1. Introduction

Multispectral images usually present complimentary information such as visual-band imagery and infrared imagery (near infrared or long wave infrared). There are strong evidences that the fused multispectral imagery (in gray scales) increases the reliability of interpretation (Rogers & Wood, 1990; Essock et al., 2001) and thus good for machine analysis (computer vision); whereas the colorized multispectral imagery improves observer performance and reaction times (Toet et al. 1997; Varga, 1999; Waxman et al., 1996) and thus good for visual analysis (human vision).

Imagine a nighttime navigation task that may be executed by an aircraft equipped with a multispectral imaging system. Analyzing the synthesized (fused or colorized) multisensory image will be more informative 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, which may be displayed either on several split panels on a big screen or on several small screens. The focus of this chapter is how to synthesize a color presentation of multispectral images in order to enhance night vision. It is anticipated that the successful applications of night vision colorization techniques will lead to improved performance of remote sensing, nighttime navigation, target detection, and situational awareness. This colorization approaches mentioned here involve two main techniques, image fusion and colorization, which are briefly 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). The discussions of image fusion here are limited to *pixel-level fusion*.

Two commonly used 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 (aDWT) method (Zheng et al., 2004) has been proposed, which incorporates principal component analysis (PCA) and morphological processing into a regular DWT fusion algorithm. The aDWT method can produce a better fused image in comparison with pyramid methods and regular DWT methods. Image fusion is a necessary step for the following color fusion and colorization methods.

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 (i.e., “segmentation-based”), and also provide automatic association between the source and target images (i.e., avoiding the manual scene-matching in global coloring). This local coloring method is also referred to as “segmentation-based” colorization in contrast with “channel-based” color fusion introduced later.

In this chapter, we will discuss and explore how to enhance human night vision by presenting a color image with a set of multispectral images. Certainly, a color presentation of multispectral night vision images can provide a better visual result for human users. We would prefer the color images resembling natural daylight pictures that we are used to; meanwhile the coloring process shall be efficient enough ideally for real time applications. A segmentation-based colorization procedure is first reviewed, and a channel-based color fusion is then introduced. The remainder of this chapter is organized as follows. The multispectral image preprocessing, registration and fusion are described in Section 2. Next, the segmentation-based colorization method is completely discussed in Section 3. Then, a new channel-based color fusion method is introduced in Section 4. The experiments and discussions are given in Section 5. Conclusions are finally drawn in Section 6.
2. Multispectral image preprocessing

The multispectral images that we acquired include visible (RGB color) images, image intensified (II, enhanced visible) images, near infrared (NIR; spectral range: 0.9–1.7 μm) images, and long-wave infrared (LWIR; spectral range: 7.5–13 μm) images. Before performing multispectral colorization, image preprocessing, image registration, and image fusion are required.

2.1 Standard preprocessing

Standard image preprocessing such as denoising, normalization and enhancement can favorite the following processes, i.e., image registration, fusion, and colorization. The noise in digital images may be caused by imperfection of imaging sensors, scene contents in FOV (field of view, extremely cold or hot objects for infrared imaging), environment (atmosphere) disturbance, or poor illumination (for visible band imaging). Noise can be reduced according to the nature of the noise sources that depends on a particular application. For example, pepper-and-salt noise can be removed by a median filter; periodic noise may be reduced by a designed frequency filter in Fourier transformed domain; and a random noise can be suppressed by a Gaussian filter or a nonlinear diffusion filter.

Night-vision images (NIR and LWIR) were acquired under different background and conditions, which may cause images to have different background (brightness) and contrast (dynamic range). We employed a general image normalization (also called contrast stretching) to standardize all multispectral images.

\[ I_N = (I_0 - I_{\text{Min}}) \frac{L_{\text{Max}} - L_{\text{Min}}}{I_{\text{Max}} - I_{\text{Min}}} + L_{\text{Min}} \]  

(1)

where \( I_N \) is the normalized image, \( I_0 \) is the original image; \( I_{\text{Min}} \) and \( I_{\text{Max}} \) are the maximum and minimum pixel values in \( I_0 \), respectively; \( L_{\text{Min}} \) and \( L_{\text{Max}} \) are the expected maximum and minimum pixel values in \( I_N \), which normally equal 0 and 1, respectively. After image normalization, \( I_N \in [0, 1] \).

