Qualitative and quantitative comparisons of multispectral night vision colorization techniques

Yufeng Zheng
Wenjie Dong
Erik P. Blasch

Qualitative and quantitative comparisons of multispectral night vision colorization techniques

Yufeng Zheng
Alcorn State University
Department of Advanced Technologies
Lorman, Mississippi 39096
E-mail: yzheng@alcorn.edu

Wenjie Dong
University of Texas-Pan American
Department of Electrical Engineering
Edinburg, Texas 78539

Erik P. Blasch
US Air Force Research Laboratory
Information Directorate
Rome, New York 13441

Abstract. Multispectral images enable robust night vision (NV) object assessment over day-night conditions. Furthermore, colorized multispectral NV images can enhance human vision by improving observer object classification and reaction times, especially for low light conditions. NV colorization techniques can produce the colorized images that closely resemble natural scenes. Qualitative (subjective) and quantitative (objective) comparisons of NV colorization techniques proposed in the past decade are made and two categories of coloring methods, color fusion and color mapping, are discussed and compared. Color fusion directly combines multispectral NV images into a color-version image by mixing pixel intensities at different color planes, of which a channel-based color fusion method is reviewed. Color-mapping usually maps the color properties of a false-colored NV image (source) onto that of a true-color daylight target picture (reference). Four coloring-mapping methods—statistical matching, histogram matching, joint histogram matching, and look-up table (LUT)—are presented and compared, including a new color-mapping method called joint-histogram matching (JHM). The experimental NV imagery includes visible (Red-Green-Blue), image-intensified, near infrared, and long-wave infrared images. The qualitative evaluations are conducted by visual inspections of the colorized images, whereas the quantitative evaluations are achieved by a newly proposed metric, objective evaluation index. From the experimental results according to both qualitative and quantitative evaluations, the following conclusions can be drawn: the segmentation-based colorization method produces very impressive and realistic colors but requires intense computations; color fusion and LUT-based methods run very fast but with less realistic results; the statistic-matching method always provides acceptable results; histogram matching and joint-histogram matching can generate impressive and vivid colors when the color distributions between source and target are similar; and the statistic-matching then joint-histogram matching (SM-JHM) method is a reliable and efficient method recommended from both qualitative and quantitative evaluations. © *2012 Society of Photo-Optical Instrumentation Engineers (SPIE).* [DOI: 10.1117/1.OE.51.8.087004]

Subject terms: color fusion; color mapping; multispectral images; night vision colorization; joint histogram matching; statistic matching then joint-histogram matching; objective evaluation index.

Paper 120786 received May 30, 2012; revised manuscript received Jul. 22, 2012; accepted for publication Jul. 26, 2012; published online Sep. 5, 2012.

1 Introduction

Multispectral images present complementary information that typically includes visual-band (e.g., RGB or intensified) imagery and infrared imagery [e.g., near infrared (NIR) and long-wave infrared (LWIR)]. Imagine a night-time object assessment task that may be executed on the ground or by aircraft equipped with a multispectral imaging system. Multispectral images enable night vision (NV) but it is inconvenient to directly observe and analyze multiple images of a scene. Instead, analyzing the synthesized (fused or colorized) multisensory image is more informative and efficient for target recognition.[1] The fused multispectral imagery (in gray scale) can increase the reliability of interpretation[2,3] and supports machine analysis (computer vision), whereas the colorized multispectral imagery (in colors) improves observer situational awareness,[4] reaction time,[5] and perceptual analysis (human vision).[6] This paper focuses on the discussion and comparisons of several NV coloring methods using multispectral images.

A NV colorization technique can produce colorized imagery with a naturalistic and stable color appearance by processing multispectral NV imagery. Although appropriately false-colored imagery is often helpful for human observers in improving their performance on scene classification and reaction time tasks,[6,7] inappropriate color mappings can also be detrimental to human performance.[5,8] A possible reason is lack of physical color constancy.[5] Another drawback with false coloring is that observers need specific training with each of the false color schemes so that they can correctly and quickly recognize objects. With colorized night-time imagery rendered with natural colors, users should be able to readily recognize and identify objects without any training.

Toet[9] proposed an NV colorization method that transfers the color characteristics of daylight imagery into multispectral NV images. Essentially, this 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). Thus, this method is referred as to "statistic matching." However, this color-mapping method colorizes the image regardless of scene content, and thus the accuracy of the coloring is dependent on how well the target and source images are matched. In other words, the statistic-matching method weighs the local regions of the source image by the "global" color statistics of the target image, and thus yields less naturalistic results (e.g., biased colors) for images containing regions that differ significantly in their colored content.

To address this bias problem in global coloring, Zheng and Essock[10] presented a "local coloring" method that can colorize the NV images to resemble daylight imagery. The local-coloring method renders the multispectral images with natural colors segment by segment (i.e., "segmentation-based"), and also provides an automatic association between the source and target images (i.e., by avoiding the manual scene-matching in global coloring).

The segmentation-based colorization can usually produce a vivid NV image closely resembling the colors in a natural scene. However, the segmentation-based coloring procedure involves many processes and heavy computations, such as image segmentation and pattern classification. Zheng[11] recently introduced a channel-based color-fusion method, which is fast enough for real-time applications. Note that the term "color fusion" in this paper refers to combing multispectral images into a color-version image with the purpose of resembling natural scenes.

Hogervorst and Toet[12,13] recently proposed a new color-mapping method using a look-up table (LUT). The LUT is created between a false-colored image (formed with multispectral NV images) and its color reference image (aiming at the same scene but taken during the daytime). The colors in the resulting colored NV image resemble the colors in the daytime color image. This LUT-mapping method, which runs fast for real-time implementations, is summarized along with the statistic-matching method in their recent paper.[13]

The quality of colorized images can be assessed by subjective/objective measures. However, subjective evaluation normally costs time and resources and should be related to a standard, such as the National Imagery Interpretability Rating Scale (NIIRS). Moreover, the subjective evaluation methods cannot be readily and routinely used for real-time and automated systems. On the other hand, objective evaluation metrics can automatically and quantitatively measure the image qualities.[14] In the past decade, many objective metrics for grayscale image evaluations have been proposed.[15–17] However, the metrics for grayscale images cannot be directly extended to the evaluations of colorized images. Recently, some objective evaluations of color images have been reported in the literature. To objectively assess a color-fusion method, Tsagaris[18] proposed a color image fusion measure (CIFM) by using the amount of common information between the source images and the colorized image, and also the distribution of color information. Yuan et al.[19] presented an objective evaluation method for visible and infrared color

fusion utilizing four metrics: image sharpness metric, image contrast metric, color colorfulness metric, and color naturalness metric. In this paper, we introduce an objective evaluation index (OEI) to quantitatively evaluate the colorized images. Given a reference (daylight color) image and several versions of the colorized NV images from different coloring techniques, all color images are first converted into International Commission on Illumination (CIE) LAB space, with dimension L for lightness and a and b for the color-opponent dimensions.[20] Then the OEI metric is computed with the four established metrics, phase congruency metric (PCM), gradient magnitude metric (GMM), image contrast metric (ICM), and color natural metric (CNM).

Certainly, a color presentation of multispectral NV images can provide a better visual input for human users. Users expect the colored images to closely resemble natural daylight pictures; along with a coloring process fast enough for real-time applications. In this paper, six NV coloring methods (i.e., color fusion, statistic matching, histogram matching, joint histogram matching, LUT-mapping, segmentation-based coloring) are explored and compared, using both qualitative and quantitative evaluations and employing a new color-mapping method of joint-histogram matching that is developed in the paper. Conclusions are drawn based the experimental results. The rest of paper is organized as follows. Multispectral image preprocessing and color space transform are briefly described in Sec. 2, channel-based color fusion is reviewed in Sec. 3, color-mapping methods are presented in Sec. 4, quantitative evaluation metrics are defined in Sec. 5 and the experiments and discussions are given in Sec. 6. Conclusions are drawn in Sec. 7.

2 Preprocessing and Color Space Transform

All NV colorization methods require preprocessing and color space transform, which are briefly reviewed in this section.

2.1 Multispectral Image Preprocessing

Multispectral images include visible (RGB), image-intensified (enhanced visible), NIR, and LWIR images. Before performing multispectral colorization, preliminary preprocessing, registration, and image fusion methods are required.[21] Standard image preprocessing such as denoising, normalization, and enhancement can support image registration, fusion, and colorization. Noise can be reduced according to the nature of the clutter that depends on a particular application. For example, a random noise can be suppressed by a Gaussian filter applied to noisy RGB and NIR images.

Night-vision images (NIR and LWIR) were acquired under different background and lighting 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 as:

$$\mathbf{I}_N = (\mathbf{I}_0 - I_{\text{Min}}) \frac{L_{\text{Max}} - L_{\text{Min}}}{I_{\text{Max}} - I_{\text{Min}}} + L_{\text{Min}}, \tag{1}$$

where $\mathbf{I}_N$ is the normalized image, $\mathbf{I}_0$ is the original image; I_{Min} and I_{Max} are the maximum and minimum pixel values in $\mathbf{I}_0$, respectively; L_{Min} and L_{Max} are the expected minimum and maximum pixel values in $\mathbf{I}_N$, which normally

equal 0 and 1, respectively. After image normalization, $I_N \in [0, 1]$, a common reference is established.

The image contrasts of NIR images are significantly affected by illumination conditions. Nonlinear enhancements, such as histogram equalization or histogram matching, usually increase noise while enhancing an NIR image. A linear enhancement such as piecewise contrast stretching is preferred. Equation (1) is still applicable but just applied within each range of an 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 simplify the notation, this transform can be denoted as $S_{[0,0.8]}^{[0,1.0]}$ hereafter, where S is the scaling operation.

