21

Iterative Multiscale Fusion and Night Vision Colorization of Multispectral Images

Yufeng Zheng
Alcorn State University
USA

1. Introduction

Multispectral images usually present complimentary information such as visual-band imagery and infrared imagery (near infrared or long wave infrared). There is strong evidence that the fused multispectral imagery increases the reliability of interpretation (Rogers & Wood, 1990; Essock et al., 2001); whereas the colorized multispectral imagery improves observer performance and reaction times (Toet et al. 1997; Varga, 1999; Waxman et al., 1996). A fused image in grayscale can be automatically analyzed by computers (for target recognition); while a colorized image in color can be easily interpreted by human users (for visual analysis).

Imagine a nighttime navigation task that may be executed by an aircraft equipped with a multisenor imaging system. Analyzing the combined or synthesized multisensory data will be more convenient and more efficient than simultaneously monitoring multispectral images such as visual-band imagery (e.g., image intensified, II), near infrared (NIR) imagery, and infrared (IR) imagery. In this chapter, we will discuss how to synthesize the multisensory data using image fusion and night vision colorization techniques in order to improve the effectiveness and utility of multisensor imagery. It is anticipated that the successful applications of such an image synthesis approach will lead to improved performance of remote sensing, nighttime navigation, target detection, and situational awareness. This image synthesis approach involves two main techniques, image fusion and night vision colorization, which is reviewed as follows, respectively.

Image fusion combines multiple-source imagery by integrating complementary data in order to enhance the information apparent in the respective source images, as well as to increase the reliability of interpretation. This results in more accurate data (Keys et al., 1990) and increased utility (Rogers & Wood, 1990; Essock et al., 1999). In addition, it has been reported that fused data provides far more robust aspects of operational performance such as increased confidence, reduced ambiguity, improved reliability and improved classification (Rogers & Wood, 1990; Essock et al., 2001). A general framework of image fusion can be found in Reference (Pohl & Genderen, 1998). In this chapter, our discussions focus on *pixel-level* image fusion. A quantitative evaluation of fused image quality is important for an objective comparison between the respective fusion algorithms, which measures the amount of useful information and the amount of artifacts introduced in the fused image.

Two common fusion methods are the discrete wavelet transform (DWT) (Pu & Ni, 2000; Nunez et al., 1999) and various pyramids (such as Laplacian, contrast, gradient, and morphological pyramids) (Jahard et al., 1997; Ajazzi et al., 1998), which both are *multiscale fusion* methods. Recently, an advanced wavelet transform (*a*DWT) method (Zheng et al., 2004) has been proposed, which incorporates principal component analysis (PCA) and morphological processing into a regular DWT fusion algorithm. The *a*DWT method can produce a better fused image in comparison with pyramid methods and regular DWT methods. Experiments also reveal an important relationship between the fused image quality and the wavelet properties. That is, a higher level of DWT decomposition (with smaller image resolution at a higher scale) or a lower order of wavelets (with a shorter length) usually results in a more sharpened fused image. This means that we can use the level of DWT decomposition and the length of a wavelet as the control parameters of an *iterative* DWT-based fusion algorithm.

So far, only a few metrics are available for quantitative evaluation of the quality of fused imagery. For example, the root mean square error (RMSE) may be the natural measure of image quality if a "ground truth" image is available. Unfortunately, for realistic image fusion applications there are no ground truths. Piella et al. (2003) presented an image fusion metric, the image quality index (IQI), which measures how similar the fused image is to both input images. More recently, Zheng et al. (2007) proposed an image quality metric, termed as "the ratio of SF error (rSFe)", which is a relative measurement regardless of the type of image being analyzed. The rSFe metric is defined upon "spatial frequency" (SF) (Eskicioglu & Fisher, 1995). In addition, the rSFe value can show the fusion status (i.e., under-fused or over-fused). Refer to Section 2.3 for a review of fusion metrics.

On the other hand, a *night vision colorization* technique can produce colorized imagery with a naturalistic and stable color appearance by processing multispectral night-vision imagery. Although appropriately false-colored imagery is often helpful for human observers in improving their performance on scene classification, and reaction time tasks (Essock et al., 1999; Waxman et al., 1996), inappropriate color mappings can also be detrimental to human performance (Toet & IJspeert, 2001; Varga, 1999). A possible reason is lack of physical color constancy (Varga, 1999). Another drawback with false coloring is that observers need specific training with each of the unnatural false color schemes so that they can correctly and quickly recognize objects; whereas with colorized nighttime imagery rendered with natural colors, users should be able to readily recognize and identify objects.

Toet (2003) proposed a night vision (NV) colorization method that transfers the natural color characteristics of daylight imagery into multispectral NV images. Essentially, Toet's natural color-mapping method matches the statistical properties (i.e., mean and standard deviation) of the NV imagery to that of a natural daylight color image (manually selected as the "target" color distribution). However, this color-mapping method colorizes the image regardless of scene content, and thus the accuracy of the coloring is very much dependent on how well the target and source images are matched. Specifically, Toet's method weights the local regions of the source image by the "global" color statistics of the target image, and thus will yield less naturalistic results (e.g., biased colors) for images containing regions that differ significantly in their colored content. Another concern of Toet's "*global-coloring*" method is that the scene matching between the source and target is performed manually. To address the aforementioned bias problem in global coloring, Zheng et al. (2005; 2008) presented a "*local coloring*" method that can colorize the NV images more like daylight imagery. The local-coloring method will render the multispectral images with natural colors segment by segment (also referred to as "segmentation-based"), and also provide automatic association between the source and target images (i.e., avoiding the manual scene-matching in global coloring).

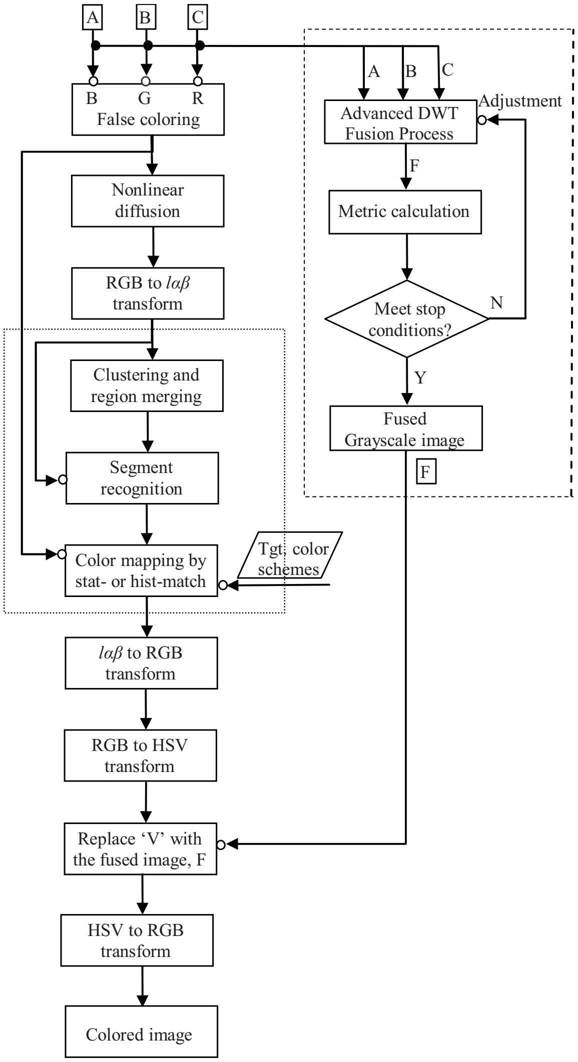


Fig. 1. The diagram of image fusion and night vision colorization. The *iterative image fusion* (shown within the right dashed box) takes multispectral images (A, B, C) as inputs, and fuses them into a grayscale image, F. The *night vision colorization* (shown in the left column) takes the same multispectral images (A, B, C) and also the fused image F as inputs, and generates a colored image. Three steps shown inside a dotted rectangle are performed in the $l\alpha\beta$ color space.