The image contrasts of near infrared (NIR) images are significantly affected by illumination conditions. Nonlinear enhancement like histogram equalization or histogram matching usually increases noises while enhancing a NIR image. A linear enhancement such as piecewise contrast stretching is preferred. Eq. (1) is still applicable but just applied within each piece of intensity interval. For example, given \([I_{\text{Min}}, I_{\text{Max}}] = [0, 0.8]\), and \([L_{\text{Min}}, L_{\text{Max}}] = [0, 1.0]\), after piecewise contrast stretching, the pixels within \([0, 0.8]\) will be linearly scaled to \([0, 1.0]\), while those pixels originally within \([0.8, 1.0]\) are unchanged. To simply writing, this transform can be notated as \([0,1.0]_S \) thereafter.

2.2 Image registration

Image registration is a required preprocess by image fusion and image colorization. In general, image registration aligns multiple images by performing affine transformations that allows translation, rotation, and scaling. Similarity metrics are used to decide the optimized transformation parameters. Normalized mutual information (NMI) turns out to be the robust metric for noisy and multi-modality image registration (Hill & Batchelor, 2001). The computation complexity increases with the number of degrees of freedom. For 2D image
registration, Fourier-Mellin transform (FMT, Chen et al., 1994) is much faster than NMI-based registration, but FMT is sensitive to noise. For multispectral night-vision image registration, we utilize two registration algorithms, i.e., use the FMT method for translation registration, and then use the NMI-based method for scaling and rotation registration.

We used the FMT method only accounting for translation alignment although it can be alternated for scaling and rotation (but not reliable). The image alignment by scaling and rotation is accomplished with affine transforms using NMI metric. The image transforming parameters can be estimated by maximizing the NMI value. Calculation of NMI and interpolation of transforming (e.g., fractional scaling) are quite time consuming. However, the searching spaces of parameters (for scaling and rotation) are small because two cameras are sitting on the same fixture by turns and aiming at the same target. This expedites the registration process on the other hand.

Different FOV of multispectral images is another challenge for image registration. For example, FLIR SC620 camera (used in our experiments) is a two-band imaging device with a LWIR camera (640×480 pixels; FOV: 24°) and a built-in visible camera (2048×1536 pixels; FOV: 32°). Before registration with LWIR image cropping the visible image is desired. To find the matched block (region) of LWIR on the visible image, (i) scan the visible image block by block with step movement 5~10 pixels (Left to Right, Top to Bottom), where the block image is of size (960×720, estimated according to view angles); (ii) compute the NMI between the scanning block (on the visible image) and LWIR; (iii) select the scanning block (region) with maximal NMI as the matched block for the following registration. The framework of general image registration was documented elsewhere (Brown, 1992), and the details of our proposed coarse-to-fine registration method will be discussed in a separate paper.

2.3 Image fusion

Image fusion is a necessary step for the color fusion discussed in this chapter. Image fusion serves to combine multiple-source imagery using advanced image processing techniques. Laplacian pyramid and DWT-based fusion methods are briefly reviewed, while the details of image fusion were documented elsewhere (Zheng et al., 2005).

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.

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 (aDWT) method (Zheng et al., 2004), we apply PCA (principle component analysis) 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. (2) below:

\[ C_f = \left( a_1 \cdot C_A + a_2 \cdot C_B \right) / \left( a_1 + a_2 \right) \]

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. (2) 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.

3. Segmentation-based colorization

In segmentation-based colorization (i.e., local coloring) method, multispectral night vision imagery is rendered segment-by-segment with the statistical color properties of natural scenes by using the color mapping technique. Eventually, the colorized images resemble daylight pictures. The main steps of segmentation-based colorization 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, the techniques of nonlinear diffusion, clustering, and region merging. A set of “clusters” are formed by analyzing the histograms of the three components of the diffused image in \( l\alpha\beta \) color space. Those clusters are merged to “segments” if their similarity values in \( l\alpha\beta \) space are greater than a preset threshold. (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 subsection, the RGB to LMS (long-wave, medium-wave and short-wave) transform is discussed first. Then, an \( l\alpha\beta \) 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 \( \beta \)) 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.

\[
\begin{pmatrix}
X \\
Y \\
Z \\
\end{pmatrix} = \begin{pmatrix}
0.5141 & 0.3239 & 0.1604 \\
0.2651 & 0.6702 & 0.0641 \\
0.0241 & 0.1228 & 0.8444 \\
\end{pmatrix} \begin{pmatrix}
R \\
G \\
B \\
\end{pmatrix}
\]  \hspace{1cm} (3)