Image registration is a required preparation for image fusion and colorization. In general, image registration aligns multiple images (e.g., NIR and LWIR) by performing affine transformations that allow translation, rotation, and scaling. Similarity metrics are used to decide the optimized transformation parameters. The details of image registration can be found elsewhere.[22,23]

Image fusion is a necessary step for most coloring methods. For example, the fused image is expected in color fusion (Sec. 3) and segmentation-based colorization (Sec. 4.5). Image fusion actually combines multisensory images into one image. An advanced discrete wavelet transform (aDWT) fusion method is used in our experiments, where the details of image fusion are documented elsewhere.[24]

2.2 Color Space Transform

All color-mapping methods are performed in a transformed color space, called $l\alpha\beta$ space. 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, α, and β) so that the manipulation (such as statistic matching and histogram matching) on each color component can be performed independently. Inverse transforms (e.g., $l\alpha\beta$ space to the LMS, LMS to LMS, LMS to RGB) are needed to complete the NV colorization process.[9]

The actual conversion (matrix) from RGB tristimulus to device-independent XYZ tristimulus values depends on the characteristics of the display being used. Fairchild[25] suggested a "general" device-independent conversion (without a priori knowledge about the display device) that maps white in the chromaticity diagram to white in the RGB space and vice versa.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.5141 & 0.3239 & 0.1604 \\ 0.2651 & 0.6702 & 0.0641 \\ 0.0241 & 0.1228 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \qquad (2)$$

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

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \qquad (3)$$

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

$$\boldsymbol{L} = \log L, \qquad \boldsymbol{M} = \log M, \qquad \boldsymbol{S} = \log S. \qquad (4)$$

Ruderman et al.[26] presented a color space, named $l\alpha\beta$ (Luminance-Alpha-Beta), which can decorrelate the three axes in the LMS space:

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.5774 & 0.5774 & 0.5774 \\ 0.4082 & 0.4082 & -0.8165 \\ 1.4142 & -1.4142 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{M} \\ \boldsymbol{S} \end{bmatrix}. \qquad (5)$$

The three axes can be considered as an achromatic direction ($l \propto r + g + b$), a yellow-blue opponent direction ($\alpha \propto r + g + b$), and a red-green opponent direction ($\beta \propto r - b$). The $l\alpha\beta$ space has the characteristics of compact, symmetrical and decorrelation, which highly facilitates the subsequent process of color-mapping (see Sec. 4).

3 Color Fusion Methods

A fast color fusion method, termed as channel-based color fusion, was introduced to facilitate real-time applications.[11] The term of "color fusion" means combing multispectral images into a color-version image with the purpose of resembling natural scenes. Relative to the "segmentation-based colorization" (refer to Sec. 4.5), color fusion performs a trade-off between color realism with processing speed.

The general framework of channel-based color fusion is as follows: prepare for color fusion, preprocessing (denoising, normalization, and enhancement) and image registration; form a color fusion image by properly assigning multispectral images to red, green, and blue channels; then fuse multispectral images (gray fusion) using aDWT algorithm;[24] and replace the value component of color fusion in color natural metric (HSV) color space with the gray-fusion image, and finally transform the fused image back to RGB space.

In NV imaging, there may be several bands of images available, for example, visible (RGB), image intensified (II), NIR, medium-wave infrared (MWIR), and LWIR. On the basis of the available images and the context, we only discuss two-band color fusions: (II $\oplus$ LWIR), and (NIR $\oplus$ LWIR), with the symbol '$\oplus$' denotes the fusion of multiband images.

3.1 Color Fusion of (II $\oplus$ 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,

$$F_R = S_{[0,1.0]}^{[0,0.7]}(I_{\text{LWIR}}), \qquad (6a)$$

$$F_G = S_{[0.1,\text{IGmax}]}^{[0.2,1]}(I_{\text{II}}), \qquad (6b)$$

$$F_B = S_{[0,1.0]}^{[0.1,0.75]}([1.0 - I_{\text{LWIR}}] \bullet I_{\text{II}}), \qquad (6c)$$

$$V_F = \text{Fus}(I_{\text{II}}, I_{\text{LWIR}}), \tag{6d}$$

where $S_{[0.1,I_{\text{Gmax}}]}^{[0.2,1]}$ denotes piecewise contrast stretching defined in Eq. (1); $I_{\text{Gmax}} = \min([\mu_{\text{II}} + 3\sigma_{\text{II}}], 0.8)$ are the mean and standard deviation of an II image; $[1.0 - I_{\text{LWIR}}]$ is to invert LWIR image; symbol '•' means element-by-element multiplication; V_F is the value component of F_C in HSV space; and $Fus()$ means image fusion operation using the aDWT algorithm.[24] Although the limits given in contrast stretching are obtained empirically according to the NV images available, it is viable to formulate the expressions and automate the fusion based upon a set of conditions (e.g., imaging devices, imaging time, and application location). Notice the transform parameters in Eqs. (6a) to (6d) are applied to all color fusion operations in our experiments.

3.2 Color Fusion of (NIR ⊕ LWIR)

A color fusion of NIR and LWIR is formulated by,

$$F_R = S_{[0,1.0]}^{[0.2,0.9]}(I_{\text{LWIR}}), \tag{7a}$$

$$F_G = S_{[0.1,\text{IGmax}]}^{[0.2,1]}(I_{\text{NIR}}), \tag{7b}$$

$$F_B = S_{[0,1.0]}^{[0.1,0.7]}([1.0 - L_{\text{LWIR}}] \bullet I_{\text{NIR}}), \tag{7c}$$

$$V_F = \text{Fus}(I_{\text{NIR}}, I_{\text{LWIR}}), \tag{7d}$$

where $I_{\text{Gmax}} = \min([\mu_{\text{NIR}} + 2\sigma_{\text{NIR}}], 0.8)$ and $\min()$ is an operation to get the minimal number. Other parameters and operators are the same as that in Eqs. (6a) to (6d).

4 Color-Mapping Methods

All color-mapping methods described in this section are performed in $l\alpha\beta$ color space. Thus, the color space conversion from RGB to $l\alpha\beta$ must be done prior to color mapping, and then the inverse transform to RGB space is necessary after the mapping. Certainly, two images, a source image and a target image, are involved in a color-mapping process (refer to Fig. 1). The source image is usually a color-fusion image (in Secs. 4.1 to 4.3) or a false-colored image (in Sec. 4.4); while the target image is normally a daylight picture containing the similar scene. The target image may have a different resolution as depicted in Secs. 4.1 to 4.3; however, the LUT described in Sec. 4.4 is established using the registered target (reference) image.

4.1 Statistic Matching

A statistic matching (stat-match) is used to transfer the color characteristics from natural daylight imagery to false color night-vision imagery, which is formulated as:

$$I_C^k = (I_S^k - \mu_S^k)\frac{\sigma_T^k}{\sigma_S^k} + \mu_T^k, \qquad \text{for } k = \{l, \alpha, \beta\}, \tag{8}$$

where I_C is the colored image, I_S is the source (false-color) image in $l\alpha\beta$ space; μ denotes the mean and σ denotes the standard deviation; the subscripts 'S' and 'T' refer to the
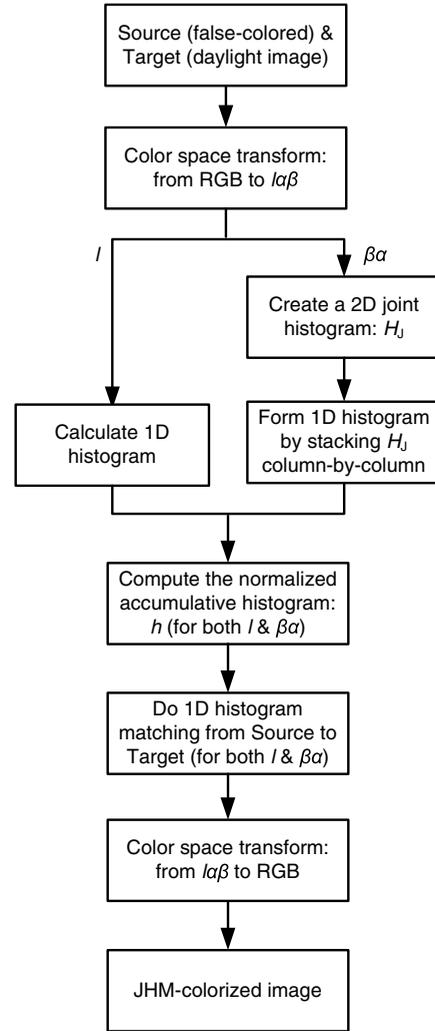

**Fig. 1** Diagram of joint histogram matching [demonstrated with joint-HM($\beta\alpha$)].

source and target images, respectively; and the superscript 'k' is one of the color components: $\{l, \alpha, \beta\}$.

After this transformation, the pixels comprising the multispectral source image have means and standard deviations that conform to the target daylight color picture in $l\alpha\beta$ space. The colored image is transformed back to the RGB space [refer to Fig. 2(e)] through the inverse transforms [i.e., $l\alpha\beta$ space to the **LMS**, exponential transform from **LMS** to LMS, and LMS to RGB, refer to Eqs. (2) to (5].[9]

4.2 Histogram Matching

Histogram matching (i.e., histogram specification) is usually used to enhance an image when histogram equalization fails.[27] Given the shape of the histogram that we want the enhanced image to have, histogram matching can generate a processed (i.e., matched) 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.