In this chapter, a joint approach that incorporates image fusion and night vision colorization is presented to synthesize and enhance multisensor imagery. This joint approach provides two sets of synthesized images, fused image in grayscale and colored image in colors using the image fusion procedure and night vision colorization procedure. As shown in Fig. 1, the *image fusion* (shown in the right dashed box) takes multispectral images (A, B, C) as inputs and fuses them into a grayscale image (F). The *night vision colorization* (shown in the left column) takes the same multispectral images (A, B, C) and the fused image (F) as an input and eventually generates a colored image. The image fusion process can take more than three bands of images; whereas the night vision colorization can accept three (or less) bands of images. If there are more than three bands of images available, (e.g. II, NIR, MWIR (medium-wave IR) and LWIR (long-wave IR)), we may choose a visual band image (II) and two bands of IR images for the following colorization (refer to Section 4 for a detailed discussion). Two procedures are discussed respectively in Sections 2 and 3. Note that in this chapter, the term "multispectral" is used equivalently to "multisensory"; and by default the term "IR" means "LWIR" unless specified.

The remainder of this chapter is organized as follows: The multiscale image fusion methods are discussed in Section 2. Image quality metrics are also reviewed in this section. The night vision colorization methods are fully described in Section 3. The experiments and discussions are given in Section 4. Finally, conclusions are made in Section 5.

2. Multiscale image fusion

Image fusion serves to combine multiple-source imagery using advanced image processing techniques. In this section, Laplacian pyramid and DWT fusion methods are briefly reviewed, then an advanced discrete wavelet transform (*a*DWT) method is introduced. Next, several established fusion metrics are discussed. With an established fusion metric, an iterative *a*DWT fusion process (i.e., *a*DWTi) can be formed. Finally, a newly proposed orientation based fusion is described. The orientation based fusion is performed by using Gabor wavelet transforms, which may be considered a multiscale procedure in frequency domain.

2.1 Laplacian pyramid

The Laplacian pyramid was first introduced as a model for binocular fusion in human stereo vision (Burt & Adelson, 1985), where the implementation used a Laplacian pyramid and a maximum selection rule at each point of the pyramid transform. Essentially, the procedure involves a set of band-pass copies of an image is referred to as the Laplacian pyramid due to its similarity to a Laplacian operator. Each level of the Laplacian pyramid is recursively constructed from its lower level by applying the following four basic steps: blurring (low-pass filtering); sub-sampling (reduce size); interpolation (expand); and differencing (to subtract two images pixel by pixel) (Burt & Adelson, 1983). In the Laplacian pyramid, the lowest level of the pyramid is constructed from the original image.

2.2 The advanced DWT method

The regular DWT method is a multi-scale analysis method. In a regular DWT fusion process, DWT coefficients from two input images are fused pixel-by-pixel by choosing the average of the approximation coefficients (i.e., the low-pass filtered image) at the highest transform scale; and the larger absolute value of the detail coefficients (i.e., the high-pass filtered images) at each transform scale. Then, an inverse DWT is performed to obtain a fused image. At each DWT scale of a particular image, the DWT coefficients of a 2D image consist of four parts: approximation, horizontal detail, vertical detail, and diagonal detail. In the *advanced DWT* (*a*DWT) method (Zheng et al., 2004), we apply PCA to the two input images' approximation coefficients at the highest transform scale. That is, we fuse them using the principal eigenvector (corresponding to the larger eigenvalue) derived from the two original images, as described in Eq. (1) below:

$$C_F = (a_1 \cdot C_A + a_2 \cdot C_B)/(a_1 + a_2), \tag{1}$$

where C_A and C_B are approximation coefficients (image matrices) transformed from input images A and B. C_F represents the fused coefficients; a_1 and a_2 are the elements (scalars) of the principal eigenvector, which are computed by analyzing the original input images. Note that the denominator in Eq. (1) is used for normalization so that the fused image has the same energy distribution as the original input images.

For the detail coefficients (the other three quarters of the coefficients) at each transform scale, the larger absolute values are selected, followed by neighborhood morphological processing, which serves to verify the selected pixels using a "filling" and "cleaning" operation (i.e., the operation fills or removes isolated pixels locally). Such an operation (similar to smoothing) can increase the consistency of coefficient selection thereby reducing the distortion in the fused image.

2.3 Image quality metrics

2.3.1 Image quality index

The *image quality index* (IQI) was introduced by Wang and Bovik (2002). Given two sequences $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, let $\bar{x}$ denote the mean of x, and σ_x and σ_{xy} denote the variance of x and covariance of x and y, respectively. The global quality index of two vectors is defined as

$$Q_0(x,y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\overline{xy}}{(\bar{x}^2 + \bar{y}^2)} \cdot \frac{2\sigma_x \sigma_y}{(\sigma_x^2 + \sigma_y^2)} = \frac{4\sigma_{xy}\overline{xy}}{(\bar{x}^2 + \bar{y}^2)(\sigma_x^2 + \sigma_y^2)}, \tag{2}$$

Note that $Q_0 \in [0, 1]$ can reflect the correlation (similarity), luminance distortion, and contrast distortion between vectors x and y, which correspond to the three components (factors) in Eq. (2). Keep in mind that for the image quality evaluation with Q_0, the values x_i, y_i are positive grayscale values. The maximum value $Q_0 = 1$ is achieved when x and y are identical.

Then, the fused image quality metric (i.e., the image quality index) (Wang & Bovik, 2002; Piella & Heijmans, 2003) can be defined as

$$Q_w = \lambda Q_0(I_A, I_F) + (1-\lambda)\, Q_0(I_B, I_F), \tag{3}$$

where subscripts A, B, and F denote the input images (A, B) and the fused images (F); and weight $\lambda = S(I_A) / [S(I_A) + S(I_B)]$. $S(I_A)$ denotes the "saliency" of image A, which may be the local variance, $S(I_A) = \sigma_A$. Since image signals are generally non-stationary, it is more appropriate to measure the weighted image quality index Q_w over local regions (e.g., using a sliding window) and then combine the different results into a single measure.

2.3.2 Spatial frequency and the ratio of spatial frequency error

The metric "*spatial frequency*" (SF) (Eskicioglu & Fisher, 1995; Li et al., 2001) is used to measure the overall activity level of an image. The spatial frequency of an image is defined as

$$SF = \sqrt{[(RF)^2 + (CF)^2 + (MDF)^2 + (SDF)^2]/(4-1)}\ , \tag{4}$$

where RF and CF are row frequency and column frequency, respectively; and MDF and SDF represent the main diagonal SF and the secondary diagonal SF. Eq. (4) is a revision of the original definition (Zheng et al., 2007) for spatial frequency by introducing two diagonal SFs and also the normalization of the degree of freedom. Four directional spatial frequencies are defined as follows:

$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[I(i,j)-I(i,j-1)]^2}\ , \tag{5a}$$

$$CF = \sqrt{\frac{1}{MN}\sum_{j=1}^{N}\sum_{i=2}^{M}[I(i,j)-I(i-1,j)]^2}\ ; \tag{5b}$$

$$MDF = \sqrt{w_d \cdot \frac{1}{MN}\sum_{i=2}^{M}\sum_{j=2}^{N}[I(i,j)-I(i-1,j-1)]^2}\ , \tag{5c}$$

$$SDF = \sqrt{w_d \cdot \frac{1}{MN}\sum_{j=1}^{N-1}\sum_{i=2}^{M}[I(i,j)-I(i-1,j+1)]^2}\ ; \tag{5d}$$

where $w_d = 1/\sqrt{2}$ in Eqs. (5c-d) is a distance weight; similarly, it can be considered that $w_d = 1$ in Eqs. (5a-b). M and N are the image size (in pixels). Notice that the term "spatial frequency" is computed in the spatial domain as defined in Eqs. (4-5), does not correspond with the Fourier transform where the spatial frequency is measured in frequency domain with the unit of "cycles per degree" or "cycles per millimeter".