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

\[
\begin{pmatrix}
L \\
M \\
S \\
\end{pmatrix} = \begin{pmatrix}
0.3897 & 0.6890 & -0.0787 \\
-0.2298 & 1.1834 & 0.0464 \\
0.0000 & 0.0000 & 1.0000 \\
\end{pmatrix} \begin{pmatrix}
X \\
Y \\
Z \\
\end{pmatrix}
\]  \hspace{1cm} (4)

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

\[
L = \log L, \quad M = \log M, \quad S = \log S. \hspace{1cm} (5)
\]

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

\[
\begin{pmatrix}
l \\
\alpha \\
\beta \\
\end{pmatrix} = \begin{pmatrix}
0.5774 & 0.5774 & 0.5774 \\
0.4082 & 0.4082 & -0.8165 \\
1.4142 & -1.4142 & 0 \\
\end{pmatrix} \begin{pmatrix}
L \\
M \\
S \\
\end{pmatrix}
\]  \hspace{1cm} (6)

The three axes can be considered as an achromatic direction (l), a yellow-blue opponent direction (α), 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 laβ 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 laβ 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 smooths 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 laβ color space. Each component (l, a or β) of the diffused image is clustered in the laβ space by individually analyzing its histogram. Specifically, for each intensity component (image) l, a or β, (i) normalize the intensity onto [0,1]; (ii) bin the normalized intensity to a certain number of levels NBin 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, a & 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., Qw(x,y) > TQ (a predefined threshold)), these two clusters will be merged. Qw(x,y) is a similarity metric between two clusters, x and y, which is defined in the laβ color space as follows:

\[ Q_w(x,y) = \sum_{k=[l,a,β]} [w_k \cdot Q_k(x,y)] \]  

(7a)

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

\[ Q_k(x,y) = \frac{2\pi y}{(\bar{x}^2 + \bar{y}^2)} \cdot \frac{2\sigma_x \sigma_y}{(\sigma_x^2 + \sigma_y^2)} \]  

(7b)

where \( \bar{x} \) and \( \sigma_x \) 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. (7)) 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 ($s_i$) into a known color group ($C_i$), i.e., $C_i = T(s_i)$, (i ≤ j), where $s_i$ is a feature vector that distinguishes 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 ($s_i$) 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, $\{s_i\}$ can be computed and the LUT (mapping) between $\{s_i\}$ and $\{C_i\}$ can be manually set up upon the experimental results. In a classifying (testing) stage, all $Q_w(x_k,s_j)$ values (for $j = 1, 2, 3, \ldots$) are calculated, where $x_k$ means the current classified segment and $s_j$ represents one of the existing segments from the training stage. Certainly, $x_k$ is automatically classified into the color group of the largest $Q_w$ (similarity). For example, if $Q_w(x_1,s_5)$ is the maximum, then the segment of $x_1$ will be colorized using the color scheme $T(s_5)$ that is the color used to render the segment of $s_5$ 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:

$$I^k_c = (I^k_s - \mu^k_s) \frac{\sigma^k_s}{\sigma^k_T} + \mu^k_T, \text{ for } k = \{l, a, \beta\},$$

(8)

where $I_c$ is the colored image, $I_s$ is the source (false-color) image in $l\alpha\beta$ space; $\mu$ denotes the mean and $\sigma$ 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: $\{l, a, \beta\}$. After this transformation, the pixels comprising the multispectral source image have means and standard deviations that conform to the target daylight color image in $l\alpha\beta$ space. The color-mapped image is transformed back to the RGB space through the inverse transforms ($l\alpha\beta$ space to the LMS, exponential transform from LMS to LMS, and LMS to RGB, refer to Eqs. (3-6)) (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 laβ space. Histogram matching and statistic matching can be
applied separately or jointly.