Histogram matching can be implemented as follows. First, the normalized cumulative histograms of source
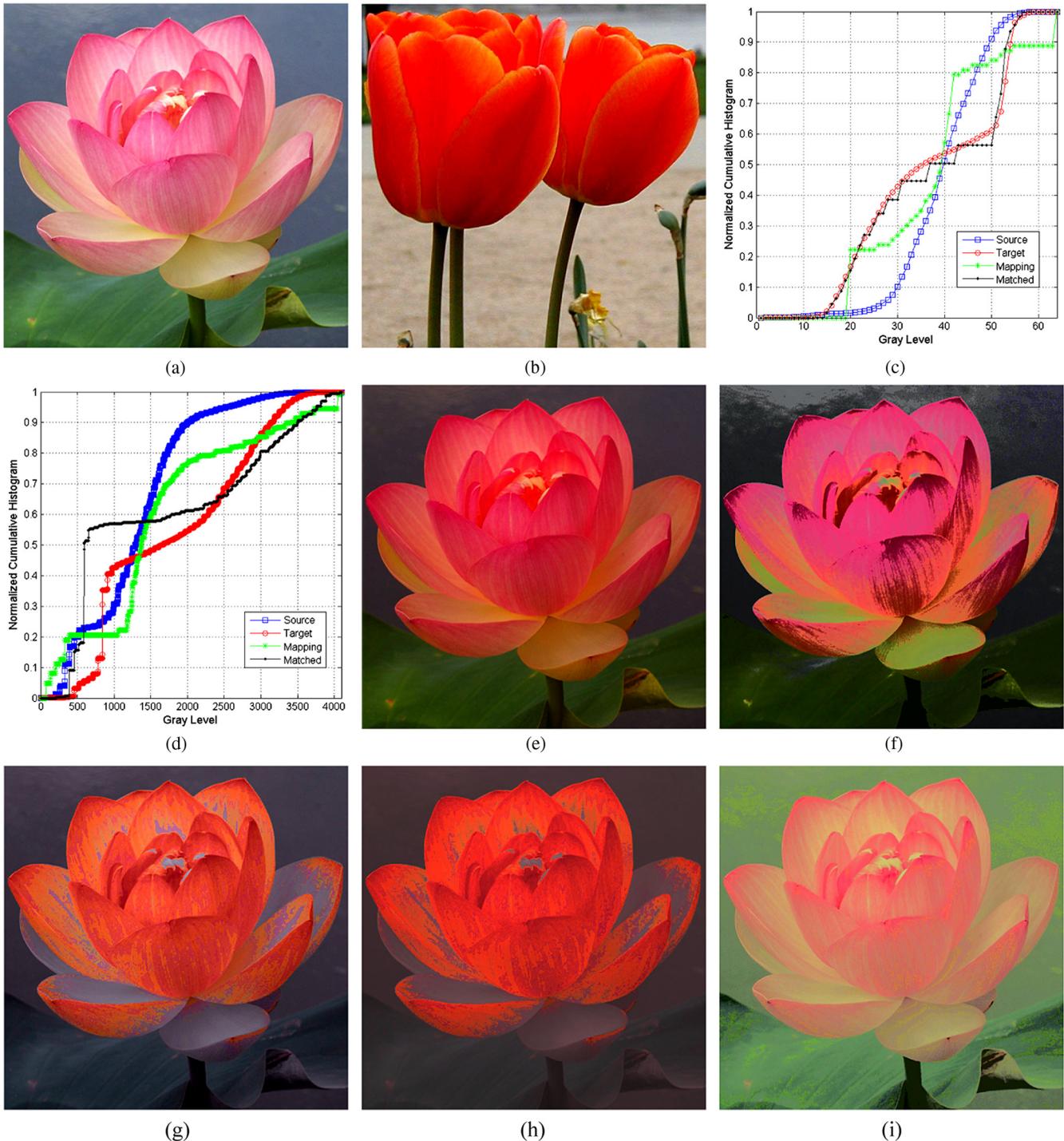
Fig. 2 Illustration of color-mapping techniques using only daylight RGB pictures: (a) and (b) source (Lotus) and target (Tulip) images; (c) and (d) the cumulative histograms of *Luminance* (*l*), and the 1-D histograms from the 2-D JHs of *Beta-Alpha* (*βα*) in processing (g); (e) and (f) the mapped images using statistic-matching, and histogram-matching (HM), respectively; (g)–(i) the mapped images using joint-HM (*βα*), statistic-matching then joint-HM (*βα*), and joint-HM (*lα*), respectively. Notice that the contrasts of all color images were increased by 10%.

image $[h_S = S(u_k)]$ and target image $[h_T = T(v_k)]$ are calculated, respectively.

$$h_S = S(u_k) = (L - 1) \cdot \sum_0^{L-1} \frac{n_k}{N}, \qquad (9)$$

where N is the total number of pixels in the image, n_k is the number of pixels that have gray level u_k, and L is the number of gray (bin) levels in the image. Typically, $L = 256$ for a

digital image. But we can round the image down to m ($m < L$, e.g., $m = 64$) levels, and thus its histogram is called a m-bin histogram. Clearly, $S(u_k)$ is a nondecreasing function. Similarly, $h_T = T(v_k)$, where v_k is the gray level in the target image, can be computed [see the "Target" curve in Fig. 2(c)].

Second, considering $h_S = h_T$ [i.e., $S(u_k) = T(v_k)$] for histogram matching, the matched image is accordingly computed as

$$v_k = T^{-1}[S(u_k)], \qquad k = 0, 1, 2, ..., L-1. \tag{10}$$

It is then straightforward to find a discrete solution of the inverse transform, $T^{-1}[S()]$ [see the "Mapping" curve in Fig. 2(c)] as both $T()$ and $S()$ can be implemented with look-up tables (LUT).

Similar to the statistic matching (described in Sec. 4.1), histogram matching also serves for color-mapping and is performed component-by-component in $l\alpha\beta$ space [refer to Fig. 2(f)]. Specifically, with each color component (say the α component, treated as a grayscale image) of a false-colored image, we can compute $S(u_k)$. With a selected target image, $T(v_k)$ can be calculated with regard to the same color component (say α). Using Eq. (10), the histogram matching can be completed regarding that color component (α). Histogram matching and statistic matching can be applied separately or jointly and when applied together, for instance, it is referred to "statistic matching then histogram matching."

4.3 Joint Histogram Matching

As described in Sec. 4.2, histogram matching is applied to each color component (plane) separately. It is possible to distort the color distributions of the mapped image [see Fig. 2(f)]. To avoid color distortion, we introduce a new color-mapping method, joint histogram matching (joint-HM).

In $l\alpha\beta$ space, α and β represent the color distributions; while l is the intensity component. In this paper, a joint histogram (also called two-dimensional (2-D) histogram) of "two color planes" (α versus β) is calculated and then matched from the source to the target. The intensity component (l) is matched individually (using the same procedure as described in Sec. 4.2). A diagram of joint histogram matching is illustrated in Fig. 1. In the literature, the joint histogram usually means the joint (2-D) intensity distribution of "two grayscale images," which is often used to compute the joint entropy[22] for image registration.

How to calculate the normalized cumulative histogram (denoted as h) from a 2-D joint histogram (denoted as H_J) needs further discussion. For histogram matching, h is expected to be a nondecreasing function. As illustrated in Fig. 1, we propose to form a one-dimensional (1-D) histogram by stacking H_J column-by-column and then perform histogram matching, as defined in Eq. (10). Of course, to correctly index a 1-D transform [$T^{-1}()$], the proper calculation of u_m (with m bins) using two gray levels (e.g., β and α) is expected. If H_J is computed as (β versus α), its matching process is denoted as joint-HM($\beta\alpha$) [see Figs. 1 and 2(g)]. Theoretically joint-HM($\beta\alpha$) and joint-HM($\alpha\beta$) should be the same, but our process (the formation of 1-D histogram from a 2-D H_J) makes them eventually different. Another interesting aspect of joint-HM($l\alpha$) is presented in Fig. 2(i). As shown in Fig. 2(d), the histogram of the mapped image (the "Mapped" curve) is sort of a trade-off between two histograms, "Source" and "Target." This is expected since we want no color distortion (i.e., preserving its own colors to some extent) during color mapping. In addition, joint-HM can be also applied together with statistic matching such as "statistic matching then joint-HM," referred to as SM-JHM [see Fig. 2(h)].

4.4 Look-Up Table

Hogervorst and Toet[12,13] proposed a color-mapping method using a look-up table (LUT). The LUT is created using one image pair: a false-colored image (formed with two-band NV images) and a reference (i.e., target) daylight image. This method yields a colored NV image similar to the daytime image colors. The implementation of this LUT method is described as follows.

(1) Create a false-colored image (of three color planes) by assigning LWIR to R, NIR image to G plane, and zeros to B, respectively;

(2) Build RG color map (i.e., a 256×256 LUT) and convert the false-colored image to an indexed image (0 to 65535) associated with the RG color map;

(3) For all pixels in the indexed false-colored image whose index value equals 0:

   (a) Locate all corresponding pixels in the reference (i.e., target) color image (that must be strictly aligned with the false-colored image);

   (b) Calculate the averaged $l\alpha\beta$ values of those corresponding pixels and then convert them back to RGB values;

   (c) Assign the RGB values to index 0 in the lookup table;

(4) Vary the index value from 1 to 65535 and repeat the processes described in Step 3. At the end, the LUT will be established.

Once the LUT is created, the LUT-based mapping procedure is simple and fast [see Fig. 3(i)], and thus can be deployed in real-time. However, the LUT creation thoroughly relies on the aligned reference image corresponding to the same scene. Any misalignment, using a different reference color image, or coloring a different NV imagery (i.e., aiming at a different direction), will usually result a degraded colorization. To make the LUT colorization independent of viewing direction, Hogervorst and Toet[12,13] further suggested deriving the LUT table from more than one corresponding image pair (false-colored versus daylight) such that all materials relevant for a given surroundings are represented in the imagery from which the LUT is derived.

