Article

Visible and Near-Infrared Image Acquisition and Fusion for Night Surveillance

Hyuk-Ju Kwon and Sung-Hak Lee *

School of Electronic and Electrical Engineering, Kyungpook National University, 80 Deahakro, Buk-Gu, Daegu 702-701, Korea; olin1223@ee.knu.ac.kr
*   Correspondence: shak2@ee.knu.ac.kr; Tel.: +82-53-950-7216

Abstract: Image fusion combines images with different information to create a single, information-rich image. The process may either involve synthesizing images using multiple exposures of the same scene, such as exposure fusion, or synthesizing images of different wavelength bands, such as visible and near-infrared (NIR) image fusion. NIR images are frequently used in surveillance systems because they are beyond the narrow perceptual range of human vision. In this paper, we propose an infrared image fusion method that combines high and low intensities for use in surveillance systems under low-light conditions. The proposed method utilizes a depth-weighted radiance map based on intensities and details to enhance local contrast and reduce noise and color distortion. The proposed method involves luminance blending, local tone mapping, and color scaling and correction. Each of these stages is processed in the LAB color space to preserve the color attributes of a visible image. The results confirm that the proposed method outperforms conventional methods.

Keywords: infrared; image fusion; tone mapping; luminance blending; night surveillance

1. Introduction

The visible light spectrum spans from 400 to 700 nm, while the near-infrared (NIR) spectrum between 700 and 1100 nm. The human visual system can perceive only visible light. The visibility of a photograph is greatly influenced by the weather conditions, such as haze, smoke, and fog, etc. These bad weather conditions attenuate the contrast and detail of the captured image by scattering light. Nevertheless, the NIR light can be transmitted through haze [1]. The NIR light response offers a higher contrast and richer textural details than that of visible light. Most commercial image sensors can detect both visible and NIR light, rendering them useful for night vision and surveillance applications [2].

Surveillance cameras are used to capture objects under various lighting conditions, such as in dark rooms, in tunnels, in daylight, and in the presence of vehicle headlights. A surveillance camera system consists of an image sensor, a hot mirror filter, and an infrared (IR) light-emitting diode (LED) lamp. The hot mirror is an IR cutoff filter to prevent distortion or undesired artifacts caused by NIR light. The IR lamp serves as an auxiliary component to detect objects that cannot be perceived by human vision. The image sensor uses the hot mirror filter to take a visible image as it blocks NIR light during the day, while also uses IR flashes to illuminate dark regions. Visible images tend to include numerous colors, but a high level of noise is generated by high ISO settings under low-light conditions [3]. In contrast, NIR images demonstrate sharper detail than visible images under low-light conditions. This is due to the use of the IR LED lamp, which tends to make them monotonically reddish.

Surveillance cameras at nighttime often use IR LED lights that do not have a visual effect for security or crime prevention as auxiliary light. In general, since the amount of infrared light emitted by the lamp is fixed, the IR image becomes too bright when capturing close objects, making it very hard to recognize an object. In particular, at night, the hot mirror filter for blocking infrared rays in the daytime is removed, and a complex

visible-near-infrared image is captured. The overlap of visible and strong IR lighting affects the image's color and saturation distortion [4].

Figure 1 demonstrates images taken under various IR LED luminosity conditions. When photographing a distant object, the LED output increases to extend the reach of the light, whereas IR light saturation of a close object becomes relatively higher. When the brightness is reduced to facilitate identification of a nearby object, low-level saturation occurs as in the red arrow area of Figure 1. The quality of an image taken at night faces a number of problems due to low-light environments, excessive or insufficient IR light, and inappropriate exposure times.

Figure 1. Infrared super-saturation or under-saturation effects.

Lately, a wealth of research has investigated how to clearly identify objects in dark spaces under low-light conditions using new LED lighting and camera controls. In order to capture a clear night image, a technology that provides adjustable illumination to different zones of a field of view of an image sensor with a plurality of vertical cavity surface emitting lasers, or to prevent IR oversaturation, it has been proposed to adjust an auto-exposure target in an auto-exposure operation on a sequence of images by obtaining a histogram of at least one of the images [5,6]. However, it is difficult to adjust the camera parameters and the IR LED light adaptively according to the distance of the subject and amount of light in various environments, including severe environmental changes such as moving photography. The hardware configuration of a camera system corresponding to various functions becomes complicated. In addition, increasing just the camera sensitivity makes it difficult to capture an image at a level in which the human vision adapts and perceives a wide range of luminance conditions and sudden lighting changes experienced in everyday life. The proposed method can improve the camera performance for object identification, while preserving color information in various environments by using photography using IR LED pulses of a constant pattern regardless of the lighting environment and the distance of the object.

Visible and IR image fusion methods have been recently found using surveillance cameras. The literature has proposed a large number of fusion methods, including multi-scale transform [7,8], sparse representation [9], neural network [10–12], subspace [13], and saliency methods [14], as well as hybrid models [15–17]. These approaches improve image quality and provide useful information by overcoming the disadvantages of visible and IR images. The multi-scale transform-based methods divide the input image into layers of multiple levels using wavelet, pyramid, and curvelet transforms, as well as an edge-preserving filter. Each layer is processed to analyze the image information and extract the image features and is recombined using a specific rule of each method. Spare representation-based methods convert the source images into a dictionary composed of a few atoms. This method extracts image features using a spare coefficient from the dictionary, fuses the features according to the fusion rule, and reconstructs the fused image. Neural network-based methods are developed based on pulse couple neuron network (PCNN) [18]. PCNN depicts the perception behavior of the human brain to deal with neural information. PCNN-based methods are similar to multi-scale transform methods. Source infrared and visible images are decomposed into low and high frequency subbands and each subband is fused using various methods, such as wavelet, contourlet, and curvelet transforms. The final fused image is reconstructed by an inverse transform. Recently, deep learning-based neural network methods have been introduced in various fields [19–22]. The deep learning-based methods are applied not only to object detection