With Eq. (4) we can calculate the SFs of input images (SF_A and SF_B) or of the fused image (SF_F). Now we determine how to calculate a reference SF (SF_R) with which the SF_F can be compared. The four differences (inside square brackets) defined in Eqs. (5a-d) are actually the four first-order gradients along four directions at that pixel, denoted as *Grad*[I(i,j)]. The four reference gradients can be obtained by taking the maximum of absolute gradient values between input image A and B along four directions:

$$Grad^D[I_R(i,j)] = \max\{|Grad^D[I_A(i,j)]|,\ |Grad^D[I_B(i,j)]|\},$$

for each of four directions, i.e., D = {H, V, MD, SD}, (6)

where 'D' denotes one of four directions (Horizontal, Vertical, Main Diagonal, and Secondary Diagonal). Substituting the differences (defined inside square brackets) in Eqs. (5a-d) with $Grad^D[I_R(i,j)]$, four directional reference SFs (i.e., RF_R, CF_R, MDF_R and SDF_R) can be calculated. For example, the reference row frequency, RF_R, can be calculated as follows:

$$RF_R = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[Grad^H(I_R(i,j))]^2}\ . \tag{7}$$

Similar to Eq. (4), the overall *reference spatial frequency*, SF_R, can be computed by combining four directional reference SFs (SF_R is not formulated here). Note that the notation of

"$Grad^H[I_R(i,j)]$" is interpreted as "the horizontal reference gradient at point (i,j)", and no reference image is needed to compute the SF_R value.
Finally, the *ratio of SF error* (rSFe) is defined as follows:

$$\mathrm{rSFe} = (SF_F - SF_R) / SF_R, \tag{8}$$

where SF_F is the spatial frequency of the fused image; whereas SF_R is the reference spatial frequency. Clearly, an ideal fusion has rSFe = 0; that is, the smaller the rSFe's absolute value, the better the fused image. Furthermore, rSFe > 0 means that an *over-fused* image, with some distortion or noise introduced, has resulted; rSFe < 0 denotes that an *under-fused* image, with some meaningful information lost, has been produced.

2.4 The iterative *a*DWT method

The IQI (Wang & Bovik, 2002; Piella & Heijmans, 2003) value is calculated to measure the fused image quality by the *a*DWT. It is then fed back to the fusion algorithm (*a*DWT) in order to achieve a better fusion by directing the parameter adjustment. Previous experiments (Zheng et al., 2004) showed that a higher level DWT decomposition (with lower image resolution at higher scale) or a lower order of wavelets (with shorter length) usually resulted in a more sharpened fused image. The IQI value usually tends to be larger for a fused image with a lower level decomposition or a higher order of wavelets. This means that we can use the level of DWT decomposition and the length of a wavelet as control parameters of an *iterative aDWT* (*a*DWTi) algorithm. With the definition of IQI, we know that it has an ideal value, 1, i.e., $0 < \mathrm{IQI} \le 1$. The level of DWT decomposition (L_d) is a more significant factor than the length of wavelet (L_w) in the sense of the amplitude of IQI changing. The iterative *a*DWT algorithm optimized by the IQI is denoted as *a*DWTi-IQI. Similarly, *a*DWTi-rSFe means an iterative *a*DWT optimized by rSFe metric.
Of course, some termination conditions are needed in order to stop the fusion iteration. The following conditions are demonstrated with IQI metric. For example, the fusion iteration stops when (1) it converges at the ideal value – the absolute value of (IQI-1) is smaller than a designated small tolerance error, i.e. $|\mathrm{IQI}-1| < \varepsilon$; (2) there is no significant change of the IQI value between two consecutive iterations; (3) the IQI value is generally decreasing for subsequent iterations; or (4) the parameters' boundaries are reached. In implementing an iterative fusion procedure, appropriate parameters' initializations and boundaries' restrictions should be designated upon the definition of parameters and the context, which will help reduce the number of iterations (N_i). The details of implementation are described in Reference (Zheng et al., 2005).
The iterative *a*DWT algorithm hereby described can be combined with the rSFe metric (*a*DWTi-rSFe) (Zheng et al., 2007) or other fusion IQ metrics.

2.5 Orientation based fusion

Gabor wavelet transforms (GWT) have received considerable attentions because the characteristics of certain cells in the visual cortex of some mammals can be approximated by these filters. Further, biological research suggests that the primary visual cortex performs a similar orientational and Fourier space decomposition (Jones & Palmer, 1987), so they seem to be sensible for a technical vision or a recognition system. The details of GWT implementation are described elsewhere (Zheng & Agyepong, 2007).

In the *orientation-based fusion* algorithm, the Gabor wavelet transforms are performed with each input image at M spatial frequencies by N orientations, notated as $M\times N$. For a 16×16 GWT, a total of 256 pairs (magnitudes and phases) of filtered images are extracted with 256 Gabor wavelets (also called Gabor kernels, or Gabor filter bank) distributed along 16 bands (located from low to high frequencies) by 16 orientations (0.00°, 11.25°, 22.50°, ..., 157.50°, 168.75°). The size of each Gabor filter should match the image size being analyzed. If all input images are of the same size, then the set of 256 Gabor wavelets are only computed once. Instead of doing spatial convolution, the GWT can be accomplished in frequency domain by using fast Fourier transforms (FFT) that will significantly speed up the process.

Many GWT coefficients are produced, for example, 512 coefficients (256 magnitudes plus 256 phases) per pixel in an 16×16 GWT. Suppose a set of $M\times N$ GWT are performed with two input images (I_A and I_B). At each frequency band ($b = 1, 2, \ldots, M$), the index of *maximal* GWT magnitude between two images is selected pixel by pixel; and then two index frequencies, $H_A(b)$ and $H_B(b)$, are calculated as its index accumulation along N orientations, respectively. The final H_A and H_B are the weighted summations through M bands, where the band weights (W_b) are given empirically. Eventually, the fused image (I_F) is computed as

$$I_F = (I_A .* H_A + I_B .* H_B)/(H_A + H_B), \tag{9}$$

where '.*' denotes element-by-element product of two arrays; and

$$H_A = \sum_{b=1}^{M} W_b H_A(b)', \tag{10a}$$

$$H_B = \sum_{b=1}^{M} W_b H_B(b)', \tag{10b}$$

where W_b are the band weights decided empirically. The middle frequency bands (Hollingsworth et al., 2009) in GWT (by suppressing the extreme low and extreme high frequency bands) usually give a better representation and consistency in image fusion, especially for noisy input images.

The orientation-based fusion algorithm can be further varied by either keeping DC (direct current) or suppressing DC in GWT. "Keeping DC" will produce a contrast-smooth image (suitable for contrast-unlike images); while "suppressing DC" (i.e., forcing DC = 0.0) will result a sharpened fusion (suitable for contrast-alike images). *Color fusion* can be achieved by replacing the red channel of a color image with the fused image of red channel and LWIR image, which is suitable for poorly illuminated color images.