4.5 Segmentation-Based Colorization

In segmentation-based colorization (also called "local coloring") method,[9] multispectral NV imagery is rendered "segment-by-segment" with the statistical color properties of natural scenes by using either statistic matching or histogram matching. Therefore, this is not a new color-mapping technique but just uses the existing methods differently. Eventually, the colorized images resemble daylight pictures. The main steps of segmentation-based colorization are summarized below, but the details are given elsewhere.[9] 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. Then, the false-colored image is segmented using the features of color properties, and the techniques of nonlinear diffusion, clustering, and region merging. A set
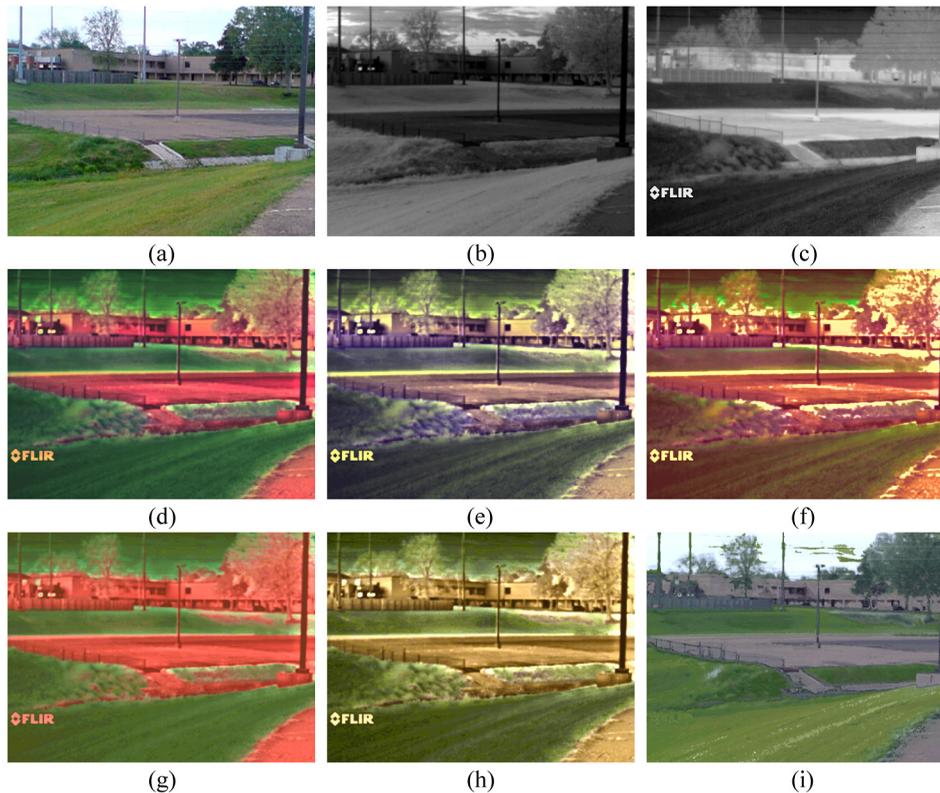
Fig. 3 Night-vision coloring comparison (Case# ST014 in NV-set 1—taken at sunset; 640 × 480 pixels): (a-c) Color RGB, NIR, and LWIR images, respectively; (d)–(f) the colorized images using channel-based color fusion of (NIR ⊕ LWIR), statistic-matching, and histogram-matching, respectively; (g)–(i) the colorized images using joint-HM, statistic-matching then joint-HM, and LUT-mapping [reference = (a)], respectively. The settings in the color-mappings of (e)–(h) are source = (d) and target = (a). Notice that the contrasts of all color images were increased by 10%.

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. 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. 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. The color-mapping procedures (statistic-matching then histogram-matching) are carried out to render natural colors onto the false-colored image segment by segment. The mapped image is then transformed back to the RGB space. Finally, the mapped image is transformed into HSV 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.

5 Quantitative Metrics for Colorization Evaluations

Three image quality metrics for grayscale images and one metric for color images are reviewed in Sec. 5.1. A new objective metric, termed object evaluation index[28] (OEI), is introduced in Sec. 5.2, which is defined with the four

metrics. The color-related metrics are defined in the CIELAB space, where CIE stands for the International Commission on Illumination and LAB is for $L^*a^*b^*$. The perceptually uniform CIELAB space consists of an achromatic luminosity component L^* (black-white) and two chromatic values a^* (green-magenta) and b^* (blue-yellow). The coordinates $L^*a^*b^*$ (CIE 1976) can be calculated using the CIE XYZ tri-stimulus values.[20]

5.1 Four Image Quality Metrics

5.1.1 Phase congruency metric

The phase congruency (PC) model is also called the "local energy model" developed by Morrone et al.[29] This model postulates that the features in an image are perceived at the points where the Fourier components are maximal in phase. Based on the physiological and psychophysical evidences, the PC theory provides a simple but biologically plausible model of how mammalian visual systems detect and identify the features in an image.[29–32] The PC can be considered as a significance measure of local structures in an image.

According to the definition of PC,[29] there are many different implementations developed so far. A widely-used method developed by Kovesi[30] is adopted in this paper. Given a 1-D image $f(x)$, M_n^e and M_n^o represent the even-symmetric and odd-symmetric filters at scale n, respectively. M_n^e and M_n^o form a quadrature pair: $e_n(x)$ and $o_n(x)$. Responses of the quadrature pair form a response vector:

$$\begin{bmatrix} e_n(x) \\ o_n(x) \end{bmatrix} = \begin{bmatrix} f(x) * M_n^e \\ f(x) * M_n^o \end{bmatrix}, \tag{11a}$$

and the local amplitude at scale n is

$$A_n(x) = \sqrt{e_n^2(x) + o_n^2(x)}. \tag{11b}$$

Let

$$F(x) = \sum_n e_n(x), \qquad H(x) = \sum_n o_n(x). \tag{11c}$$

The 1-D PCM can be computed as

$$PC(x) = \frac{\sqrt{F^2(x) + H^2(x)}}{\sum_n A_n(x) + \varepsilon}, \tag{11-d}$$

where ε is a small positive constant.

In order to calculate the quadrature pair of filters M_n^e and M_n^o, Gabor filters[33] or log-Gabor filters[34] can be applied. In this paper, we use log-Gabor filters (e.g., wavelets at scale $n = 4$) due to its following two features: log-Gabor filters, by definition, have no direct current (DC) component; and the transfer function of the log-Gabor filter has an extended tail at the high frequency end, which makes it more capable to encode natural images than ordinary Gabor filters.[35] The transfer function of a log-Gabor filter in the frequency domain is

$$G(\omega) = e^{-\frac{[\log(\omega/\omega_0)]2,}{2\sigma_r^2}}, \tag{12a}$$

where ω_0 is the filter's center frequency and σ_r controls the filter's bandwidth.

To compute the PCM of 2-D grayscale images, we can apply the 1-D analysis over several orientations and then combine the results according to some rules that optimize performance. The 1-D log-Gabor filters described above can be extended to 2-D ones by applying a Gaussian function across the filter perpendicular to its orientation.[30,34,36,37] The 2-D log-Gabor function has the following transfer function

$$G_2(\omega, \theta_j) = e^{-\frac{[\log(\omega/\omega_0)]^2}{2\sigma_r^2}} \cdot e^{-\frac{(\theta-\theta_j)^2}{2\sigma_\theta^2}}, \tag{12b}$$

where $\theta_j = \frac{j\pi}{2J}$ and $j = 0, 1, 2, \ldots, J - 1$ is the number of orientations and σ_θ determines the filter's angular bandwidth. By modulating ω_0 and θ_j and convolving G_2 with the 2-D image, we get a set of responses at each point (x, y) as $[e_{n,\theta_j}(x, y), o_{n,\theta_j}(x, y)]$. The local amplitude at scale n and orientation θ_j is

$$A_{n,\theta_j} = \sqrt{e_{n,\theta_j}^2(x, y) + o_{n,\theta_j}^2(x, y)} \tag{13a}$$

and the local energy along orientation θ_j is

$$E_{\theta_j} = \sqrt{F_{\theta_j}^2(x, y) + H_{\theta_j}^2(x, y)}, \tag{13b}$$

where

$$F_{\theta_j}(x, y) = \sum_n e_{n,\theta_j}(x, y), \qquad H_{\theta_j}(x, y) = \sum_n o_{n,\theta_j}(x, y). \tag{13c}$$

The 2-D PCM at (x, y) is defined as

$$PC_{2D}(x, y) = \frac{\sum_j E_{\theta_j}(x, y)}{\sum_n \sum_j A_{n,\theta_j}(x, y) + \varepsilon}, \tag{13d}$$

where ε is a small positive constant. It should be noted that $PC_{2D}(x, y)$ is a real number within [0,1]. The PCM of an image is defined as

$$PCM = \frac{1}{MN} \sum_{(x,y)} PC_{2D}(x, y)$$
$$= \frac{1}{MN} \sum_{(x,y)} \frac{\sum_j E_{\theta_j}(x, y)}{\sum_n \sum_j A_{n,\theta_j}(x, y) + \varepsilon}, \tag{13e}$$

where $M \times N$ is the size of the image. The range of PCM is [0,1].

5.1.2 Gradient magnitude metric

The image gradient magnitude (GM) is computed to encode contrast information. So, PC and GM are complementary and reflect different aspects of the human visual system in assessing the local image quality, with GM measuring the sharpness of an image. The perception of sharpness is related to an image's clarity of detail. Image gradient computation is a traditional topic in image processing and gradient operators can be expressed by convolution masks. One commonly used gradient operator is the Sobel operator. The partial derivatives of image $f(x, y)$, G_x and G_y, along horizontal and vertical directions using the Sobel operators are

$$G_x = \frac{1}{4} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * f(x, y),$$
$$G_y = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * f(x, y). \tag{14a}$$

The GM of $f(x, y)$ at pixel (x, y) is defined as

$$G(x, y) = \sqrt{G_x^2 + G_y^2}. \tag{14b}$$

The averaged GM over all pixels is called image GMM,

$$GMM = \frac{1}{MN} \sum_{x,y} G(x, y) = \frac{1}{MN} \sum_{x,y} \sqrt{G_x^2 + G_y^2}, \tag{14c}$$

where $M \times N$ is the size of the image.

5.1.3 Image contrast metric

An image with excellent contrast has a wide dynamic range of intensity level and appropriate intensity. Both the dynamic range of intensity level or the overall intensity distribution of the image can be provided by a histogram. A global contrast

metric[19] is proposed using the histogram character. The histogram of image with levels in the range $[0, N-1]$ is a frequency-distribution function defined as the overall intensity distribution of an image

$$h(X_k) = n_k, \tag{15a}$$

where X_k is the k'th level of input and n_k is the number of the pixels in the image having level X_k. The probability density function (PDF) is computed by

$$P(X_k) = n_k/n, \tag{15b}$$

where n is the total number of the pixels of the image. The dynamic range value β is defined as

$$\beta = \sum_{k=0}^{L-1} S(X_k), \tag{15c}$$

where

$$S(X_k) = \begin{cases} 1, & \text{if } P(X_k) > 0 \\ 0, & \text{otherwise} \end{cases}. \tag{15d}$$

The dynamic range matrix α of histogram is defined as

$$\alpha = \frac{\beta}{2\,N - \beta}, \tag{15e}$$

where $\alpha \in [0, 1]$. Note that a larger value of α means a wider dynamic range in the histogram, which leads to better contrast. The local image contrast metric is defined as

$$C = \alpha \sum_{k=0}^{N-1} \frac{X_k}{N} P(X_k). \tag{15f}$$

For color images, the image contrast metric is determined by both gray contrast and color contrast. Because human perception is more sensitive to the luminance on contrast evaluation, we employ the L^* channel in the CIELAB space to evaluate the color contrast. Thus, image contrast is determined by the histogram of gray intensity and the histogram of color luminance L^* (see Fig. 4). For the gray intensity I, the gray contrast metric is defined as

$$C_g = \alpha_I \sum_{k=0}^{N_I-1} \frac{I_k}{N} P(I_k), \tag{16a}$$

where α_I and $P(I_k)$ can be calculated as above for gray intensity. For the L^* channel, the color contrast metric is

$$C_c = \alpha_c \sum_{k=0}^{N_{L^*}-1} \frac{L_k^*}{N_{L^*}} P(L_k^*), \tag{16b}$$

where α_c and $P(L_k^*)$ can be calculated as above for the L^* channel. The global image contrast metric (ICM) is defined as

$$\text{ICM} = \sqrt{\omega_1 C_g^2 + \omega_2 C_c^2}, \tag{16c}$$

where ω_1 and ω_2 are the weights of C_g and C_c. For simplicity, we choose $\omega_1 = \omega_2 = 0.5$. ICM varies within [0,1]. The evaluation of image contrast metric of color fusion image is shown in Fig. 4.