4. Channel-based color fusion

The segmentation-based colorization described in Section 3 can usually produce colorized
night-vision images closely resembling natural daylight pictures. However, this segmentation-
based coloring procedure involves many processes and heavy computations such as histogram
analysis, color space transform, image segmentation, and pattern classification. It will be a
grand challenge for real time applications. Therefore, we propose a fast color fusion method,
termed as channel-based color fusion, which is efficient enough ideally for real time applications.
Notice that the term of “color fusion” means combing multispectral images into a color-
version image with the purpose of resembling natural scenes. Relative to “night vision
colorization”, color fusion trades the realism of colors with speed. On the other hand, false
coloring techniques have no intention of resembling natural color scenery.
The general framework of channel-based color fusion is as follows, (i) prepare for color
fusion, preprocessing (denoising, normalization and enhancement) and image registration;
(ii) form a color fusion image by properly assigning multispectral images to red, green, and
blue channels; (iii) then fuse multispectral images (gray fusion) using DWT algorithm (see
Section 2.3); and (iv) replace the value component of color fusion in HSV color space with the
gray-fusion image, and finally transform back to RGB space.
In night vision imaging, there may be two or several bands of images available, for example,
visible (RGB), image intensified (II), near infrared (NIR), medium wave infrared (MWIR),
long wave infrared (LWIR, also called thermal). The discussions of following subsections
focus on how to form a channel-wise color fusion with the available multispectral images.

4.1 Color fusion with two-band images

Upon the available images and common applications, we will discuss two-band color fusion
of (II ⊕ LWIR), (NIR ⊕ LWIR), (RGB ⊕ LWIR), and (RGB ⊕ NIR), although other
combinations of two bands may be possible in some applications. The symbol ‘⊕ ’ denotes
the fusion of multiband images.

4.1.1 Color fusion of (II ⊕ LWIR)

Suppose a color fusion image (F_C) consists of three color planes, F_R, F_G, F_B, the color fusion of
II and LWIR images are formed by using the following expressions,

\[ F_R = S_{[0,1]}^{[0.0,0.7]} (I_{LWIR}) \]  
\[ F_G = S_{[0,1]}^{[0.2,1]} (I_{II}) \]  
\[ F_B = S_{[0,1]}^{[0.1,0.75]} ([1.0-I_{LWIR}] \cdot I_{II}) \]  
\[ V_F = Fus(I_{II}, I_{LWIR}) \]
where \( S_{\[0.2,1\]}^{[0.1,1]} \) denotes piecewise contrast stretching defined in Eq. (1) and \( I_{Gmax} = min([\mu_II + 3\sigma_II],0.8) \), \( \mu \) and \( \sigma \) are the mean and standard deviation of an II image; \([1.0- I_{LWIR}]\) is to invert LWIR image; symbol ‘•’ means element-by-element multiplication; \( V_F \) is the value component of \( F_C \) in HSV space, \( Fus() \) means image fusion operation using a DWT algorithm. Although the limits given in contrast stretching are obtained empirically according to the night vision images that we had, it is viable to formulate the expressions and automate the fusion based upon a set of conditions (imaging devices, imaging time, and application location). Notice the transform parameters in Eqs. (9) were applied to all color fusions in our experiments.

### 4.1.2 Color fusion of (NIR ⊕ LWIR)

A color fusion of NIR and LWIR is formulated by,

\[
F_R = S_{\[0.2,0.9\]}^{[0.1,1]} (I_{LWIR}), \quad (10a)
\]

\[
F_C = S_{\[0.2,1\]}^{[0.1,1]} (I_{NIR}), \quad (10b)
\]

\[
F_B = S_{\[0.1,0.7\]}^{[0.0,1]} ([1.0- I_{LWIR}] \cdot I_{NIR}); \quad (10c)
\]

\[
V_F = Fus(I_{NIR}, I_{LWIR}); \quad (10d)
\]

where \( I_{Gmax} = min([\mu_{NIR} + 2\sigma_{NIR}],0.8) \), \( min() \) is an operation to get the minimal number. Other notes are the same as that in Eqs. (9).

### 4.1.3 Color fusion of (RGB ⊕ LWIR)

Two-band color fusion of RGB and LWIR is described as follows,

\[
F_R = S_{\[0.1,1\]}^{[0.1,0]} (max[I_{LWIR}, I_{Red}]), \quad (11a)
\]

\[
F_C = I_{Green}; \quad (11b)
\]

\[
F_B = I_{Blue}; \quad (11c)
\]

where \( I_{Red}, I_{Green} \) and \( I_{Blue} \) are the three channel images of a RGB image; \( I_{Rmax} = min([\mu_{Red} + 8\sigma_{Red}],0.6) \), \( min() \) is an operation to get the minimal number; \( max[I_{LWIR}, I_{Red}] \) is to take the maximal pixel values between \( I_{LWIR} \) and \( I_{Red} \). In fact, this color fusion only modifies the red channel in a RGB image, where the piecewise contrast stretching in Eq. (11a) is to keep a good color balance and contrast. No image fusion is used because the RGB images captured at night time are usually very noisy. Of course, no HSV transform is performed.