and classification, but also to image processing such as multi-focus image fusion, remote sensing imagery, panchromatic image fusion as well as infrared and visible image fusion. Subspace-based methods reduce high-dimensional input data to low-dimensional space using principal component analysis, non-negative matrix factorization, and independent component analysis. The low-dimensional space has intrinsic information of input data without redundant information. This method is used for image fusion because it requires less memory and computational time and can improve generalization. Saliency-based methods extract salient object regions from an input image and generate salient layers and fuse the image based on these layers. The salient object region is more significant than the background of the image. These methods maintain the integrity of salient object regions and improve the visual quality of the fused images. Hybrid models combine the advantages of the multiple fusion methods to improve image quality [23].

Sugimura et al. proposed an RGB and NIR image fusion method that relies on blending different exposure times to synthesize a noise- and blur-free image under extremely low-light conditions [24]. The researchers used an adaptive smoothness map depending on the gradient and color correlations of each region. Vanmali et al. proposed a multi-resolution fusion process that is guided by weighting maps for visible and NIR images [1]. These weighting maps are generated based on the local entropy, local contrast, and visibility of each image. Ma et al. proposed a multi-scale fusion method based on a visual saliency map (VSM) and weighted least squares (WLS) optimization models [16]. The Ma et al. method decomposes the visible and NIR image into base and detailed layers. The base layer of each source is then recombined by the VSM. The VSM is used to extract salient structures, regions, and objects from an image. The detailed layer is fused using WLS optimization, which selects more useful details and reduces noise. Li et al. proposed deep-learning architecture, DenseFuse, for visible and IR image fusion [12]. DenseFuse entails encoding and decoding processes. The encoding process combines convolution layers, fusion layers, and dense blocks to obtain useful features from input images. The decoding process is used to reconstruct the fused image.

In the image captured at night using a conventional surveillance camera, as the object gets closer to the camera, over-saturation occurs due to IR light reflections. It is difficult to accurately identify the object's color and shape. To address this problem, it is important to design a complex hardware configuration to measure the amount of surrounding light and the distance between the camera and the object. Another way is to increase the camera sensitivity, but responding in real time to the instantaneous changes in the lighting environment is challenging. Therefore, a capturing that confirms the object shapes regardless of the distance between the object and the camera under various environmental conditions is highly essential.

In this paper, we propose an IR image fusion method combining high- and low-intensities using an IR image acquisition system to enhance image quality under low-illumination conditions. The input high- and low-intensity IR image pairs are taken sequentially for the same scene using an alternating IR lamp. The high-intensity IR image captures the objects with deeper distance depth, whereas the low-intensity IR image captures the overall color information and recognizes the near objects. The proposed method involves luminance blending, local tone mapping, and color scaling and correction (CsC) [25,26]. The luminance blending is used to extract and enhance the detail information of input sources. The luminance blending stage combines the low- and high-intensity IR luminance channels to create a single radiance map using four types of weighting maps: chroma, intensity, depth information based on intensity, and variance differences. Each weighting map is designed to extract the local details of the low- and high-intensity IR images. The local tone mapping is used to improve the visibility of the captured image. A local tone mapping technique, the multi-scale luminance adaptation transform (MLAT) [27], is applied to the radiance map to enhance local contrast. CsC is then used with scaling factors to preserve the original color of the low-intensity IR image and to reduce the noise and color distortion caused by luminance blending and local tone mapping.

2. MLAT

In our previous work, the MLAT was proposed to enhance local contrast and preserve the color of tone-mapped images [27]. The MLAT comprises the multiple luminance adaptation transform (LAT) with different surroundings. The LAT is processed in the LAB color space to decompose the luminance and chrominance channels. The LAT consists of local tone mapping in the luminance channel and color compensation in the chrominance channel. Local tone mapping uses a visual-brightness function [28,29] to enhance local contrast and image rendition depending on the scale of the surroundings. A narrower scope enhances detail, while a wider scope improves tonal rendition. To balance the detail and tonal rendition of an image, a multi-scale method consisting of the weighted sum of several single-scale methods with different surroundings is applied to the LAT. The equations for the MLAT are the following:

$$L_{min} = 0.021 + 0.0185L_{an}^{1.0314}, \tag{1}$$

$$L_{max} = 25.83 + 30.82L_{an}^{0.6753}, \tag{2}$$

$$\gamma_v = 0.444 + 0.045\ln(L_{an} + 0.6034), \tag{3}$$

$$f(x,y) = \left| \frac{L_n(x,y) - L_{min}(x,y)}{L_{max}(x,y) - L_{min}(x,y)} \right|^{\gamma_v(x,y)}, \tag{4}$$

$$\text{LAT} : L_o = I_{gain}f(x,y) + I_{offset}, \tag{5}$$

$$\text{MLAT} = \sum_{n=0}^{N-1} \omega_n \text{LAT}_n, \tag{6}$$

where L_{an} is the normalized adaptation luminance, the Gaussian-blurred image of the input luminance channel L_n. The maximum value of L_{an} is 100. L_{min} and L_{max} are the minimum and maximum luminance for L_{an}, respectively. γ_v is the visual gamma function, $f(x,y)$ is the visual gamma image for pixel position (x,y). L_o is the output luminance channel. I_{gain} and I_{offset} are applied to adjust the intensity range of L_o. MLAT is the multi-scale LAT, and ω is the weight.

