is the difference between subsequent values and rate of change of the function. The formula for the 2nd order derivative is

$$\frac{\partial^2 f}{\partial^2 x} = f(x+1) + f(x-1) - 2f(x) \quad (3.12)$$

2nd order derivative gives the fine details for e.g thin lines. For getting the details of the image laplacian filter is used. The calculation for the laplacian filter is as follows:

The laplacian is defined as

$$\nabla^2 f = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y} \quad (3.13)$$

2nd order derivative along x-axis

$$\frac{\partial^2 f}{\partial^2 x} = f(x+1, y) + f(x-1, y) - 2f(x, y) \quad (3.14)$$

2nd order derivative along y-axis

$$\frac{\partial^2 f}{\partial^2 y} = f(x, y+1) + f(x, y-1) - 2f(x, y) \quad (3.15)$$

So the the resultant laplacian is

$$\nabla^2 f = [f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1)] - 4f(x, y) \quad (3.16)$$

The filter based on this calculation is

0	1	0
1	-4	1
0	1	0

Fig (a) First filter

After this when laplacian filter is applied to an image and it will give the image with highlight edges. But for an enhanced image the laplacian image is subtracted from an original image to get the sharp image. The overall calculation will be like this

$$\begin{aligned}
 g(x, y) &= f(x, y) - \nabla^2 f \\
 &= f(x, y) - [f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y)] \\
 &= 5f(x, y) - f(x+1, y) - f(x-1, y) - f(x, y+1) - f(x, y-1)
 \end{aligned} \tag{3.17}$$

0	-1	0
-1	5	-1
0	-1	0

Fig (b) Resultant filter

After getting the laplacian filtered image 1D and 2D median filtering technique can be applied to denoise the image from impulse noise.

After applying the median filter and getting the sharp and clear image laplacian pyramid technique is implemented. This is how the result of the laplacian pyramid technique is improved.

d. Gradient Pyramid

The Gradient Pyramid constructs Gaussian pyramids by convolution and down sampling and then applies gradient filters in the horizontal, vertical and two diagonal directions[6][7].

The regions of high spatial variance across one image are computed by thresholding the intensity gradients, $G = (G_X, G_Y)$, for the horizontal and vertical directions using a simple forward difference. The regions of high temporal variance between two images are computed by comparing the intensity gradients of corresponding pixels from the two images. We then compute an importance image (a weighting function) W , by processing the gradient magnitude $|G|$. The weighted combination of input gradients gives us the gradient of the desired output. The basic steps are described in [12]

Basic algorithm

For each input image I_i

Find gradient field $G_i = I_i$

Compute importance image W_i from $|G_i|$

and for each pixel (x, y)

Compute mixed gradient field $G(x, y) =$

$$\sum_i W_i(x, y) G_i(x, y) / \sum_i W_i(x, y)$$

Reconstruct image I from gradient field G

Normalize pixel intensities in I to closely match $\sum_i W_i I_i$

4. EXPERIMENTAL RESULTS

Following are the results observed after implementing fusion techniques mentioned in Section 3. The MOS is conducted by 30 highly experienced individuals who graded each resultant image placed randomly side by side, with respect to clarity, texture, boundary edges, resolution and brightness.

4.a EXPERIMENTAL RESULTS FOR IMAGE SET A

The source images of NIR and FIR are shown in Figure 1 and 2 respectively. Figures 3,4,5 and 6 are processed images by using Laplacian, Wavelet, Gradient and Quincunx Lifting Scheme respectively. The MOS test as for given set of pictures is provided in Table 1, indicates that the Gradient method scored second highest rating 3.86, while Wavelet achieved a 2.25 rating-the lowest one



Figure 1. The FIR Source Image [6]



Figure 2. The NIR Source Image [6]



Figure 3. Fusion Using Laplacian Method



Figure 4. Fusion Using Wavelet Method



Figure 5. Fusion Using Gradient Method



Figure 6. Quincunx Lifting Scheme Method

4.b EXPERIMENTAL RESULTS FOR IMAGE SET B

In this experiment two clock images were taken as source images with FIR and NIR sensors as shown in Figure 7 and 8 respectively of this set “B”. Images obtained by using fusion methods are represented in Figure 9-12. The MOS test as given in Table 2, indicate that the Laplacian technique has been scored the highest rating 4.37, while the lowest Quincunx Lifting method scored lowest i.e 1.9.