3. Night vision colorization

The aim of *night vision colorization* is to give multispectral (NV) images (source) the appearance of normal daylight color images. The proposed "local coloring" method renders the multispectral images segment-by-segment with the statistical properties of natural scenes using the color mapping technique. The main steps of the *local coloring* procedure are given below: (1) A false-color image (source image) is first formed by assigning multispectral (two or three band) images to three RGB channels. The false-colored images usually have an unnatural color appearance. (2) Then, the false-colored image is segmented using the features of color properties, and the techniques of nonlinear diffusion, clustering, and region merging. (3) The averaged mean, standard deviation, and histogram of a large sample of natural color images are used as the target color properties for each color scheme.

The target color schemes are grouped by their contents and colors such as plants, mountain, roads, sky, water, buildings, people, etc. (4) The association between the source region segments and target color schemes is carried out automatically utilizing a classification algorithm such as the nearest neighbor paradigm. (5) The color mapping procedures (statistic-matching and histogram-matching) are carried out to render natural colors onto the false-colored image segment by segment. (6) The mapped image is then transformed back to the RGB space. (7) Finally, the mapped image is transformed into HSV (Hue-Saturation-Value) space and the "value" component of the mapped image is replaced with the "fused NV image" (a grayscale image). Note that this fused image replacement is necessary to allow the colorized image to have a proper and consistent contrast.

3.1 Color space transform

In this section, the RGB to *LMS* (long-wave, medium-wave and short-wave) transform is discussed first. Then, an *laβ* space is introduced from which the resulting data representation is compact and *symmetrical*, and provides a higher *decorrelation* than the second order. The reason for the color space transform is to decorrelate three color components (i.e., *l*, *a* and *β*) so that the manipulation (such as statistic matching and histogram matching) on each color component can be performed independently. Inverse transforms (*laβ* space to the ***LMS*** and *LMS* to RGB) are needed to complete the proposed segmentation-based colorization, which are given elsewhere (Zheng & Essock, 2008).

The actual conversion (matrix) from RGB tristimulus to device-independent XYZ tristimulus values depends on the characteristics of the display being used. Fairchild (1998) suggested a "general" device-independent conversion (while non-priori knowledge about the display device) that maps white in the chromaticity diagram to white in the RGB space and vice versa.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.5141 & 0.3239 & 0.1604 \\ 0.2651 & 0.6702 & 0.0641 \\ 0.0241 & 0.1228 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \cdot \tag{11}$$

The XYZ values can be converted to the *LMS* space using the following equation

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \cdot \tag{12}$$

A logarithmic transform is employed here to reduce the data skew that existed in the above color space:

$$\boldsymbol{L} = \log L, \quad \boldsymbol{M} = \log M, \quad \boldsymbol{S} = \log S. \tag{13}$$

Ruderman et al. (1998) presented a color space, named *laβ* (Luminance-Alpha-Beta), which can decorrelate the three axes in the ***LMS*** space:

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.5774 & 0.5774 & 0.5774 \\ 0.4082 & 0.4082 & -0.8165 \\ 1.4142 & -1.4142 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{M} \\ \boldsymbol{S} \end{bmatrix} \cdot \tag{14}$$

The three axes can be considered as an achromatic direction (*l*), a yellow-blue opponent direction (*a*), and a red-green opponent direction (*β*). The *laβ* space has the characteristics of compact, symmetrical and decorrelation, which highly facilitate the subsequent process of color-mapping (see Section 3.4).

3.2 Image segmentation

The nonlinear diffusion procedure has proven to be equivalent to an adaptive smoothing process (Barash & Comaniciu, 2004). The diffusion is applied to the false-colored NV image here to obtain a smooth image, which significantly facilitates the subsequent segmentation process. The clustering process is performed separately on each color component in the *lαβ* color space to form a set of "clusters". The region merging process is used to merge the fragmental clusters into meaningful "segments" (based on a similarity metric defined in 3D *lαβ* color space) that will be used for the color-mapping process.

3.2.1 Adaptive smoothing with nonlinear diffusion

Nonlinear diffusion methods have been proven as powerful methods in the denoising and smoothing of image intensities while retaining and enhancing edges. Barash and Comaniciu (2004) have proven that nonlinear diffusion is equivalent to adaptive smoothing and bilateral filtering is obtained from an extended nonlinear diffusion. Nonlinear diffusion filtering was first introduced by Perona and Malik (1990). Basically, diffusion is a PDE (partial differential equation) method that involves two operators, smoothing and gradient, in 2D image space. The diffusion process smoothes the regions with lower gradients and stops the smoothing at region boundaries with higher gradients. Nonlinear diffusion means the smoothing operation depends on the region gradient distribution. For color image diffusion, three RGB components of a false-colored NV image are filtered separately (one by one). The number of colors in the diffused image will be significantly reduced and will benefit the subsequent image segmentation procedures – clustering and merging.

3.2.2 Image segmentation with clustering and region merging

The diffused false-colored image is transformed into the *lαβ* color space. Each component (l, α or β) of the diffused image is *clustered* in the *lαβ* space by individually analyzing its histogram. Specifically, for each intensity component (image) l, α or β, (i) normalize the intensity onto [0,1]; (ii) bin the normalized intensity to a certain number of levels N_{Bin} and perform the histogram analysis; (iii) with the histogram, locate local extreme values (i.e., peaks and valleys) and form a stepwise mapping function using the peaks and valleys; (iv) complete the clustering utilizing the stepwise mapping function.

The local extremes (peaks or valleys) are easily located by examining the crossover points of the first derivatives of histograms. Furthermore, "peaks" and "valleys" are expected to be interleaved (e.g., valley-peak-valley-…-peak-valley); otherwise, a new valley value can be calculated with the midpoint of two neighboring peaks. In addition, two-end boundaries are considered two special valleys. In summary, all intensities between two valleys in a histogram are squeezed in their peak intensity and the two end points in the histogram are treated as valleys (rather than peaks). If there are n peaks in a histogram, then an n-step mapping function is formed. If there are two or more valley values (including the special valley at the left end) at the left side of the leftmost peak, then use the special (extreme) valley intensity.

Clustering is done by separately analyzing three components (l, α & β) of the false-colored image, which may result in inconsistent clusters in the sense of colors. *Region merging* is necessary to incorporate the fragmental "clusters" into meaningful "segments" in the sense of colors, which will improve the color consistency in a colorized image. If two clusters are similar (i.e., $Q_w(x,y) > T_Q$ (a predefined threshold)), these two clusters will be merged. $Q_w(x,y)$ is a similarity metric (derived from the IQI metric described in Section 2.3.1) between two clusters, x and y, which is defined in the $l\alpha\beta$ color space as follows:

$$Q_w(x,y) = \sum_{k=\{l,\alpha,\beta\}} [w_k \cdot Q_k(x,y)], \tag{15a}$$

where w_k is a given weight for each color component. $Q_k(x,y)$ is formulated below:

$$Q_k(x,y) = \frac{2\overline{x}\overline{y}}{(\overline{x}^2 + \overline{y}^2)} \cdot \frac{2\sigma_x\sigma_y}{(\sigma_x^2 + \sigma_y^2)}, \tag{15b}$$

where $\overline{x}$ and σ_x are the mean and the standard deviation of cluster x in a particular component, respectively. Similar definitions are applied to cluster y. The sizes (i.e., areas) of two clusters (x and y) are usually unequal. Notice that $Q_k(x,y)$ is computed with regard to the diffused false-color image.

3.3 Automatic segment recognition

A *nearest neighbor* (NN) paradigm (Keysers et al., 2002) is demonstrated to classify the segments obtained from the preceding procedure (described in Section 3.2). To use the NN algorithm, a distance measure between two segments is needed. The similarity metric $Q_w(x,y)$ (as defined in Eqs. (15)) between two segments, x and y, is used as the distance measure. Thus, the closer two segments in $l\alpha\beta$ space, the larger their similarity.