The color compensation stage reduces the desaturation effect after local tone mapping. It does so by improving the chrominance gain based on the ratio of the input luminance to the LAT, as follows:

$$C_o = C_i \left(\frac{L_o}{L_i} \right), \tag{7}$$

where C_o is the compensated chrominance channel using the input chrominance channel C_i and the ratio of the output luminance channel L_o to the input luminance L_i. The chrominance channel C represents the a and b channels of the LAB color space.

3. Visible and NIR Image Acquisition and Fusion

The proposed method assumes an indoor scene under ambient light, which contains less outdoor IR light and less ambient IR effect on the camera. This condition considers capturing that is only affected by the emitted IR light by the designed system. First, we designed an image acquisition system to capture IR image pairs using an alternating IR lamp with low and high intensities. The low level is for closer objects, and the high level is for relatively distant objects. The hot mirror filter was removed from the camera to capture IR images. The IR light can affect the visible image, but the low-intensity IR lamp under normal illumination has been adjusted for minimal impact. The proposed acquisition system is illustrated in Figures 2 and 3. In Figure 2, the proposed hardware consists of a wide-band camera that includes the visible and near-infrared band (OmniVision OV5640 CMOS, 1920 (W) × 1080 (H) pixels), two IR lamps (5 V–3 W), an IR lamp control circuit, and a camera trigger circuit. The IR lamps were synchronized with the capturing cycle and alternately light up in two levels, color information and near-field subject information were acquired at the low-level intensity of IR lamps for short distance, and detailed luminance

information of distant subjects were acquired at the high-intensity of IR lamps for long distance, then images were synthesized. As shown in Figure 3, in dark surroundings, the high level of IR light can effectively capture a distant subject, and the low level of IR light can effectively capture a close subject. However, as the surround level increases, the lower level-IR does not significantly affect the image, and information in the visible light band without distortion, including color information, can be captured. Against this backdrop, the proposed system captures two types of IR images: high-intensity IR and low-intensity IR. High- and low-intensity IR images are captured when the alternating level of the IR lamp is high and low, respectively. These IR lamps are synchronized with the capturing period. The IR lamp pulse rate and the image frame rate are synchronized by the camera trigger circuit. As a result, the IR lamps alternately light up in two levels during high–low intensity IR image captures. The camera acquires color information and near-subject contour information at a low level and obtains distant subject contour information at a high level.
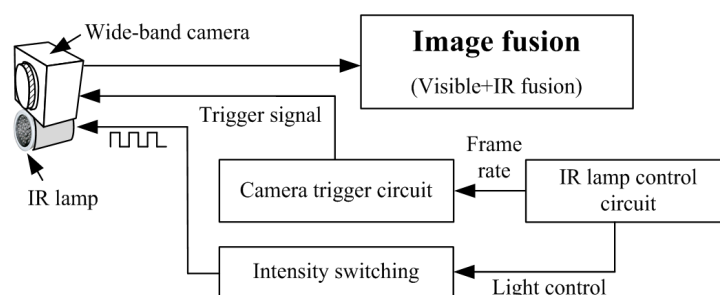
Figure 2. IR illumination modulation diagram for capturing and fusion.
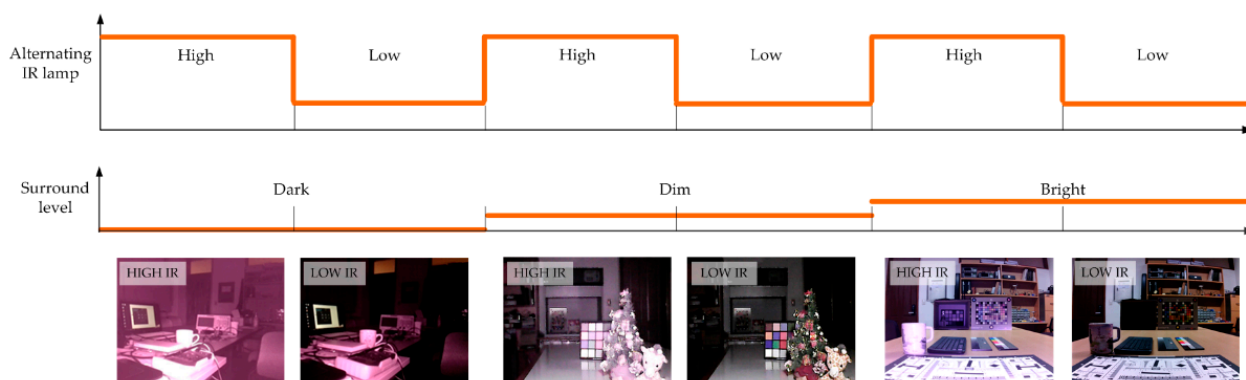

Figure 3. IR image acquisition system.

The images captured by the proposed acquisition system demonstrate the following features. In a dark scene, an image obtained with high-intensity IR is generally shot with a brighter image than that of low-intensity IR, but with a saturation of an adjacent object. Conversely, an image obtained with low-intensity IR has less light, containing details of the saturated region of high-intensity IR. In a bright scene, the low-intensity IR image contains less IR information, and most of the captured image has a visible light band. This enables the color of the scene to be preserved. In our experiments, we captured indoor images to reduce the impact of IR light found in natural light.