5.1.4 *Color natural metric*

Given a daylight image $f_1(x, y)$ and a colorized image $f_2(x, y)$, if a colorized image is similar to the daylight image then the colorized image is considered good quality. Since a human is sensitive to hue in addition to luminance, we compare the a^* and b^* channels of the reference image with that of the colorized image using the gray relational analysis (GRA) theory.[38]

We first convert two images, f_1 and f_2, to $L^*a^*b^*$ space. $L_i^*(x, y)$, $a_i^*(x, y)$, and $b_i^*(x, y)$ are the $L^*a^*b^*$ values of f_i at pixel (x, y). The gray relation coefficient between a_1^* and a_2^* at pixel (x, y) is defined as

$$\xi_a(x, y) = \frac{\min_i \min_j |a_1^*(i, j) - a_2^*(i, j)| + 0.5 \max_i \max_j |a_1^*(i, j) - a_2^*(i, j)|}{|a_1^*(x, y) - a_2^*(x, y)| + 0.5 \max_i \max_j |a_1^*(i, j) - a_2^*(i, j)| + \varepsilon}, \tag{17a}$$

where ε is a small positive constant. The gray relation coefficient between b_1^* and b_2^* at pixel (x, y) is defined as

$$\xi_b(x, y) = \frac{\min_i \min_j |b_1^*(i, j) - b_2^*(i, j)| + 0.5 \max_i \max_j |b_1^*(i, j) - b_2^*(i, j)|}{|b_1^*(x, y) - b_2^*(x, y)| + 0.5 \max_i \max_j |b_1^*(i, j) - b_2^*(i, j)| + \varepsilon}. \tag{17b}$$

In the definitions of $\xi_a(x, y)$ and $\xi_b(x, y)$, $\min(\cdot)$ and $\max(\cdot)$ are operated over whole image. However, it is possible that $\min(\cdot)$ and $\max(\cdot)$ are operated over a small neighborhood of (x, y). The gray rational degrees of a^* and b^* information for two images are defined as

$$R_a = \sum_{(x,y)} \omega(x, y) \xi_a(x, y), \tag{17c}$$

$$R_b = \sum_{(x,y)} \omega(x, y) \xi_b(x, y), \tag{17d}$$

where $\omega(x, y)$ is the weight of the gray rational coefficient, which satisfies

$$\sum_{(x,y)} \omega(x, y) = 1. \tag{17e}$$

For simplicity, we choose $\omega(x, y) = \frac{1}{M \times N}$ where M and N are the length of vectors x and y, respectively. The CNM is defined as
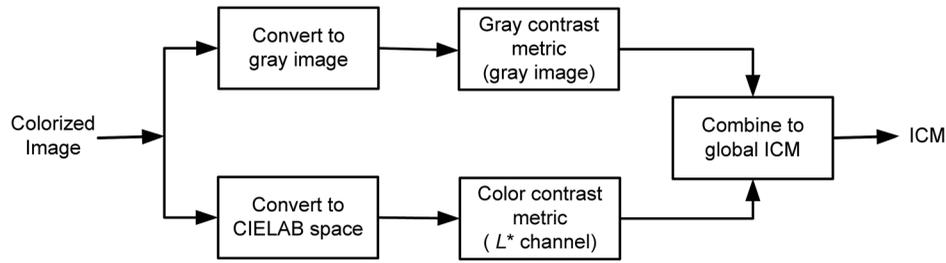
$$\text{CNM} = \sqrt{R_a R_b}. \tag{17f}$$

Fig. 4 Diagram of calculation of the contrast metric.

CNM varies within [0,1]; the larger the CNM, the more similar the two images.

5.2 Objective Evaluation Index

With the four metrics defined in Sec. 5.1, a new OEI is proposed to quantitatively evaluate the qualities of colorized images. Given the reference image f_1 and the colorized image f_2, the OEI is calcualted in two steps. First the local similarity maps of the two images are computed, and then the similarity maps are integrated into a single similarity score.

The two images are first converted into $L^*a^*b^*$ space. For L^* information, the PC maps are calculated and denoted as PC_1 and PC_2 for f_1 and f_2 images, respectively. The similarity measure, S_{PC}, between PC_1 and PC_2 at pixel (x, y) is defined as

$$S_{PC}(x,y) = \frac{2\,PC_1(x,y)PC_2(x,y) + K_1}{PC_1^2(x,y) + PC_2^2(x,y) + K_1}, \tag{18a}$$

where K_1 is a positive constant. In practice, the determination of K_1 depends on the dynamic range of PC values. S_{PC} varies within [0,1]. Similarly, the similarity measure based on the two GM values is defined as

$$S_G(x,y) = \frac{2G_1(x,y)G_2(x,y) + K_2}{G_1^2(x,y) + G_2^2(x,y) + K_2}, \tag{18b}$$

where K_2 is a positive constant. S_G varies within [0,1]. Then, $S_{PC}(x,y)$ and $S_G(x,y)$ are combined into one similarity measure, $S_L(x)$, where subscripts L is for $L*a*b$ space, as follows

$$S_L(x,y) = [S_{PC}(x,y)]\lambda_1 [S_G(x,y)]\lambda_2, \tag{18c}$$

where λ_1 and λ_2 are parameters to adjust the relative importance of PC and GM features.

With the aid of the similarity $S_L(x, y)$ at each pixel (x, y), the overall similarity between f_1 and f_2 can be calculated with the averaged $S_L(x, y)$ over all pixels. However, the image saliency (i.e., local significance) usually varies with the pixel location. For example, edges convey more crucial information than smooth areas. Specifically, a human is sensitive to phase congruent structures,[39] and thus the larger $PC(x, y)$ value between f_1 and f_2 implies a higher impact on evaluating the similarity between f_1 and f_2 at location (x, y). Therefore, we use $PC_{max}(x, y) = \max[PC_1(x, y), PC_2(x, y)]$ to weigh the importance of $S_L(x, y)$ in formulating the overall similarity. Accordingly, the OEI between f_1 and f_2 is defined as follows

$$OEI = \left(\frac{\sum_{(x,y)}PC_{max}(x,y)S_L(x,y)}{\sum_{(x,y)} PC_{max}(x,y)} \right)^{\gamma_1} \times (S_{ICM})^{\gamma_2}$$
$$\times (CNM)^{\gamma_3}, \tag{19a}$$

where

$$PC_{max}(x,y) = \max[PC_1(x,y), PC_2(x,y)], \tag{19b}$$

$$S_{ICM} = \frac{2\,ICM(f_1) \times ICM(f_2) + K_3}{ICM(f_1)2 + ICM(f_2)2 + K_3}, \tag{19c}$$

CNM is previously defined in Eq. (17f) and K_3 and γ_i ($i = 1, 2, 3$) are positive constants. The diagram of calculating OEI is shown in Fig. 5. The range of OEI is [0,1]. The larger the OEI value of a colorized image is, the more similar (i.e., the better) the colorized image is to the reference image. Error pooling is the integration of methods with tradeoffs between γ_1, γ_2, and γ_3.

γ_1, γ_2, and γ_3 are the weights of three components in the OEI metric, with the selection of γ_i being critical for the OEI calculation. The values of γ_i are empirically decided, and the typical values of γ_1 and γ_2 are between $0.8 \sim 1.1$ and γ_3 is between $0.05 \sim 0.2$. K_i ($i = 1, 2, 3$) are constants to increase the metric stability. In our experiments presented in Sec. 6, we chose $\gamma_1 = \gamma_2 = 1$, $\gamma_3 = 0.2$; $K_1 = 0.85$, $K_2 = 160$, $K_3 = 0.001$; and $\lambda_1 = \lambda_2 = 1$.

6 Experimental Results and Discussions

To demonstrate and compare different color-mapping methods, two daylight pictures were used as source [Fig. 2(a); Lotus] and target [Fig. 2(b); Tulip] images, respectively (collected by authors). The colored results using statistic matching (stat-match), histogram matching (HM), joint histogram matching [joint-HM($\beta\alpha$)], and stat-match then joint-HM ($\beta\alpha$) are presented in Fig. 2(e) to 2(h), respectively. Figure 2(e) shows the background (water, leaf) painted with the Tulip's colors; whereas Fig. 2(f) appears oversaturated in colors (i.e., color distortion). The Lotus shown in Fig. 2(g) [or Fig. 2(h)] has the closest colors to the Tulip but its background colors are altered. Two histograms of "Mapped" and "Target" shown Fig. 2(d) indicate the color difference between the mapped image and the target image. This result may imply that a source image can be ideally colorized when its color distribution (e.g., histogram) is similar with that of a target image. Another result of the joint-HM($l\alpha$) is exhibited in Fig. 2(i), where Lotus shows in light and pure colors but its background (water) is distorted. Notice that the following
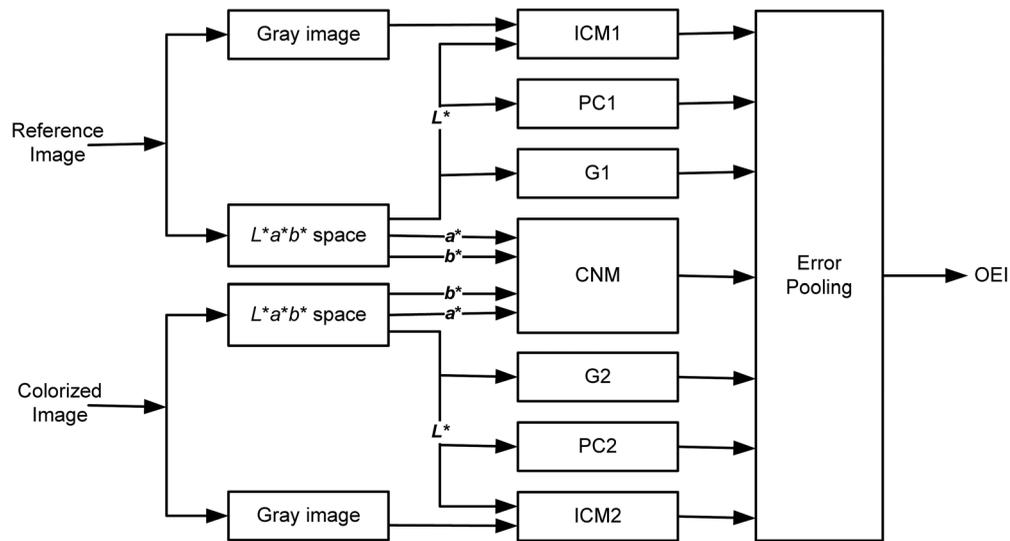
**Fig. 5** Diagram of calculating OEI in $L^*a^*b^*$ space.

experimental results of joint histogram matching were conducted by using joint-HM($\beta\alpha$).