Similar to a training process, a look up table (LUT) has to be built under supervision to classify a given segment ($\mathbf{s}_j$) into a known color group (C_i), i.e., $C_i = T(\mathbf{s}_j)$, ($i \leq j$), where $\mathbf{s}_j$ is a feature vector that distinguishingly describes each segment; C_i stands for a known color scheme (e.g., sky, clouds, plants, water, ground, roads, etc.); and T is a classification function (i.e., a trained classifier). We use segment color statistics (e.g., mean and deviation of each channel) as features (of six statistical variables). The statistical features ($\mathbf{s}_j$) are computed using the diffused false-color images and the color mapping process is carried out between a false-color segment and a daylight color scheme. The reason for using the diffused false-color images here is because the diffused images are less sensitive to noise. In a training stage, a set of multispectral NV images are analyzed and segmented such that a sequence of feature vectors, $\{\mathbf{s}_j\}$ can be computed and the LUT (mapping) between $\{\mathbf{s}_j\}$ and $\{C_i\}$ can be manually set up upon the experimental results. In a classifying (testing) stage, all $Q_w(\mathbf{x}_k, \mathbf{s}_j)$ values (for j = 1, 2, 3, …) are calculated, where x_k means the classifying segment and sj represents one of the existing segments from the training stage. Certainly, $\mathbf{x}_k$ is automatically classified into the color group of the largest Q_w (similarity). For example, if $Q_w(\mathbf{x}_1, \mathbf{s}_5)$ is the maximum, then the segment of $\mathbf{x}_1$ will be colorized using the color scheme $T(\mathbf{s}_5)$ that is the color used to render the segment of $\mathbf{s}_5$ in the training stage.

3.4 Color mapping

3.4.1 Statistic matching

A "*statistic matching*" is used to transfer the color characteristics from natural daylight imagery to false color night-vision imagery, which is formulated as:

$$\boldsymbol{I}_C^k = (\boldsymbol{I}_S^k - \boldsymbol{\mu}_S^k) \cdot \frac{\boldsymbol{\sigma}_T^k}{\boldsymbol{\sigma}_S^k} + \boldsymbol{\mu}_T^k, \qquad \text{for } k = \{\, l, \alpha, \beta \,\}, \tag{16}$$

where $\mathbf{I}_C$ is the colored image, $\mathbf{I}_S$ is the source (false-color) image in *lαβ* space; **μ** denotes the mean and **σ** denotes the standard deviation; the subscripts 'S' and 'T' refer to the source and target images, respectively; and the superscript '*k*' is one of the color components: { *l, α, β*}. After this transformation, the pixels comprising the multispectral source image have means and standard deviations that conform to the target daylight color image in *lαβ* space. The color-mapped image is transformed back to the RGB space through the inverse transforms (*lαβ* space to the ***LMS***, exponential transform from ***LMS*** to *LMS*, and *LMS* to RGB, refer to Eqs. (11-14)) (Zheng & Essock, 2008).

3.4.2 Histogram matching

Histogram matching (also referred to as histogram specification) is usually used to enhance an image when histogram equalization fails (Gonzalez & Woods, 2002). Given the shape of the histogram that we want the enhanced image to have, histogram matching can generate a processed image that has the specified histogram. In particular, by specifying the histogram of a target image (with daylight natural colors), a source image (with false colors) resembles the target image in terms of histogram distribution after histogram matching. Similar to statistic matching, histogram matching also serves for color mapping and is performed component-by-component in *lαβ* space. Histogram matching and statistic matching can be applied separately or jointly.

4. Experimental results and discussions

One pair of off-focal images (two clocks) captured at different focal planes (Fig. 2) and five pairs of multispectral images (Figs. 3-6 and Fig. 10) were tested and compared by using the presented multiscale image fusion algorithms and night vision colorization algorithm. Two-band images are image intensified (II) versus infrared (IR) as shown in Figs. 3-5, or visible versus IR image as shown in Fig. 6 and Fig. 10. Note that there was no post-processing imposed on the fused images. The fusion process illustrated here was accepting two input images. In fact, the fusion procedure can accept more than two input images (e.g., three or more images) that will go through the same fusion rules to yield a fused image.

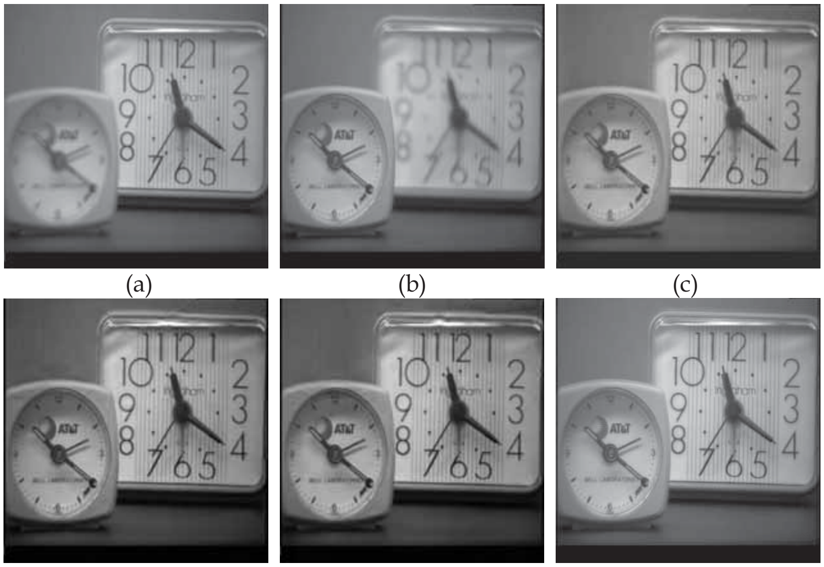


Fig. 2. Image fusion with off-focal image pair (512×512 pixels): (a) and (b) are the input images; (c) Fused image by Laplacian pyramid; (d) Fused image by *a*DWTi-IQI; (e) Fused image by *a*DWTi-rSFe; (f) Orientation-based fusion (16×16) without DC (i.e., suppressing DC). The IQI values of four fused images shown in (c, d, e, f) are 0.8887, 0.9272, 0.9222, 0.*9391*.

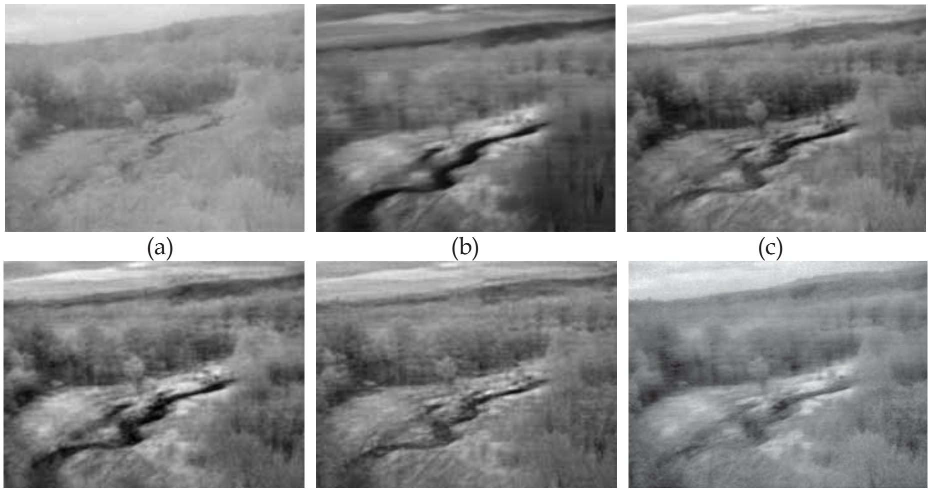
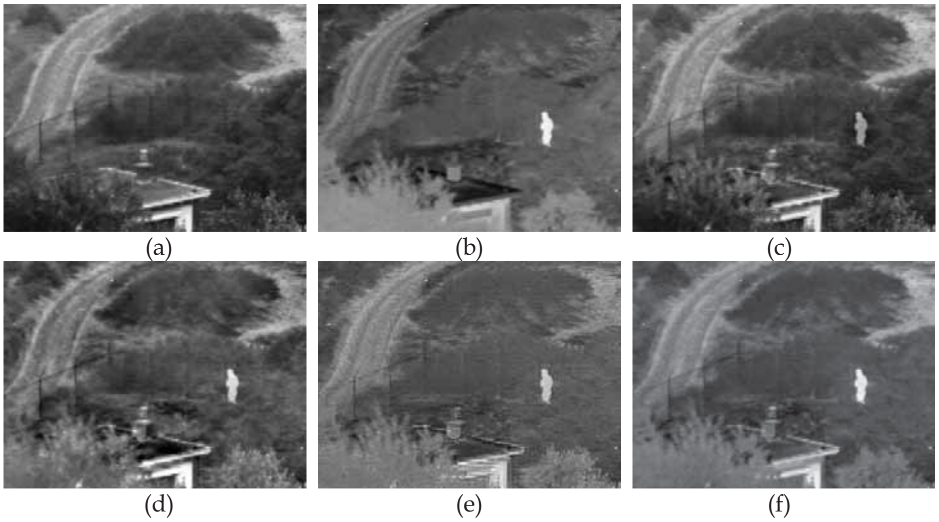