In Figure 4, the left-hand image depicts the high-intensity IR image, which includes little color information, whereas the right-hand image depicts the low-intensity IR image. The low-intensity IR image has color information resulting from ambient light. The flower details and colors clearly appear in the low-intensity IR image, but some regions close to the camera in the high-intensity IR image are oversaturated due to the IR light reflections. For the flowerpot under the flowers and the lightbox in the background of the scene, compared

to the low-intensity IR image, the high-intensity IR image demonstrates more details. The red boxes indicate areas with prominent details.
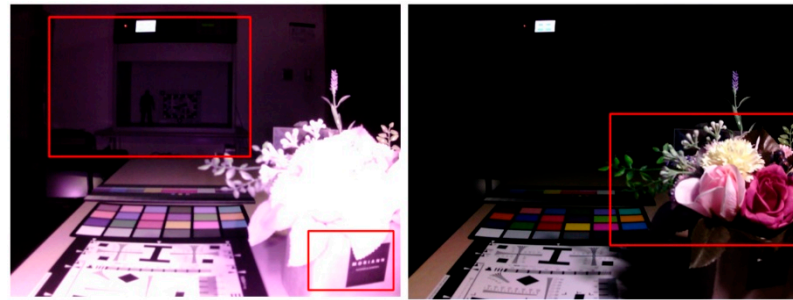
Figure 4. High-intensity IR image (**left**) and low-intensity IR image (**right**) in dim surroundings.

Figure 5 presents a block diagram of the proposed method for the captured images. It shows luminance blending and CsC. Each stage uses the LAB color space to decompose the luminance and chrominance components of the image.
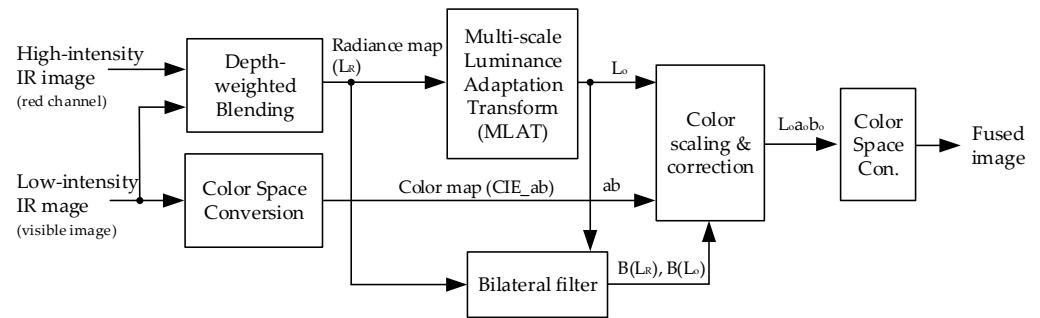
Figure 5. Block diagram of the proposed method.

3.1. Step 1: Depth-Weighted Luminance Blending

Luminance blending combines the input IR image pairs based on weighting maps to produce a radiance map with enhanced details. The weighting maps are calculated in the depth-weighted blending block of Figure 5, incorporating chroma, intensity, intensity difference, and variance difference. Each weighting map shows weights ranging from zero to one. As a weight value moves closer to one, the radiance map becomes increasingly similar to the low-intensity IR image.

The chroma weighting map uses the chrominance channels of the input low-intensity IR image to produce higher weighting values for the objects in a scene. Compared to high-intensity IR images, low-intensity IR images are more colorful and accurate, without dim surroundings. Furthermore, the color components can be used as local details for the objects in a scene. The equation for this process is the following:

$$\omega_c = N\left(\sqrt{a^2 + b^2}\right), \tag{8}$$

where $N(\cdot)$ is a normalization function that converts the range of the weighting map from zero to one, ω_c is the chroma weighting map, and a and b are the chrominance channels of the input low-intensity IR image in the LAB color space.

The intensity weighting map represents the local contrast, using the illumination component of the low-intensity IR image that can be obtained using bilateral filtering. In a simplified computer vision model, the pixel values of an image are the product of illumination and surface reflectance [30,31]. The illumination components can be represented by the weighted sum of adjacent pixels, such as through the use of a bilateral filter. A bilateral filter is an edge-preserving filter that smooths an image while retaining strong edges [32].

Such a filter reduces halo artifacts caused by local tone mapping. The intensity weighting map is calculated as follows:

$$\omega_i = N(B(L_{low})), \tag{9}$$

where ω_i is the intensity weighting map, $B(\cdot)$ is the bilateral filter function, and L_{low} is the luminance channel of the low-intensity IR image in the LAB color space.

The weighting map for the intensity difference represents the local edges between the high- and low-intensity IR images. In the captured images, a high-intensity IR image is brighter than a low-intensity IR image. The over-saturation occurs in the high-intensity IR image when an object is close to the camera. The intensity difference, ω_{id}, between input IR image pairs has higher value when the object is close to the camera. Therefore, the more local edges can be obtained by increasing the weight of the low-intensity IR image when the value of ω_{id} is close to 1:

$$\omega_{id} = N\left(B\left(L_{high}\right) - B(L_{low})\right), \tag{10}$$

where ω_{id} is the intensity difference weighting map, and L_{high} is the red channel of the high-intensity IR image.

The variance difference weighting map was calculated using the difference in the local variance between the high- and low-intensity IR images. The local variance, obtained from pixel-wise processing, represents the local details of each input image. The difference in the local variance represents the local gain at a particular pixel position for each image. The intensity difference has a positive value because the luminance of a high-intensity IR image is always higher than that of a low-intensity IR image. However, the variance difference has both positive and negative values, because local variances have different values for scenes and pixel positions. To set the range for the variance difference weighting map between zero and one, we introduce an additional normalization function with a constraint, as follows:

$$\omega_{vd} = \begin{cases} 0.5\left(\frac{Diff_{var}}{\max(Diff_{var})} + 1\right), & for\ Diff_{var} \geq 0 \\ -0.5\left(\frac{Diff_{var}}{\min(Diff_{var})} - 1\right), & for\ Diff_{var} < 0 \end{cases}, \tag{11}$$

$$Diff_{var} = Var\left(L_{high}\right) - Var(L_{low}), \tag{12}$$

where ω_{vd} is the variance difference weighting map, $Diff_{var}$ is the variance difference, and $Var(\cdot)$ is the local variance based on pixel-wise processing. In Equation (12), the variance difference weighting map ω_{vd} is rearranged based on the variance difference $Diff_{var}$. The constraint is set at 0.5, where $Diff_{var} = 0$, to assign the same weighting value for an input image pair.