Two sets of multispectral NV images were used in our experiments, which were taken at night-time and referred as to "NV-set 1" and "NV-set 2," respectively. In NV-set 1, three triplets of multispectral images (as shown Figs. 3, 6, and 7 collected at Alcorn State University), color RGB, NIR, and LWIR were colored by using different coloring methods as described in Sec. 4. The three-band input images are shown in Figs. 3, 6, and 7(a) to 7(c), respectively with the image resolutions given in the figure captions. The RGB images and LWIR images were taken by a FLIR SC620 two-in-one camera, which has a LWIR camera (of 640×480 pixel original resolution and 7.5 to 13 μm spectral range) and an integrated visible-band digital camera (2048×1536 pixel original resolution). The NIR images were taken by a FLIR SC6000 camera (640×512 pixel original resolution and 0.9 to 1.7 μm spectral range). Two cameras (SC620 and SC6000) were placed on the same fixture and turned to aim at the same location. The images were captured during sunset and dusk during autumn.

Of course, image registration and fusion as described in Sec. 2.1 were applied to the three band images shown in Figs. 3, 6, and 7, where manual alignment was employed to the RGB image shown in Fig. 7(a) since it is so dark and noisy. To better present the color images (including the daylight RGB images and the colorized NV images), contrast and brightness adjustments (as described in figure captions) were applied. Notice that piece-wise contrast stretching [Eq. (1)] was used for NIR enhancement. As referred in Eq. 7(d), the fused images (shown elsewhere[9]) were obtained using the aDWT algorithm.[24] The channel-based color fusion [defined in Eqs. (7)] was applied to the NIR and LWIR images [shown in Figs. 3, 6, and 7(b) to 7(c)], and the results are illustrated in Figs. 3, 6, and 7(d). The resulted images from two-band color fusion [Figs. 3, 6, and 7(d)] resemble natural colors, which makes scene classification easier. The paved ground appears reddish since it has strong heat radiations (at dusk) and thus causes strong responses in LWIR images. In the color-fusion

images, the trees, buildings and grass can be easily distinguished from ground (parking lots) and sky. The car is clearly identified in Fig. 7(d), where the water area (between ground and trees and shown in cyan color) is certainly noticeable. However, it is hard to realize any water area in the original images [Fig. 7(a) to 7(c)].

All color-mapping methods were applied to the three triplets and their results are presented in Figs. 3, 6, and 7. The source images are the color-fusion images [Figs. 3, 6, and 7(d)], while the target images are the color RGB images [Figs. 3 and 6(a)]. Figure 7(a) cannot be used as a target image as it is too dark and noisy. Figures 3, 6, and 7(e) show the colored images with the statistic-matching method, which are more similar to the daylight pictures in contrast with the color-fusion images. The three results [Figs. 3, 6, and 7(e)] are equivalently good, which means that the statistic matching is reliable. The histogram matching results shown in Figs. 3, 6, and 7(f) are oversaturated, which turns to be more suitable for segmentation-based colorization [see Fig. 8(c) and 8(g)]. The joint histogram matching [i.e., joint-HM($\beta\alpha$)] is illustrated in Figs. 3, 6, and 7(g), where the mapped images are better than the color fusions, but preserve the reddish colors which existed in the source images. Figure 7(i) is also a colored image using joint-HM($\beta\alpha$) by choosing [target = Fig. 6(a)], which appears slightly better than Fig. 7(g) [target = Fig. 3(a)]. The comparable results [shown in Fig. 7(g) and 7(i)] demonstrate that the color-mapping methods can flexibly choose a target image with similar scenery. The "stat-match then joint-HM" (SM-JHM) means that a joint-HM is performed with inputs of [source = the colored image from stat-match, such as Fig. 3(e); target = the RGB image such as Fig. 3(a)]. Their results are presented Figs. 3, 6, and 7(h), which are better than the results from either stat-match or joint-HM. In fact, "stat-match then joint-HM" is overall the best among all color-mapping methods described in Sec. 4. Two examples of LUT-mapping colorization are given in Figs. 3 and 6(i). Figure 3(i) (an ideal case of LUT mapping) shows impressive colors; whereas in Fig. 6(i) the exhibits are noisy and distorted. The noise in the LUT-colorized image may be caused
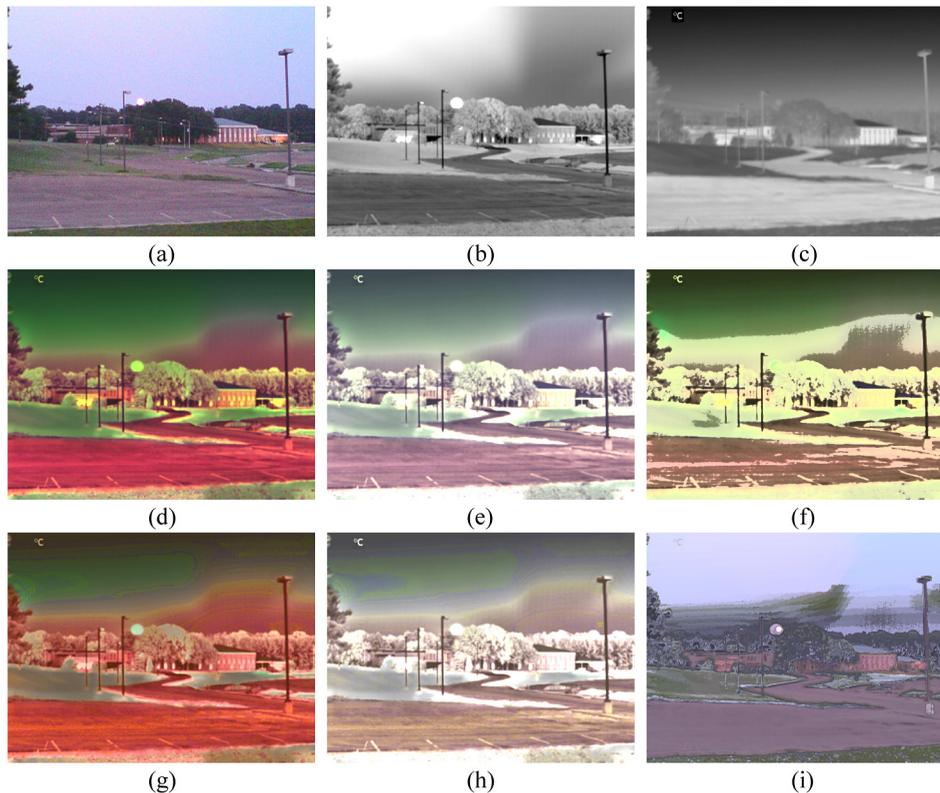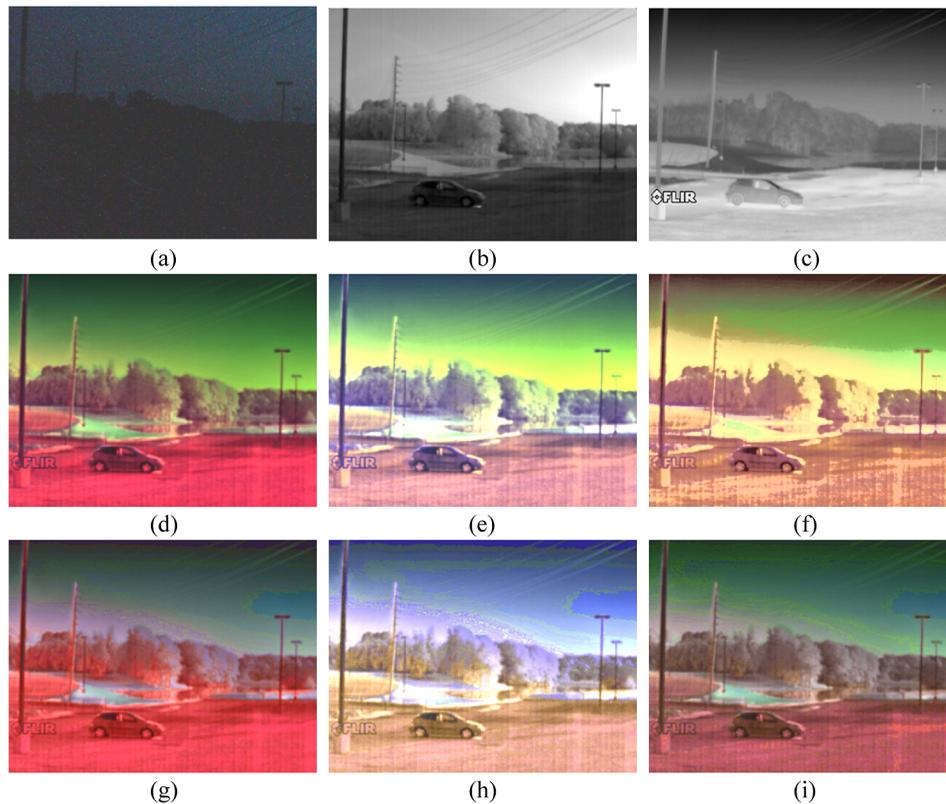
Fig. 6 Night-vision coloring comparison (Case# AT008 in NV-set 1—taken at sunset; 640 × 480 pixels): (a–c) Color RGB, NIR, and LWIR images, respectively; (d)–(f) the colorized images using channel-based color fusion of (NIR ⊕ LWIR), statistic-matching, and histogram-matching, respectively; (g)–(i) the colorized images using joint-HM, statistic-matching then joint-HM, and LUT-mapping [reference = (a)], respectively. The settings in the color-mappings of (e)–(h) are source = (d) and target = (a). Notice that the contrasts of all color images were increased by 10%, and the brightness of (a) and (i) were increased by 10%.