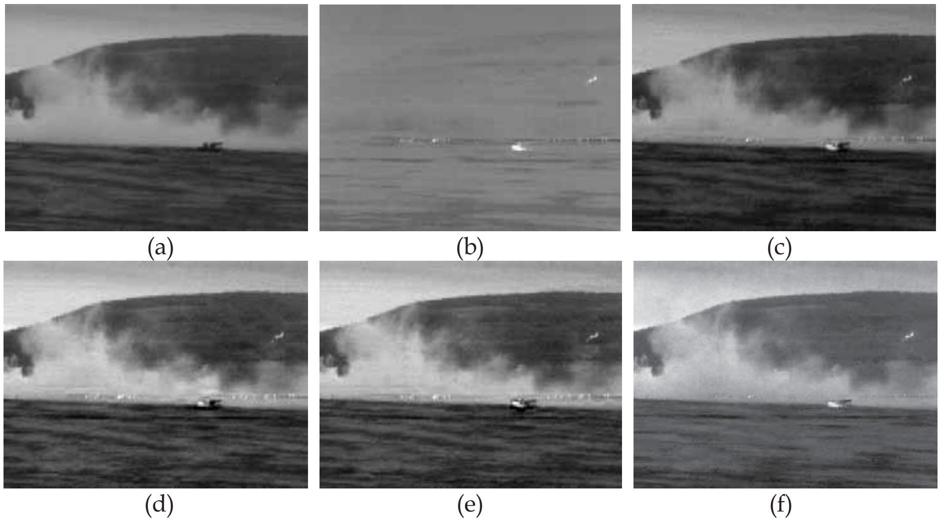
Fig. 3. Image fusion with multispectral image pair #1 (531×401 pixels): (a) and (b) are II and IR images; (c) Fused image by Laplacian pyramid; (d) Fused image by *a*DWTi-IQI; (e) Fused image by *a*DWTi-rSFe; (f) Orientation-based fusion (16×16). The IQI values of four fused images shown in (c, d, e, f) are 0.7680, 0.*7768*, 0.7132, 0.7087.



(a) (b) (c)

(d) (e) (f)

Fig. 4. Image fusion with multispectral image pair #2 (360×270 pixels): (a) and (b) are II and IR images; (c) Fused image by Laplacian pyramid; (d) Fused image by *a*DWTi-IQI; (e) Fused image by *a*DWTi-rSFe; (f) Orientation-based fusion (16×16). The IQI values of four fused images shown in (c, d, e, f) are 0.7335, 0.7089, 0.6107, 0.*7421*.



(a) (b) (c)

(d) (e) (f)

Fig. 5. Image fusion with multispectral image pair #3 (360×270 pixels): (a) and (b) are II and IR images; (c) Fused image by Laplacian pyramid; (d) Fused image by *a*DWTi-IQI; (e) Fused image by *a*DWTi-rSFe; (f) Orientation-based fusion (16×16). The IQI values of four fused images shown in (c, d, e, f) are 0.8160, *0.8347*, 0.8309, 0.8249.

For each case as demonstrated in Figs. 2-6, the IQI values of four fusions are shown in the figure captions. Actually, there were no iterations in Laplacian pyramid fusion and orientation based fusion. For the Laplacian pyramid algorithm, a pair of fixed parameters,

(L_d, L_w) = (4, 4) as typically used in literature, were used in all pyramid fusions (shown in Figs. 2-6). In general, the *a*DWTi-IQI algorithm converges at larger numbers of N_i and L_w but a smaller number of L_d; whereas the *a*DWTi-rSFe algorithm converges at a larger number of L_d but smaller numbers of N_i and L_w. Furthermore, the *a*DWTi-IQI algorithm produces a smooth image, which is especially suitable for noisy images such as multispectral NV images; whereas the *a*DWTi-rSFe algorithm yields a sharpened image, which is ideal for well exposed daylight pictures (like the two-clock image pair). On the other hand, the orientation-based fusion using Gabor wavelet transform is good for the fusion between contrast-unlike images such as visible versus IR (thermal) images.

The IQI values (the higher the better) of four fusions, as shown in the figure captions of Figs. 2-6, are used for quantitative evaluations. The IQI results showed that, the orientation-based fusion is the best in Figs. 2, 4, & 6, while the *a*DWTi-IQI fusion is the best in Figs. 3 & 5. Visual perceptions provide the same rank of fused images as the quantitative evaluations.

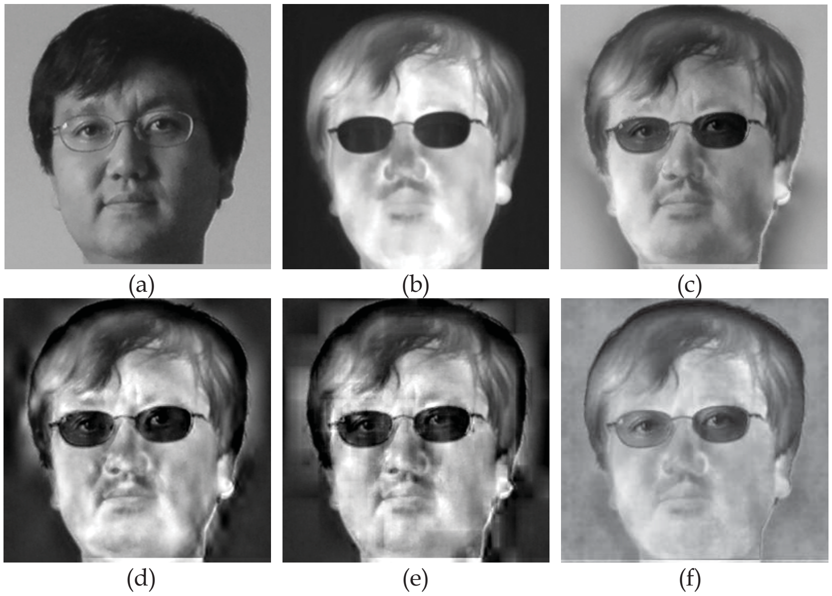

(a) (b) (c)
(d) (e) (f)

Fig. 6. Image fusion with visible and IR images (taken at daytime; 259×258 pixels). (a) Visible image; (b) IR image; (c) Fused image by Laplacian pyramid; (d) Fused image by *a*DWTi-IQI; (e) Fused image by *a*DWTi-rSFe; (f) Orientation-based fusion (16×16). The IQI values of four fused images shown in (c, d, e, f) are 0.6088, 0.6267, 0.6065, *0.6635*.