Finally, the proposed weighting map was computed from the above four types of weighting maps as follows:

$$\omega_{map} = (\omega_c + \omega_i + \omega_{id} + \omega_{vd})/4. \tag{13}$$

The radiance map can be obtained using the weighted blending of the input images:

$$L_R = \omega_{map}L_{low} + (1 - \omega_{map})E_R L_{high}, \tag{14}$$

$$E_R = mean(L_{low})/mean\left(L_{high}\right), \tag{15}$$

where L_R is the radiance map, E_R is the relative exposure gain, and $mean(\cdot)$ calculates the global mean intensity of the image.

3.2. Step 2: Local Tone Mapping

After luminance blending, we applied local tone mapping using the MLAT to the radiance map, with the following steps:

(1) In Equations (1)–(6), the luminance channel L_n and the normalized adaptation luminance L_{an} are calculated as follows:

$$L_n = 100 \left(\frac{L_R}{\max(L_R)} \right), \tag{16}$$

$$L_{an} = G(L_n, \sigma), \tag{17}$$

where the constant value of 100 is used in normalizing the output range. $G(\cdot)$ is a Gaussian filter with standard deviation σ that controls the local contrast of the LAT.

(2) The MLAT is performed using the weighted sum of three LAT versions. The weighting value ω of the MLAT is $1/3$, and the standard deviation σ is 15, 50, and 250 for the respective versions of the LAT.

3.3. Step 3: Color Scaling and Correction

After MLAT processing, CsC was used to preserve the colors in the tone-mapped image. CsC reduces the noise and the color distortion caused by luminance blending and tone scaling using scaling factors, as follows:

$$a_o = aT_s, \ b_o = bT_s, \tag{18}$$

$$T_s = B(L_o)/B(L_R), \tag{19}$$

where a and b are the chrominance channel of the low-intensity IR image in the LAB color space, and a_o and b_o are the results of using CsC. T_s is the tone scaling factor, which identifies the luminance change rates after tone mapping through bilateral filtering. L_o is the result of using the MLAT.

Finally, the tone-mapped luminance channel L_o and the compensated chrominance channels a_o and b_o are converted back from the LAB color space to the RGB color space.

4. Simulation Results

We compared the results of the proposed method with those of two conventional methods introduced by Vanmali et al. [1] and Li et al. [12]. Vanmali et al. proposed a weighted-map-guided Laplacian–Gaussian pyramid to fuse visible and NIR images. Li et al. proposed deep-learning architecture based on a convolutional neural network and dense blocks for fusing infrared and visible images. In this study, the test images were taken using the proposed image acquisition system. For the two input images, only the red channel of the high-intensity IR image was used for the IR luminance image and the RGB channels of the low-intensity IR image was used for the visible image. We have the parameters of the conventional methods by referring to their original reports. The results of the respective methods are shown in Figures 6–13.

The Vanmali approach does not consider low-light conditions. The algorithm uses the hue–saturation–value (HSV) color space to decouple the luminance and chrominance components. Weight maps are used to combine the luminance channels of the visible and NIR images. The combined luminance channel is converted back from the HSV color space to the RGB color space without color compensation for the luminance variation. As a result, color distortion and color noise appear in regions where there are brightness differences between the input images. In the results for Li, the bright regions of the NIR image display blurring and desaturation effects. This is because the reflected NIR light affects the feature maps in the encoder, and the fused image is reconstructed from the sum of each feature map.
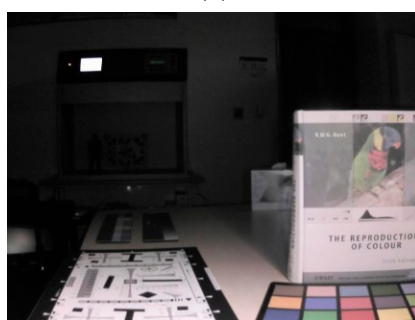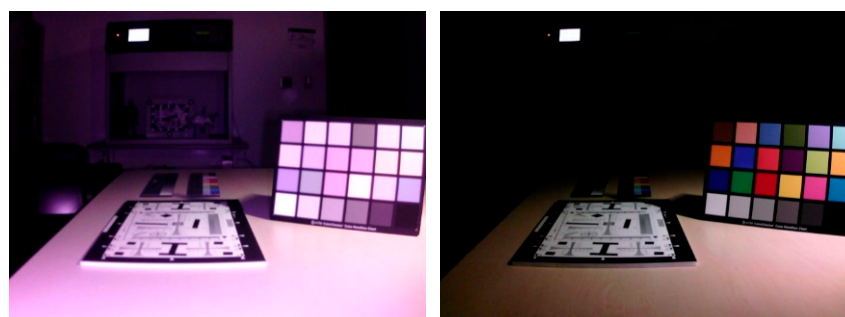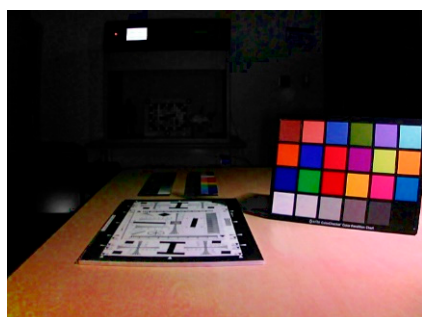
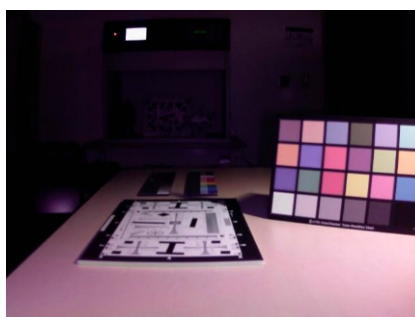**Figure 6.** Input image pair #1 and fused results: (**a**) high-intensity IR (**left**) and low-intensity IR images (**right**); (**b**) Vanmali; (**c**) Li; (**d**) proposed method.
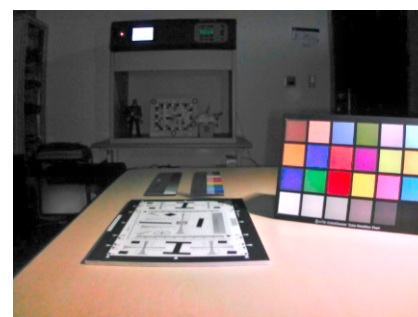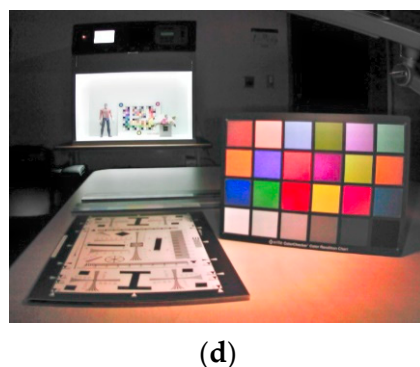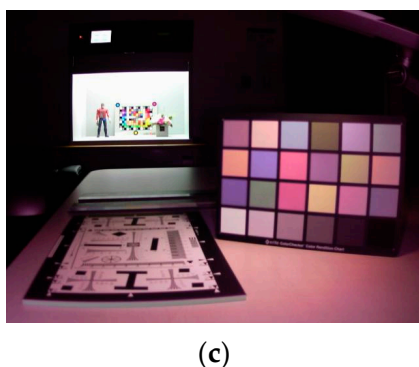