### 4.1.4 Color fusion of (RGB ⊕ NIR)

The color fusion of RGB and NIR is defined as,

\[
F_R = I_{Red}; \quad (12a)
\]
\[ F_G = S_{[0,1],G_{\text{max}}} ([\text{max}(I_{\text{NIR}}, I_{\text{Green}})], \quad (12b) \]

\[ F_B = I_{\text{Blue}}; \quad (12c) \]

where \( I_{G_{\text{max}}} = \text{min}(\mu_{\text{Green}} + 6\sigma_{\text{Green}}, 0.6) \). Other notes are the same as that in Eqs. (11). This color fusion actually modifies the green channel in a RGB image. No image fusion and no HSV transform are performed. The color fusion of (RGB \( \oplus \) NIR) is not used as often as the fusion of (RGB \( \oplus \) LWIR).

4.2 Color fusion with three-band images

Due to the available image databases, we only discuss one application of three-band color fusions, (RGB \( \oplus \) NIR \( \oplus \) LWIR). A color fusion of RGB, NIR and LWIR can be described as,

\[ F_R = S_{[0,1],0.6}(I_{\text{LWIR}}), \quad (13a) \]

\[ F_G = S_{[0,1],0.9}(I_{\text{NIR}}), \quad (13b) \]

\[ F_B = S_{[0,1],\text{B}_{\text{max}}}(I_{\text{Blue}}); \quad (13c) \]

\[ V_F = F_{\text{HIS}}(I_{\text{NIR}}, I_{\text{LWIR}}); \quad (13d) \]

where \( I_{G_{\text{max}}} = \text{min}(\mu_{\text{NIR}} + 2.5\sigma_{\text{NIR}}, 0.8) \). \( I_{B_{\text{max}}} = \text{min}(\mu_{\text{Blue}} + 3\sigma_{\text{Blue}}, 0.85) \). \( I_{\text{Blue}} \) is the blue channel image of a RGB image. The other two channels (red and green) are not used for the color fusion.

5. Experimental results and discussions

Two sets of multispectral images were used in our experiments, which were taken at night time and referred as to “NV-set 1” and “NV-set 2”. In NV-set 1, three pairs of multispectral images (as shown Figs. 1-3), image intensified (II) and long wave infrared (LWIR), were analyzed by using the aDWT fusion algorithm and segmentation-based colorization (also referred as to “local coloring”) algorithm as described in Section 3. The results of segmentation-based colorization are illustrated in Figs. 1-3. Note that there was no post-processing imposed on the resulted fusion and colorization images.

The two input images and the fused images used in the coloring process are shown in Figs. 1-3a, Figs. 1-3b and Figs. 1-3c, respectively. The image resolutions are given in figure captions. Two input images in NV-set 1 were preregistered. The false colored images (not shown in Figs. 1-3) 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 image 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 \( l\alpha\beta \) space through clustering and
Fig. 1. Segmentation-based colorization with Sample #1 (531×401 pixels) in NV-set 1: (a) and (b) are II and LWIR images; (c) Fused image by aDWT; (d) is the segmented image from a false-colored image (not shown), where 16 segments were merged from 36 clusters; (e) 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; (f) Channel-based color fusion of (IIÅ LWIR).

Fig. 2. Segmentation-based colorization with Sample #2 (360×270 pixels) in NV-set 1: (a) and (b) are II and LWIR images; (c) Fused image by aDWT; (d) is the segmented image of 12 segments merged from 21 clusters; (e) is the colored image with five auto-classified color schemes (plants, roads, ground, building and others); (f) Channel-based color fusion of (IIÅ LWIR).
Fig. 3. Segmentation-based colorization with Sample #3 (360×270 pixels) in NV-set 1: (a) and (b) are II and LWIR images; (c) Fused image by aDWT; (d) is the segmented image of 14 segments merged from 28 clusters; (e) is the colored image with three auto-classified color schemes (plants, smoke and others); (f) Channel-based color fusion of (II ⊕ LWIR).

merging operations (the clustered images are not shown in Figs. 1-3). The parameter values used in clustering and merging are \( N_{\text{Bin}} = [24 \ 24 \ 24] \), \( w_k = [0.25 \ 0.35 \ 0.40] \) and \( T_Q = 0.90 \). To emphasize two chromatic channels (due to more distinguishable among segments) in \( la\beta \) space, relatively larger weights were assigned in \( w_k \). With the segment map as shown in Figs. 1-3d, the histogram-matching and statistic-matching were performed segment by segment in \( la\beta \) space. The source region segments were automatically recognized and associated with proper target color schemes (after the training process is done). The finally colored images by segmentation-based colorization are shown in Figs. 1-3e. From a visual examination, the colored images 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 coloring process is fully automatic and well adaptive to different types of multispectral images.