Figure 7: FIR Source image [6]



Figure 8: NIR Source image [6]



Figure 9: Fusion using the Laplacian method



Figure 10: Fusion using the Wavelet method



Figure 11: Fusion using the Gradient method



Figure 12: Quincunx Lifting Scheme method

4.c EXPERIMENTAL RESULTS FOR IMAGE SET C

In this experiment FIR and NIR images of a floating ship have been taken as source images as shown in Figure 13 and 14. Figure 15-18 are the images processed by the fusion methods. The MOS test as given in Table 3, indicate that the Laplacian technique has been scored the highest rating 4.4, second highest is scored for Wavelet method i.e. 3.31, while the lowest score 1.76 was acquired by Quincunx Lifting method.



Figure 13: The FIR Source image [6]



Figure 14: The NIR image [6]



Figure 15: Fusion Using Laplacian Method



Figure 16: Fusion Using Wavelet



Figure 17: The Gradient Method



Figure 18: Quincunx Lifting Scheme

Table1: MOS results of images set A

Experiment Set A	Average MOS
Laplacian	3.68
Wavelet	2.25
Gradient	3.86
Quincunx	3.02

Table 2: MOS results of images set B

Experiment Set B	Average MOS
Laplacian	4.37
Wavelet	4.0
Gradient	3.58
Quincunx	1.9

Table 3: MOS results of images set C

Experiment Set C	Average MOS
Laplacian	4.4
Wavelet	3.31
Gradient	3.27
Quincunx	1.76

5. APPLICATIONS

The original purpose of night vision systems was to locate enemy targets at night. However, currently there are many applications of night vision and systems that include applications include military operations, hunting, wildlife observation, surveillance, security, navigation and hidden-object detection and even night vehicular driving.

Military personnel use these systems for navigation, surveillance and targeting. Police and security often use both thermal-imaging and image-enhancement technology, particularly for surveillance. Hunters and nature enthusiasts use NVDs (Night vision devices) to maneuver through the woods at night. Detectives and private investigators use night vision to watch people they are assigned to track. Many businesses have permanently-mounted cameras equipped with night vision to monitor the surroundings.

A really amazing ability of thermal imaging is that it reveals if an area has been disturbed. It can show that the ground has been dug up to bury something, even if there is no obvious sign to the naked eye. Law enforcement has used this to discover items that have been hidden by criminals, including money, drugs and bodies. Also, recent changes to areas such as walls can be seen using thermal imaging, which has provided important clues in several criminal cases [4].

6. CONCLUSION

Three different source images set were taken by FIR and NIR sensors. All three sets of images had different dimensions and contents, including fore and background contents.

On average it may be concluded that the fusion methods are rated as 4.15 for Laplacian , 3.56 for Gradient , 3.15 for Wavelet and 2.2 for Quincunx Lifting method.

REFERENCES

- [1.] Jahard.F , Fish D.A, Rio A.A and Thompson C.P (July-1997), “Far/Near Infrared Adapted Pyramid-Based Fusion for Automotive Night Vision”, Applied Mathematics and Computing Group, Cranfield University, Cranfield, UK.
- [2.] Yin Chen, Rick S. Blum,(April-2008) ”Experimental Tests of Image Fusion for Night Vision”, ECE Department ECE Department, Lehigh University Lehigh University, Bethlehem, PA 18015, US.
- [3.] “Night vision”
http://en.wikipedia.org/wiki/Night_vision accessed 6-July-2010

- [4.] Jeff Tyson ,“How Night Vision Works”,
<http://electronics.howstuffworks.com/nightvision.htm> accessed 10-July-2010

- [5.] Wolfe W.L. and Zissis G.J., The Infrared Handbook, Washington, 1985

- [6.] A matlab toolbox for image fusion (MATIFUS)

- [7.] Image Fusion Methods
http://en.wikipedia.org/wiki/Image_fusion accessed 12-July-2010

- [8.] W. T. Freeman, E. H. Adelson, “The design and use of steerable filters”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 13, pp. 891-907, 1991.

- [9.] Xiangzeng Liu, Zheng Tian (November 2009) “Registration of Remote Sensing Images with Steerable Pyramid Transform and Robust SIFT Features” Proceedings of the 2009 International Workshop on Information Security and Application (IWISA 2009) Qingdao, China

- [10.] José Alexandre Nalon, João Baptista T. Yabu-Uti (October 2002) “Compression of Quincunx Subbands” Unicamp – Universidade Estadual de Campinas DECOM – Departamento de Comunicações Av. Albert Einstein, 400 – CEP 13081-970 – Campinas SP

- [11.] Ramesh Raskar Adrian Ilie (2005) “Image fusion for context enhancement and video surrealism” Published in: · Proceeding SIGGRAPH '05 ACM SIGGRAPH