Figure 7. Input image pair #2 and fused results: (**a**) high-intensity IR (**left**) and low-intensity IR images (**right**); (**b**) Vanmali; (**c**) Li; (**d**) proposed method.
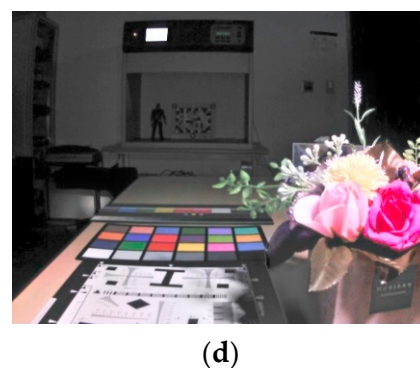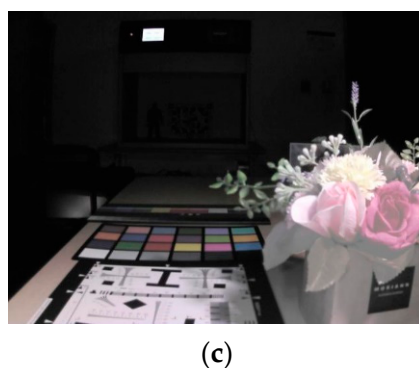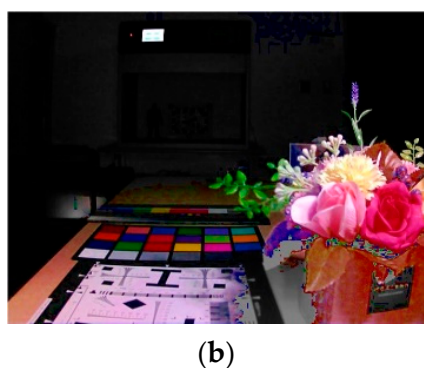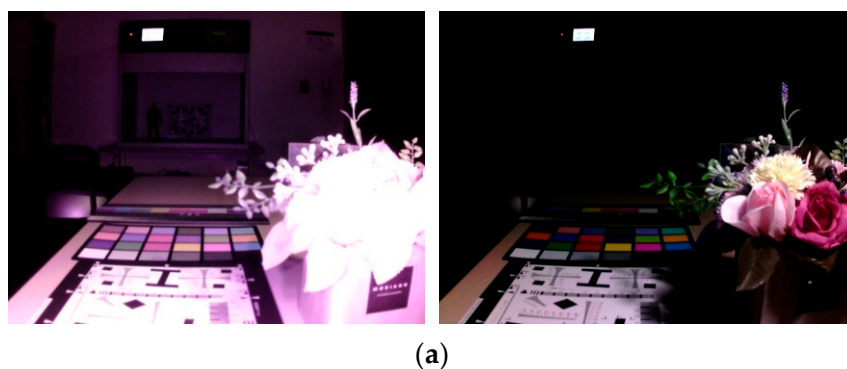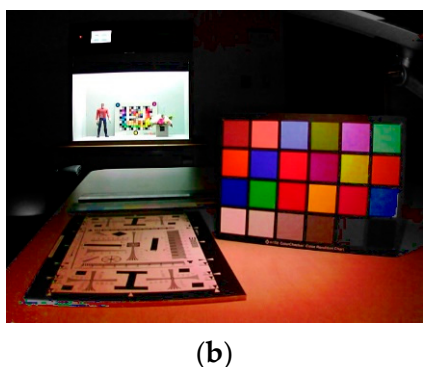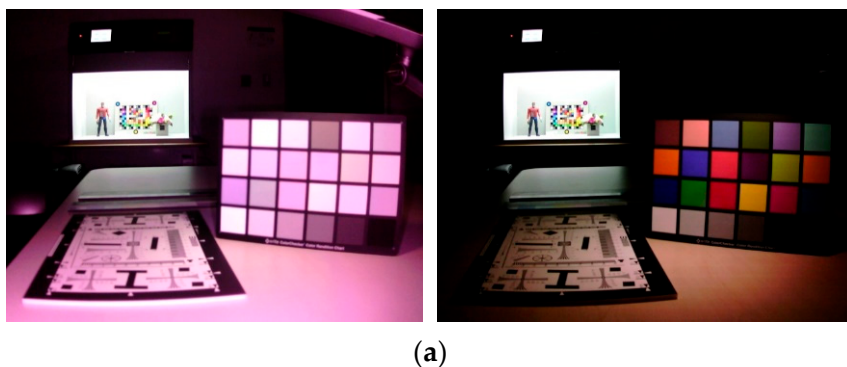
Figure 8. Input image pair #3 and fused results: (**a**) high-intensity IR (**left**) and low-intensity IR images (**right**); (**b**) Vanmali; (**c**) Li; (**d**) proposed method.