As shown in Fig. 6, Laplacian fusion (Fig. 6c) is pretty good but the eyes behind the glasses are not as clear as shown in the orientation fusion (Fig. 6f). Notice that eyes are the most important facial features in face recognition systems and applications. The iterative fusions of *a*DWTi-IQI and *a*DWTi-rSFe show overshot effect especially around the head boundary. The IQI values reveal the same rank of different fusions. The 16×16 orientation fusion (16 bands by 16 orientations, Fig. 6f) presents more details and better contrast than other multiscale fusions (Figs. 6c-e). In an $M×N$ orientation-based fusion, a larger M (number of bands) is usually beneficial to the detailed images like Fig. 6.

The tree pairs of multispectral images were completely analyzed by the presented night vision colorization algorithm; and the results using *local coloring* algorithm are illustrated in Figs. 7-9. The original input images and the fused images used in the coloring process are shown in Figs. 3-5a, Figs. 3-5b and Figs. 3-5d, respectively. The smooth images (Figs. 3-5d) fused by the *a*DWTi-IQI algorithm were used in night vision colorization because they show better contrast and less sensitive to noises. The false colored images are shown in Figs. 7-9a, which were obtained by assigning image intensified (II) images to blue channels, infrared (IR) images to red channels, and providing averaged II and IR images to green channels. The rationale of forming a false-color image is to assign a long-wavelength NV image to the red channel and to assign a short-wavelength NV to the blue channel. The number of false colors were reduced with the nonlinear diffusion algorithm with AOS (additive operator splitting for fast computation) implementation that facilitated the subsequent segmentation. The segmentation was done in *lαβ* space through clustering and merging operations (see Figs. 7-9b). The parameter values used in clustering and merging are N_{Bin} = [24 24 24], w_{k} = [0.25 0.35 0.40] and T_{Q} = 0.90. To emphasize two chromatic channels (due to more distinguishable among segments) in *lαβ* space, relatively larger weights were assigned in w_{k}. With the segment map, the histogram-matching and statistic-matching could be performed segment by segment (i.e., locally) in *lαβ* space. The source region segments were automatically recognized and associated with proper target color schemes (after the training process is done). The locally colored images (segment-by-segment) are shown in Figs. 7-9c. From a visual examination, the colored images (Figs. 7-9c) appear very natural, realistic, and colorful. The comparable colorization results by using *global coloring* algorithm are presented in Reference (Zheng & Essock, 2008). This segmentation-based *local coloring* process is fully automatic and well adaptive to different types of multisensor images. The input images are not necessary to be multispectral NV images although the illustrations given here use NV images.


(a) (b) (c)

Fig. 7. Night vision colorization with multispectral image pair #1 (531×401 pixels): Original multispectral images are shown in Figs. 3a-b, and the fused image used in colorization is shown in Fig. 3d. (a) is the false-colored image using Figs. 3a-b; (b) is the segmented image from image (a), where 16 segments were merged from 36 clusters; (c) is the colored image, where six auto-classified color schemes (sky, clouds, plants, water, ground and others) were mapped by jointly using histogram-matching and statistic-matching.


(a) (b) (c)

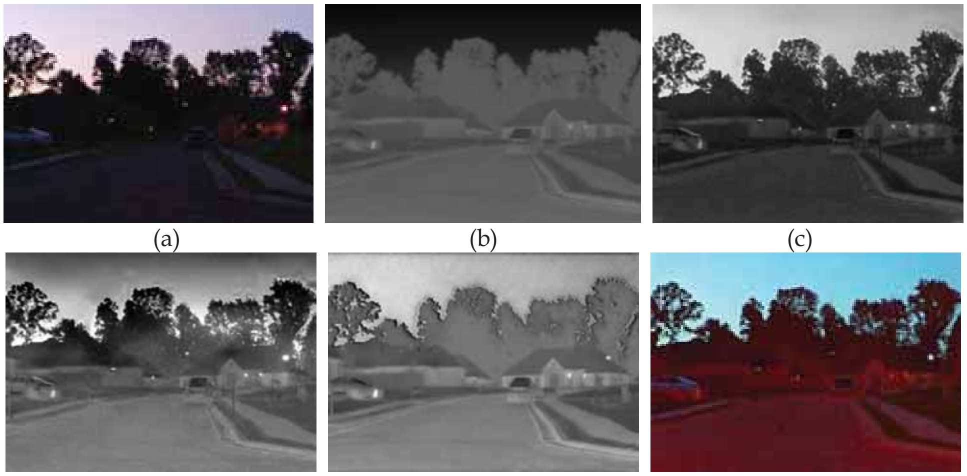
Fig. 8. Night vision colorization with multispectral image pair #2 (360×270 pixels): Refer to Figs. 4a,b,d for the original multispectral images and the fused image used in colorization. (a) is the false-colored image using Figs. 4a-b; (b) is the segmented image of 12 segments merged from 21 clusters; (c) is the colored image with five auto-classified color schemes (plants, roads, ground, building and others).


(a) (b) (c)

Fig. 9. Night vision colorization with multispectral image pair #3 (360×270 pixels): Refer to Figs. 5a,b,d for the original multispectral images and the fused image used in colorization. (a) is the false-colored image using Figs. 5a-b; (b) is the segmented image of 14 segments merged from 28 clusters; (c) is the colored image with three auto-classified color schemes (plants, smoke and others).

A different *color fusion* is illustrated in Fig. 10f by replacing the red channel image in Fig. 10a with the orientation fused images in Fig. 10e (IQI = 0.7849). The orientation-based fusion (Fig. 10e) was formed by combining the red channel image of Fig. 10a (visible band) and a IR (thermal) image (Fig. 10b), which shows a better result than Figs. 10c-d. The colors in Fig. 10f is not as natural as daylight colors but useful for human perception especially for those poorly illuminated images. For example, Fig. 10f shows a better contrast and more details than Fig. 10a and Figs. 10c-e. Note that non-uniform band weights (W_b = [0.0250 0.0250 0.0500 0.0500 0.0875 0.0875 0.0875 0.0875 0.0875 0.0875 0.0875 0.0875 0.0500 0.0500 0.0250 0.0250]) were applied to the noisy input images in order to emphasize the contents at medium frequencies meanwhile suppress the noise at high-frequencies.

The night vision colorization process demonstrated here took two-band multispectral NV images as inputs. Actually, the local-coloring procedure can accept two or three input images. If there are more than three bands of images available, we may choose the low-light intensified (visual band) image and two bands of IR images. As far how to choose two bands of IR images, we may use the image fusion algorithm as a screening process. The two selected IR images for colorization should be the two images that can produce the most (maximum) informative fused image among all possible fusions. For example, given three IR images, IR_1, IR_2, IR_3, the two chosen images for colorization, I_{C1}, I_{C2}, should satisfy the following equation: $Fus(I_{C1}, I_{C2}) = max\{Fus(IR_1, IR_2), Fus(IR_1, IR_3), Fus(IR_2, IR_3)\}$, where *Fus* stands for the fusion process and *max* means selecting the fusion of maximum information.



(a) (b) (c)

(d) (e) (f)

Fig. 10. Color fusion with visible color image and IR image (taken outdoors at dusk; 400×282 pixels). (a) Color image; (b) IR image; (c) Fused image by Laplacian pyramid (IQI = 0.7666); (d) Fused image by *a*DWTi-IQI (IQI = 0.7809); (e) Orientation-based fusion (16×16; IQI = *0.7849*) between of the red channel of (a) and LWIR image; (f) Color fusion by replacing the red channel of Image (a) with Image (e).