Two-band channel-based color fusion (described in Eqs. (9)) was applied to the II and LWIR images (shown in Figs. 1-3a, b), and the results are illustrated in Figs. 1-3f. The color fusion results are very good especially in representing vegetation. Compared to the segmentation-based colorization results, the channel-based color fusion seems less realistic such as the sky and roads shown in Figs. 1-2f. However, the processes of channel-based color fusion eliminate the needs of segmentation and classification, and also reduced the color transforms. The processing speed of is much faster than that of segmentation-based colorization.

In NV-set 2, four pairs of multispectral images (as shown Figs. 4-7), color RGB, near infrared (NIR) and long wave infrared (LWIR), were analyzed by using the channel-based color fusion algorithm as described in Section 4. The results of channel-based color fusion are presented in Figs. 4-8.
The three-band input images used in the color fusion process are shown in Figs. 4-7a, b and c, respectively. The image resolutions are given in figure captions. The RGB images and LWIR images were taken by a FLIR SC620 two-in-one camera, which has LWIR camera (of 640x480 pixel original resolution and 7.5~13 μm spectral range) and an integrated visible-band digital camera (2048x1536 pixel original resolution). The NIR images were taken by a FLIR SC6000 camera (640x512 pixel original resolution and 0.9~1.7 μm spectral range). Two cameras (SC620 and SC6000) were set on the same fixture by turns and aimed at the same direction. The images were captured during sunset time and dusk time in fall season. Of course, image registration as described in Section 2.2 was applied to the three band images shown in Figs. 4-7, where manual alignments were employed to the RGB images shown in Figs. 6-7a since those visible images are so dark and noisy. To better present the RGB images, contrast and brightness adjustments (as described in figure captions) were applied. Notice that piecewise contrast stretching (Eq. (1)) was used for NIR enhancements. The fused images using a DWT algorithm was shown in Figs. 4-7d. Two-band channel-based color fusion (Eqs. (10)) was applied to the NIR and LWIR images (shown in Figs. 4-7b, c), and the results are illustrated in Figs. 4-7e; while three-band color fusion (Eqs. (13)) of (RGB ⊕ NIR ⊕ LWIR) are shown in Figs. 4-7f. Relative to gray-fusion (Figs. 4-7d), the images shown in two-band color fusion (Figs. 4-7e) resemble natural colors, which makes scene classification much easier. In the color-fusion images, the trees and grasses can be easily distinguished from grounds (parking lots) and sky. The car and person are easily identified in Figs. 6-7e. In Fig. 6e, the water area (between ground and trees, shown in cyan color) is clearly noticeable, but it is hard to realize the water area in the gray-fusion image (Fig. 6d). There is some improvement in three-band color fusion of (RGB ⊕ NIR ⊕ LWIR) in Figs. 4-5f when the light condition is good. For example, the tree, sky and ground shown in Figs. 4-5f are represented in more realistic colors than that in Figs. 4-5e. However, there is no significant difference between two-band and three-band color fusions as shown in Figs. 6-7 because the RGB images were taken at poor lighting condition.

The two-band channel-based color fusion of (RGB ⊕ LWIR) as defined in Eq. (11) is demonstrated in Fig. 8a-c; while the color fusion of (RGB ⊕ NIR) as defined in Eq. (12) is illustrated in Fig. 8d-f. No additional brightness or contrast adjustments were applied to these color-fusion images. In Fig. 8, the top-row images appear reddish, while the bottom-row images show greenish. These color-fusion images (under poor illumination) are not very realistic but have better representations and visibilities than the original RGB images (Figs. 4-6a). No color fusions of (RGB ⊕ LWIR) or (RGB ⊕ NIR) using the images shown in Fig. 7 are presented here due to the poor quality of RGB image (Fig. 7a).