Figure 9. Input image pair #4 and fused results: (**a**) high-intensity IR (**left**) and low-intensity IR images (**right**); (**b**) Vanmali; (**c**) Li; (**d**) proposed method.

Figure 10. Input image pair #5 and fused results: (**a**) high-intensity IR (**left**) and low-intensity IR images (**right**); (**b**) Vanmali; (**c**) Li; (**d**) proposed method.
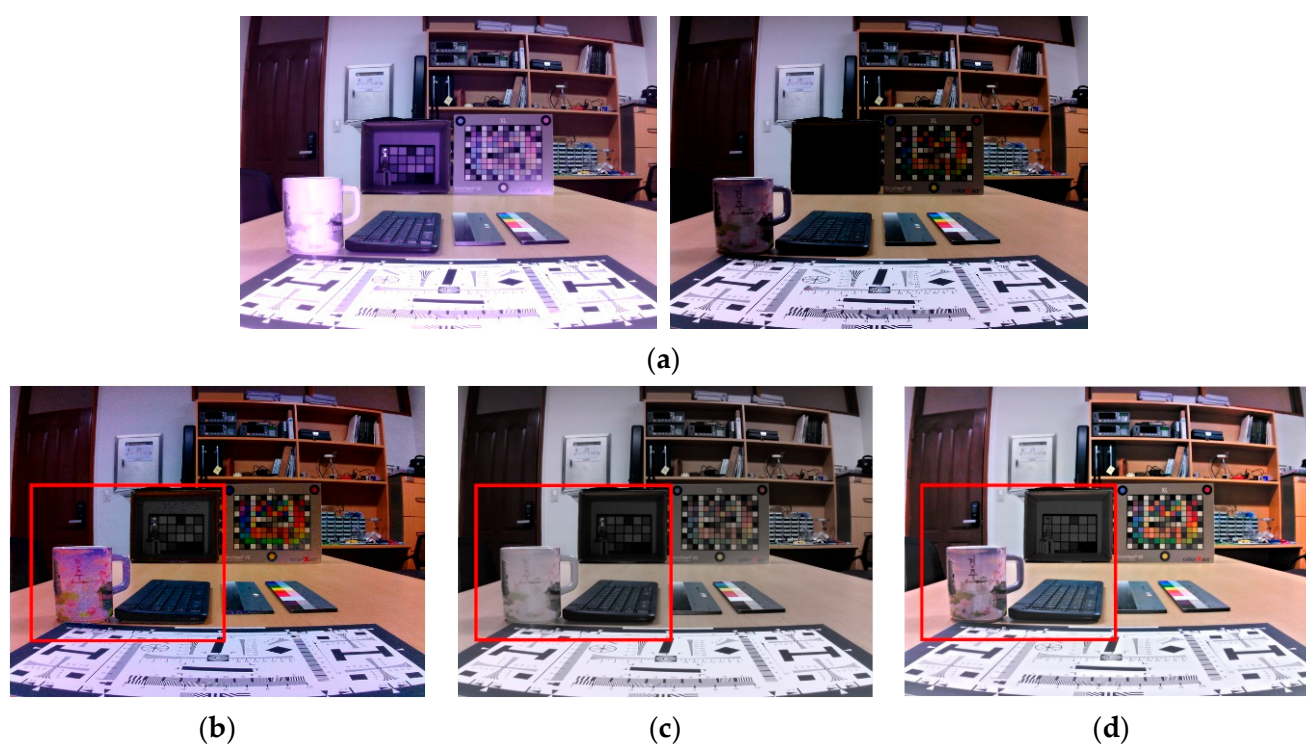