5. Conclusions

The multispectral image fusion and night vision colorization approaches presented in this chapter can be performed automatically and adaptively regardless of the image contents. Experimental results with multispectral imagery showed that the fused image is informative and clear, and the colorized image appears realistic and natural. We anticipate that the presented fusion and colorization approaches for multispectral imagery will help improve target recognition and visual analysis, especially for nighttime operations.

Specifically, the proposed approaches can produce two versions of synthesized imagery, a grayscale image and a color image. The image fusion procedure is based on multiscale analysis, and the fused image is suitable to machine analysis (e.g., target recognition). The night vision colorization procedure is based on image segmentation, pattern recognition, and color mapping. The colorized image is good for visual analysis (e.g., pilot navigation). The synthesized multispectral imagery with proposed approaches will eventually lead to improved performance of remote sensing, nighttime navigation, and situational awareness.

6. Acknowledgements

This research is supported by the U. S. Army Research Office under grant number W911NF-08-1-0404.

7. References

Ajazzi, B.; Alparone, L.; Baronti, S.; & Carla, R.; (1998). Assessment of pyramid-based multisensor image data fusion, in *Proc. SPIE 3500*, 237–248.

Barash, D. & Comaniciu, D. (2004). A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift, *Image Vision Computing* **22**(1), 73-81.

Burt, P. J. & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code, *IEEE Trans. Commun.* Com-31 (4), 532–540.

Burt, P. J. & Adelson, E. H. (1985). Merging images through pattern decomposition, *Proc. SPIE* 575, 173–182.

Eskicioglu, A. M. & Fisher, P. S. (1995). Image quality measure and their performance, *IEEE Trans. Commun.* **43**(12), 2959–2965.

Essock, E. A.; McCarley, J. S.; Sinai, M. J. & DeFord, J. K. (2001). Human perception of sensor-fused imagery, in *Interpreting Remote Sensing Imagery: Human Factors*, R. R. Hoffman and A. B. Markman, Eds., Lewis Publishers, Boca Raton, Florida.

Essock, E. A.; Sinai, M. J. & et al. (1999). Perceptual ability with real-world nighttime scenes: imageintensified, infrared, and fused-color imagery, *Hum. Factors* **41**(3), 438–452.

Fairchild, M. D. (1998). *Color Appearance Models*, Addison Wesley Longman Inc., ISBN: 0-201-63464-3, Reading, MA.

Gonzalez, R. C. & Woods, R. E. (2002). *Digital Image Processing* (Second Edition), Prentice Hall, ISBN: 0201180758, Upper Saddle River, NJ.

Hollingsworth, K. P.; Bowyer, K. W.; Flynn, P. J. (2009). The Best Bits in an Iris Code, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 964-973.

Jahard, F.; Fish, D. A.; Rio, A. A. & Thompson C. P. (1997). Far/near infrared adapted pyramid-based fusion for automotive night vision, in *IEEE Proc. 6th Int. Conf. on Image Processing and its Applications* (IPA97), pp. 886–890.

Jones J. P. & Palmer, L. A. (1987). The two-dimensional spectral structure of simple receptive fields in cat striate cortex, Journal of Neurophysiology, vol.58 (6), pp. 1187–1211.

Keys, L. D.; Schmidt, N. J.; & Phillips, B. E. (1990). A prototype example of sensor fusion used for a siting analysis, in Technical Papers 1990, ACSM-ASPRS Annual Conf. *Image Processing and Remote Sensing* **4**, pp. 238–249.

Keysers, D.; Paredes, R.; Ney, H. & Vidal, E. (2002). Combination of tangent vectors and local representations for handwritten digit recognition, Int. Workshop on Statistical Pattern Recognition, Lecture Notes in Computer Science, Vol. 2396, pp. 538-547, Windsor, Ontario, Canada.

Li, S.; Kwok, J. T. & Wang, Y. (2001). Combination of images with diverse focuses using the spatial frequency, *Information Fusion* **2**(3), 169–176.

Nunez, J.; Otazu, X.; & et al. (1999). Image fusion with additive multiresolution wavelet decomposition; applications to spot1landsat images, *J. Opt. Soc. Am. A* **16**, 467–474.

Perona, P. & Malik, J. (1990). Scale space and edge detection using anisotropic diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**, 629–639.

Piella, G. & Heijmans, H. (2003). A new quality metric for image fusion, in *Proc. 2003 Int. Conf. on Image Processing*, Barcelona, Spain.

Pohl C. & Genderen J. L. V. (1998). Review article: multisensor image fusion in remote sensing: concepts, methods and applications, *Int. J. Remote Sens.* **19**(5), 823–854.

Pu T. & Ni, G. (2000). Contrast-based image fusion using the discrete wavelet transform, *Opt. Eng.* **39**(8), 2075–2082.

Rogers, R. H. & Wood, L (1990). The history and status of merging multiple sensor data: an overview, in Technical Papers 1990, ACSMASPRS Annual Conf. *Image Processing and Remote Sensing* **4**, pp. 352–360.

Ruderman, D. L.; Cronin, T. W. & Chiao, C. C. (1998). Statistics of cone responses to natural images: implications for visual coding, *Journal of the Optical Society of America A* **15** (8), 2036–2045.

Toet, A. (2003). Natural colour mapping for multiband nightvision imagery, *Information Fusion* **4**, 155-166.

Toet, A. & IJspeert, J. K. (2001). Perceptual evaluation of different image fusion schemes, in: I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition X*, The International Society for Optical Engineering, Bellingham, WA, pp.436–441.

Toet, A.; IJspeert, J.K.; Waxman, A. M. & Aguilar, M. (1997). Fusion of visible and thermal imagery improves situational awareness, in: J.G. Verly (Ed.), *Enhanced and Synthetic Vision 1997*, International Society for Optical Engineering, Bellingham, WA, pp.177–188.

Varga, J. T. (1999). Evaluation of operator performance using true color and artificial color in natural scene perception (Report ADA363036), Naval Postgraduate School, Monterey, CA.

Wang, Z. & Bovik, A. C. (2002). A universal image quality index, *IEEE Signal Processing Letters* **9**(3), 81–84.

Waxman, A.M.; Gove, A. N. & et al. (1996). Progress on color night vision: visible/IR fusion, perception and search, and low-light CCD imaging, *Proc. SPIE Vol. 2736*, pp. 96-107, Enhanced and Synthetic Vision 1996, Jacques G. Verly; Ed.

Zheng, Y. & Agyepong, K. (2007). Mass Detection with Digitized Screening Mammograms by Using Gabor Features, Proceedings of the SPIE, Vol. 6514, pp. 651402-1-12.

Zheng, Y. & Essock, E. A. (2008). A local-coloring method for night-vision colorization utilizing image analysis and image fusion, *Information Fusion* **9**, 186-199.

Zheng, Y.; Essock, E. A. & Hansen, B. C. (2005). An advanced DWT fusion algorithm and its optimization by using the metric of image quality index, *Optical Engineering* **44** (3), 037003-1-12.

Zheng, Y.; Essock, E. A. & Hansen, B. C. (2004). An advanced image fusion algorithm based on wavelet transform—incorporation with PCA and morphological processing, *Proc. SPIE 5298*, 177–187.

Zheng, Y.; Essock, E. A.; Hansen, B. C. & Haun, A. M. (2007). A new metric based on extended spatial frequency and its application to DWT based fusion algorithms, *Information Fusion* **8**(2), 177-192.

Zheng, Y.; Hansen, B. C. & Haun, A. M. & Essock, E. A. (2005). Coloring Night-vision Imagery with Statistical Properties of Natural Colors by Using Image Segmentation and Histogram Matching, *Proceedings of the SPIE*, Vol. 5667, pp. 107-117.