partially by the noisy reference image (taken at dusk) and partially by the pixel-based process during LUT table creation. In fact, many LUT-colored results (about 50% of 30 samples) are similar with Fig. 6(i). Some cases (e.g., Fig. 7) are not directly applicable to LUT colorization since no daylight reference image can be used. When using the LUT established in a different case at daytime (but aiming at different direction at night time), the colored results (not presented in this paper) usually appear worse. For surveillance or navigation applications where cameras move around (i.e., aiming at various directions), the LUT table may be created by using several pairs of night-time/ daytime images taken on the camera moving along a path.[12]

The qualitative evaluations of six methods over three cases (shown in Figs. 3, 6, and 7) in NV-Set 1 are summarized in Table 1. Three categries of quality measurements are used for the qualitative evaluations, which are contrast, details, and colorfulness. The score of each measurement is rated from 3 to 1 to represent low, average, and high quality, respectively. Specifically, a high contrast means an adequate level of brightness and contrast, high details represent high clarity of detailed contents, and high colorfulness preserves more natural colors (i.e., closely resembles the daylight image). Columns 3 to 5 in Table 1 present the rated scores of three categories, where three scores in each cell corresponds to three cases shown in Figs. 3, 6, and 7, respectively. Notice that Fig. 7(i) is another sample of JHM and thus no score is given for LUT (shown × at the bottom row). The averaged scores are listed in the last column, where the quality rank is shown within a pair of curly brackets. It is clear that the quality order of colorization methods from the best to the worst: SM-JHM (stat-match then joint-HM), SM (stat-match), LUT, CBCF (channel-based color fusion), JHM (joint-HM), HM (histogram matching). The same acronyms of six colorization methods are used in Table 2.

The quantitative evaluations using the OEI metric defined in Eq. (19) (refer to Sec. 5.2) are presented in Table 2 (corresponding to Figs. 3 and 6, respectively), where the ranks of metric values (1 for the largest OEI) are given within round parentheses. Keep in mind that the larger the OEI value of a colorized image the better its quality. According to the OEI values in Table 2, the quality order of colorized images in Fig. 3 from the best to the worst are (i), (h), (e), (d), (f), (g); and the quality order in Fig. 6 are (e), (h), (f), (i), (d), (g). To have an overall rank, the sums of the rank numbers in Figs. 3 and 6 are calculated and shown in Table 2. The rank of colorization methods (1 for the best) is given within the curly brackets. The order of colorization methods from the best to the worst: SM-JHM, SM, LUT, HM, CBCF, JHM. For a fair comparison, the averaged scores are recalculated with the rated scores of Figs. 3 and 6, which together with their qualitative ranks (same as Table 1) are exhibited at the far right column of Table 2. Both quantitative and qualitative evaluations support each other for the top three ranks, i.e., SM-JHM, SM, LUT. Statistical matching (SM) may cause

Fig. 7 Night-vision coloring comparison (Case# AT012 in NV-set 1—taken at dusk; 640 × 480 pixels): (a–c) Color RGB, NIR, and LWIR images, respectively; (d)–(f) the colorized images using channel-based color fusion of (NIR ⊕ LWIR), statistic-matching, and histogram-matching, respectively; (g)–(i) the colorized images using joint-HM, statistic-matching then joint-HM, and joint-HM with different settings [source = (d), target = Fig. 6(a)], respectively. The settings in the color-mappings of (e)–(h) are source = (d) and target = Fig. 3(a). Notice that the contrasts of all color images were increased by 10%, and the brightness of (a) were increased by 20%.

color bias[10] when the target (daylight) image is taken at a diffrent location from the source image. The joint-HM (JHM) can prevent (or reduce) color distortion when the source and target are similar in colors (see Fig. 2). On the other hand, JHM may increase color distortion if the source significantly differs from target (refer to the buildings and parking lots in Fig. 3). The JHM is typically combined with statistic-matching (i.e., SM-JHM), which makes the NV colorization better than any individual (either SM or

Table 1 Qualitative evaluations (rated scores) of six methods over three cases (Figs. 3, 6, and 7) in NV-Set #1.(Rate: 1 = high, 2 = average, 3 = low; × = not applicable).

Method	Plot	Contrast	Details	Colorfulness	Avg. {rank}
CBCF	(d)	2, 1, 2	2, 1, 1	3, 3, 3	2.0 {3}
SM	(e)	2, 2, 1	1, 2, 1	2, 1, 1	1.4 {2}
HM	(f)	3, 3, 3	3, 3, 3	2, 3, 3	2.9 {5}
JHM	(g)	3, 1, 2	3, 1, 2	3, 2, 2	2.1 {4}
SM-JHM	(h)	1, 2, 1	1, 2, 2	1, 1, 1	**1.3 {1}**
LUT	(i)	1, 3, ×	2, 3, ×	1, 2, ×	2.0 {3}

JHM; see Fig. 3). Keep in mind the limitation of LUT method, i.e., both source and reference aiming at the same location. Although the performance of CBCF is poor, a realistic color fusion (as the source image) is always expected by other color-mapping methods. The OEI evaluations cannot be applied to Figs. 7 and 8 as no daylight images are available for the required reference images.

In NV-set 2, two pairs of multispectral images, image intensified (II) and LWIR, were analyzed by using color fusion and segmentation-based colorization methods as described in Sec. 4. The two input images are shown in Fig. 8(a) and 8(b) (provided by U.S. Army NVESD) and Fig. 8(e) and 8(f) (provided by the Netherlands TNO[9,13]), respectively. Two input images in NV-set 2 were preregistered. The false-colored images (not shown here) were obtained by assigning II images to blue channels, infrared (IR) images to red channels, and providing averaged II and IR images to green channels. The segmentation was done in $l\alpha\beta$ space through clustering and merging operations. With the segment map (not shown here), the histogram-matching and statistic-matching were performed segment by segment in $l\alpha\beta$ space. After the training process was performed, the source region segments were automatically recognized and associated with proper target color schemes. The final colored images by segmentation-based colorization are shown in Fig. 8(c) and 8(g). From a visual examination, the colored images appear natural, realistic, and colorful. The details of segmentation-based colorization and

Table 2 The OEI (order) values of six methods over two cases (Figs. 3 and 6) in NV-Set #1. (The "qualitative rank" is recalculated with the rated scores of Figs. 3 and 6 in Table 1).

Method	Plot	Fig. 3	Fig. 6	Sum {rank}	Qualitative rank
CBCF	(d)	0.5640 (4)	0.5376 (5)	9 {4}	2.0 {3}
SM	(e)	0.5965 (3)	0.5928 (1)	4 {1}	1.7 {2}
HM	(f)	0.5639 (5)	0.5607 (3)	8 {3}	2.8 {5}
JHM	(g)	0.5536 (6)	0.4825 (6)	12 {5}	2.2 {4}
SM-JHM	(h)	0.6078 (2)	0.5912 (2)	**4 {1}**	**1.3 {1}**
LUT	(i)	0.6417 (1)	0.5575 (4)	5 {2}	2.0 {3}
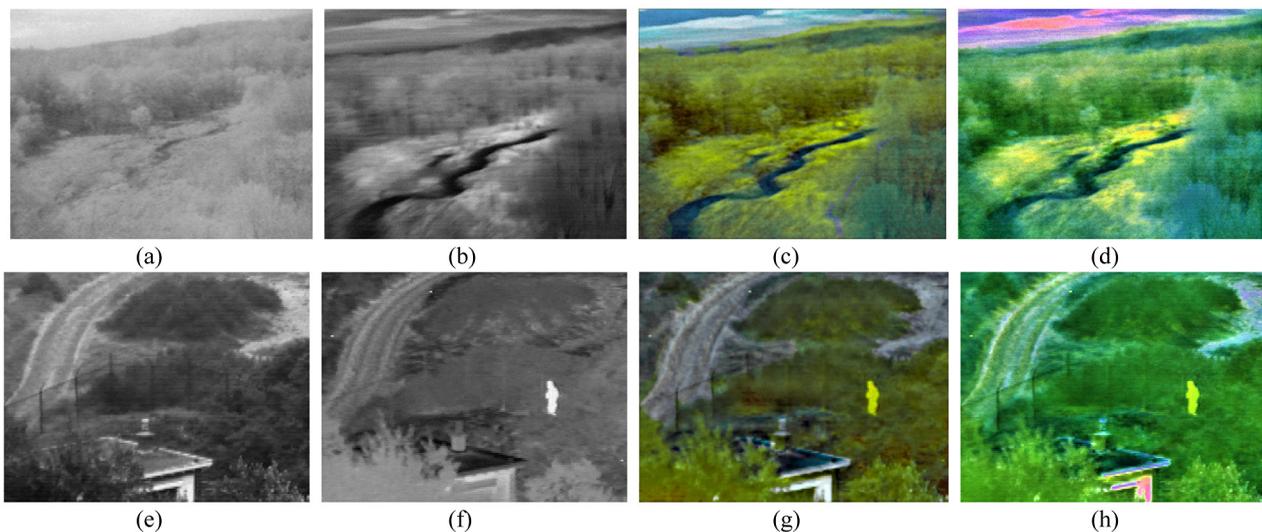
Fig. 8 Night-vision coloring comparison: (a) and (b) and (e) and (f) are two samples of II and LWIR images in NV-set 2; (c) and (g) are the segmentation-based colorizations using histogram-matching, then statistic-matching; (d) and (h) are the channel-based color fusions of (II ⊕ LWIR). Notice that there were no daylight RGB images available in NV-set 2.

experimental results (such as the colorized images with statistic-matching and histogram-matching methods) were presented in Zheng and Essock's paper.[10]

A two-band channel-based color fusion [described in Eqs. (6)] was applied to the II and LWIR images [shown in Fig. 8(a) to 8(b) and 8(e) to 8(f)], and the results are illustrated in Fig. 8(d) and 8(h). The color-fusion results are reasonably 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. However, the processes of channel-based color fusion eliminates the need for segmentation and classification, and also reduces the color transforms. The processing speed is much faster than that of segmentation-based colorization. The LUT-mapping[12,13] method may not be directly applied to NV-set 2 since no daylight reference images are available (see Fig. 8). However, other mapping methods (e.g., joint-HM, stat-match then joint-HM) are applicable here by choosing a target image of similar scenery (such as Fig. 7), but those results are not presented here due to the limited space.