The segmentation-based colorization demonstrated here took two-band multispectral images (II and LWIR) as inputs. Actually, this segmentation-based colorization procedure can accept two or three input images (e.g., II, NIR, LWIR). If there are more than three bands of images available (e.g., II, NIR, MWIR, LWIR), 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, I_{R1}, I_{R2}, I_{R3}, the two chosen images for colorization, I_{C1}, I_{C2}, should satisfy the following equation: \(Fus(I_{C1}, I_{C2}) = \max\{Fus(I_{R1}, I_{R2}), Fus(I_{R1}, I_{R3}), Fus(I_{R2}, I_{R3})\}\), where \(Fus\) stands for the fusion process and \(\max\) means selecting the fusion of maximum information.
Fig. 4. Channel-based color fusion with Sample #1 (Case# AT008 – sunset time; 640×480 pixels) in NV-set 2: (a) Color RGB image (contrast increased by 10%); (b) NIR image; (c) LWIR image; (d) Fused image of (b) & (c) by aDWT algorithm; (e) Channel-based color fusion of (NIR Å LWIR); (f) Channel-based color fusion of (RGB Å NIR Å LWIR).

Fig. 5. Channel-based color fusion with Sample #2 (Case# AT010 – after sunset; 640×480 pixels) in NV-set 2: (a) Color RGB image (brightness and contrast both increased by 10%); (b) NIR image; (c) LWIR image; (d) Fused image of (b) & (c) by aDWT algorithm; (e) Channel-based color fusion of (NIR Å LWIR); (f) Channel-based color fusion of (RGB Å NIR Å LWIR).
**Fig. 6.** Channel-based color fusion with Sample #3 (Case# AT012 – dusk time; 640×480 pixels) in NV-set 2: (a) Color RGB image (brightness and contrast both increased by 10%); (b) NIR image; (c) LWIR image; (d) Fused image of (b) & (c) by aDWT algorithm; (e) Channel-based color fusion of (NIR Å LWIR); (f) Channel-based color fusion of (RGB Å NIR Å LWIR).

**Fig. 7.** Channel-based color fusion with Sample #4 (Case# AT013 – dusk time; 640×480 pixels) in NV-set 2: (a) Color RGB image (brightness and contrast both increased by 10%); (b) NIR image; (c) LWIR image; (d) Fused image of (b) & (c) by aDWT algorithm; (e) Channel-based color fusion of (NIR Å LWIR); (f) Channel-based color fusion of (RGB Å NIR Å LWIR).
Fig. 8. Two-band channel-based color fusion: (a-c) Color fusion of (RGB ⊕ LWIR); (d-f) Color fusion of (RGB ⊕ NIR). Original images are shown in Figs. 4-6a, b, and c.

We exhibited the channel-based color fusion with possible combinations of two-band and three-band multispectral images. The processing speed of channel-based fusion is much faster than segmentation-based colorization, while the colors in channel-based fusion are less natural than the colors in segmentation-based colorization. Parameter settings in channel-based color fusion (Eqs. (9-13)) may be varied with different bands of images and with image capturing time and season, which can be conducted and stored before a field application.

6. Conclusions

In this chapter, a set of color fusion and colorization approaches are presented to enhance night vision for human users, which can be performed automatically and adaptively regardless of the image contents. Experimental results with multispectral imagery showed that the colored images contain clear information, and realistic colors. Specifically, the segmentation-based colorization (local-coloring) procedure is based on image segmentation, pattern recognition, and color mapping, which produces more colorful and more realistic colorized night-vision images. On the other hand, the channel-based color fusion procedure generates very impressive color-fusion images using linear transforms and channel assignments, which can be implemented very efficiently for real-time applications. The synthesized multispectral imagery with proposed colorizing approaches will eventually lead to improved performance of remote sensing, nighttime navigation, and situational awareness.
7. Acknowledgements

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

8. References


An Exploration of Color Fusion with Multispectral Images for Night Vision Enhancement


The purpose of this book is to provide an overview of basic image fusion techniques and serve as an introduction to image fusion applications in variant fields. It is anticipated that it will be useful for research scientists to capture recent developments and to spark new ideas within the image fusion domain. With an emphasis on both the basic and advanced applications of image fusion, this 12-chapter book covers a number of unique concepts that have been graphically represented throughout to enhance readability, such as the wavelet-based image fusion introduced in chapter 2 and the 3D fusion that is proposed in Chapter 5. The remainder of the book focuses on the area application-orientated image fusions, which cover the areas of medical applications, remote sensing and GIS, material analysis, face detection, and plant water stress analysis.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