The qualitative (subjective) evaluations of NV coloration are based on casual visual inspections with three general categories. More qualitative measurements, subjective evaluations (by a group of subjects), and statistical analysis will be introduced in the future. The quantitative (objective) evaluations using the OEI require a reference (daylight) image. Thus we will continuously improve the OEI metric by relaxing the requirement of a reference image. We will further investigate color fusion, joint-HM, and SM-JHM methods, and their interactions for speed and visualization, as well as conduct more comprehensive comparisons.

7 Conclusions

A set of qualitative and quantitative comparisons of NV colorization techniques is offered in this paper. We review a channel-based color fusion procedure; explore statistic matching, histogram matching, and LUT-based approaches; introduce new joint histogram matching and stat-match then joint-HM (SM-JHM) methods; and compare them with a segmentation-based colorization using both qualitative and quantitative evaluations. The quantitative evaluations using

the OEI are consistent with the results of qualitative evaluations.

In summary, the segmentation-based colorization generates more colorful and more realistic night-vision images, but it requires heavy computations and thus is time-consuming. The channel-based color fusion gives reasonable coloring results and can be implemented for real-time applications. The LUT method also runs fast and yields a good result when the LUT table is properly established with direction independence. Statistic matching always works reliably and produces a stable colorization. Histogram matching often causes oversaturation and thus is more suitable for segmentation-based coloring. Joint histogram matching usually preserves the existing colors in a source image, which is not ideal when the source image (e.g., a false-colored image) is very different in color from the target image.

Overall, we recommend the "stat-match then joint-HM" (SM-JHM) method that effectively and efficiently provides impressive colorization. SM-JHM also demonstrates the best trade-off between image quality and speed over the methods explored. Keep in mind that the target image (a RGB image taken at daytime) used in all color-mapping methods (except for LUT) can be freely chosen with similar scenery, which may have a different resolution and requires no alignment.

Experimental results with multispectral imagery showed that the colorized images contain comprehensive information and vivid colors. The colorized NV imagery can significantly enhance the NV targeting by human users and will eventually lead to improved performance of remote sensing, night-time perception, and situational awareness.

Acknowledgments

References

1. J. A. Ratches, "Review of Current aided/automatic target acquisition technology for military target acquisition tasks," *Opt. Eng.* **50**(7), 072001 (2011).
2. R. H. Rogers and L. Wood, "The history and status of merging multiple sensor data: an overview," in *Technical Papers 1990, ACSMASPRS Annual Conf. Image Processing and Remote Sensing*, Vol. 4, pp. 352–360 (1990).
3. E. A. Essock et al., "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 (2001).
4. A. Toet et al., "Fusion of visible and thermal imagery improves situational awareness," *Proc. SPIE* **3088**, 177–188 (1997).
5. J. T. Varga, "Evaluation of operator performance using true color and artificial color in natural scene perception," Report ADA363036 (Naval Postgraduate School, Monterey, Calif., 1999).
6. A. M. Waxman et al., "Progress on color night vision: visible/IR fusion, perception and search, and low-light CCD imaging," *Proc. SPIE* **2736**, 96–107 (1996).
7. E. A. Essock et al., "Perceptual ability with real-world nighttime scenes: image intensified, infrared, and fused-color imagery," *Hum. Factors* **41**(3), 438–452 (1999).
8. A. Toet and J. K. IJspeert, "Perceptual evaluation of different image fusion schemes," *Proc. SPIE* **4380**, 436–441 (2001).
9. A. Toet, "Natural colour mapping for multiband night vision imagery," *Inform. Fusion* **4**(3), 155–166 (2003).
10. Y. Zheng and E. A. Essock, "A local-coloring method for night-vision colorization utilizing image analysis and image fusion," *Inform. Fusion* **9**(2), 186–199 (2008).
11. Y. Zheng, "A channel-based color fusion technique using multispectral images for night vision enhancement," *Proc. SPIE* **8135**, 813511 (2011).
12. M. A. Hogervorst and A. Toet, "Method for applying daytime colors to nighttime imagery in realtime," *Proc. SPIE* **6974**, 697403 (2008).
13. A. Toet and M. A. Hogervorst, "Progress in color night vision," *Opt. Eng.* **51**(1), 010901 (2012).
14. Z. Liu et al., "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative survey," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(1), 94–109 (2012).
15. L. Alparone et al., "A global quality measurement of Pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.* **1**(4), 313–317 (2004).
16. L. Wald et al., "Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.* **63**(6), 691–699 (1997).
17. V. Tsagaris and V. Anastassopoulos, "Global measure for assessing image fusion methods," *Opt. Eng.* **45**(2), 026201 (2006).
18. V. Tsagaris, "Objective evaluation of color image fusion methods," *Opt. Eng.* **48**(6), 066201 (2009).
19. Y. Yuan et al., "Objective quality evaluation of visible and infrared color fusion image," *Opt. Eng.* **50**(3), 033202 (2011).
20. D. Malacara, *Color Vision and Colorimetry: Theory and Applications*, SPIE Press, Bellingham, WA. (2002).
21. S. S. Young et al., *Signal Processing and Performance Analysis for Image Systems*, Artech House, Norwood, MA (2008).
22. D. L. G. Hill et al., "Medical image registration," *Phys. Med. Biol.* **46**(3), R1–R45 (2001).
23. Q. Chen et al., "Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition," *IEEE Trans. Pattern Anal.Mach. Intell.* **16**(12), 1156–1168 (1994).
24. Y. Zheng et al., "An advanced DWT fusion algorithm and its optimization by using the metric of image quality index," *Opt. Eng.* **44**(3), 037003 (2005).
25. M. D. Fairchild, *Color Appearance Models*, Addison Wesley Longman, Reading, MA (1998).
26. D. L. Ruderman et al., "Statistics of cone responses to natural images: implications for visual coding," *J. Opt. Soc. Am. A* **15**(8), 2036–2045 (1998).
27. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed., Prentice Hall, Upper Saddle River, New Jersey (2002).
28. W. Dong and Y. Zheng, "An objective evaluation metric for color image fusion," *Proc. SPIE* **8401**, 84010U (2012).
29. M. C. Morrone et al., "Mach bands are phase dependent," *Nature* **324**(6094), 250–253 (1986).
30. P. Kovesi, "Image features from phase congruency," *J. Comp. Vis. Res.* **1**(3), 1–26 (1999).
31. M. C. Morrone and R. Owens, "Feature detection from local energy," *Patt. Recog. Lett.* **6**(5), 303–313 (1987).
32. M. C. Morrone and D. C. Burr, "Feature detection in human vision: a phase-dependent energy model," *Proc. R. Soc. London, Sect. B* **235**(1280), 221–245 (1988).
33. D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.* **93**(26), 429–457 (1946).
34. C. Mancas-Thillou and B. Gosselin, "Character segmentation-by-recognition using log-Gabor filters," in *Proc. Int. Conf. Pattern Recogn.* pp. 901–904, Hong Kong (2006).
35. L. Zhang et al., "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.* **20**(8), 2378–2386 (2011).
36. S. Fischer et al., "Self-invertible 2D log-Gabor wavelets," *Int. J. Comp. Vis.* **75**(2), 231–246 (2007).
37. W. Wang et al., "Design and implementation of log-Gabor filter in fingerprint image enhancement," *Patt. Recog. Lett.* **29**(3), 30–308 (2008).
38. M. Ma et al., "New method to quality evaluation for image fusion using gray relational analysis," *Opt. Eng.* **44**(8), 087010 (2005).
39. L. Henriksson et al., "Representation of cross-frequency spatial phase relationships in human visual cortex," *J. Neurosci.* **29**(45), 14342–14351 (2009).

Yufeng Zheng received his PhD in optical engineering/image processing from the Tianjin University in Tianjin, China, in 1997. He is currently an associate professor at Alcorn State University in Lorman, Mississippi. He serves as a program director of the Computer Networking and Information Technology Program, and as a director of the Pattern Recognition and Image Analysis Lab. He is the principle investigator of three federal research grants in night vision enhancement, and in thermal face recognition. He holds two patents in glaucoma classification and face recognition, and has published a book, five book chapters, and more than 50 peer-reviewed papers. His research interests include pattern recognition, biologically inspired image analysis, biometrics, information fusion, and computer-aided diagnosis. He is a Cisco Certified Network

Professional (CCNP), and a member of SPIE, and IEEE Computer Society & Signal Processing.

Wenjie Dong received an MS in automatic control from Beijing University of Aeronautics and Astronautics in 1996 and a PhD in electrical engineering from the University of California, Riverside, in 2009. He is an assistant professor in the Department of Electrical Engineering at the University of Texas-Pan American in Edinburg, Texas. His research interests include robot control, nonlinear system control, cooperative control of nonlinear systems, mobile sensor networks, and robot vision.

Erik P. Blasch received his BS in mechanical engineering from the Massachusetts Institute of Technology and MS degrees in mechanical engineering health science, and industrial engineering from Georgia Tech. He has an MBA, MSEE, MS economics, MS/PhD psychology (ABD), and a PhD in electrical engineering from Wright State University. He is also a graduate of Air War College. From 2000 to 2010, he was the information fusion evaluation tech lead for the Air Force Research Laboratory (AFRL) Sensors Directorate–COMprehensive Performance Assessment of Sensor Exploitation (COMPASS) Center, adjunct professor with Wright State University, and a reserve officer with Air Office of Scientific Research. From 2010 to 2012 he was an exchange scientist to Defence R&D Canada at Valcartier, Quebec. He is currently with the AFRL Information Directorate in the Information Intelligence Systems and Analysis Division. His research interests include target tracking, information/sensor/image fusion, pattern recognition, and biologically-inspired applications. He is a Fellow of SPIE.