Figure 11. Input image pair #6 and fused results: (**a**) high-intensity IR (**left**) and low-intensity IR images (**right**); (**b**) Vanmali; (**c**) Li; (**d**) proposed method.
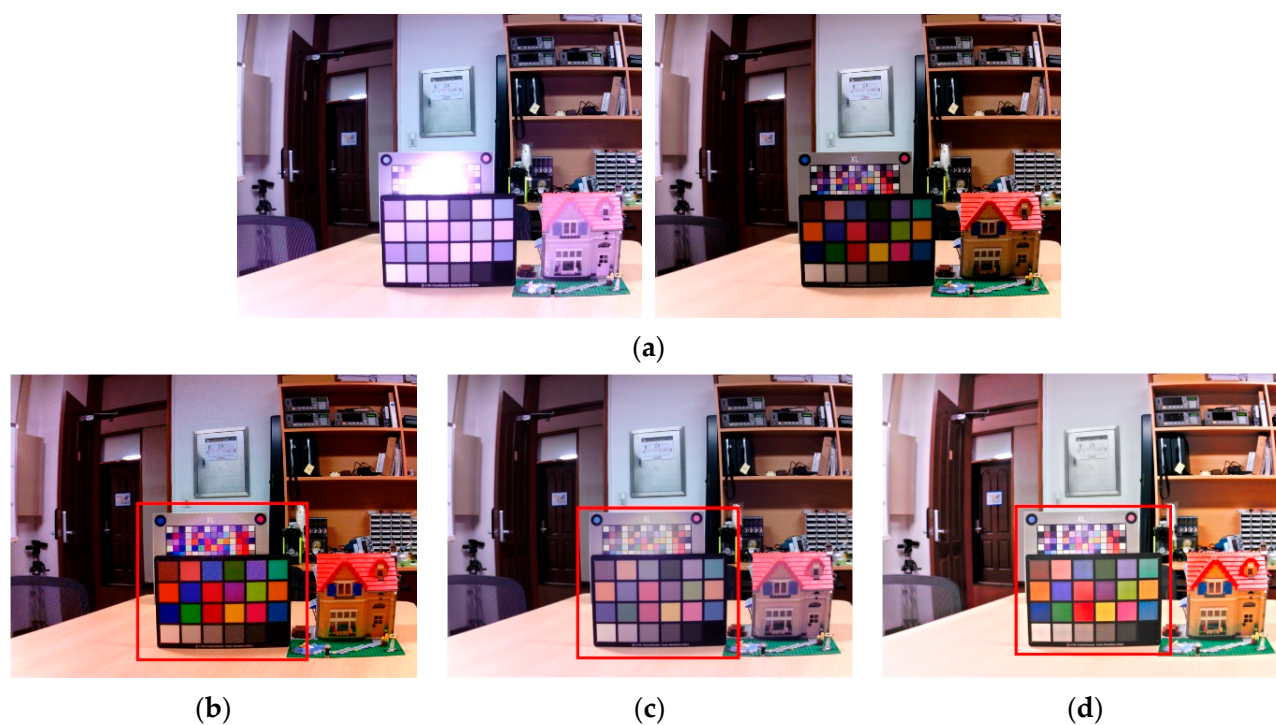
Figure 12. Input image pair #7 and fused results: (**a**) high-intensity IR (**left**) and low-intensity IR images (**right**); (**b**) Vanmali; (**c**) Li; (**d**) proposed method.
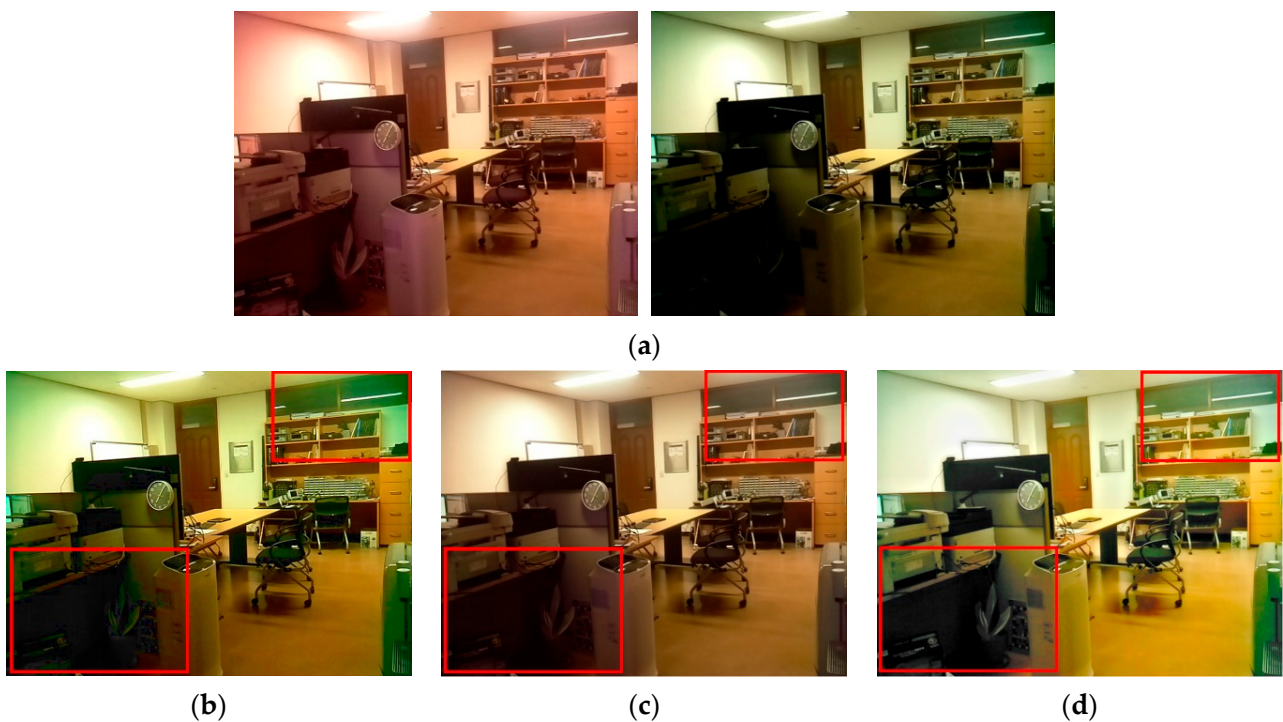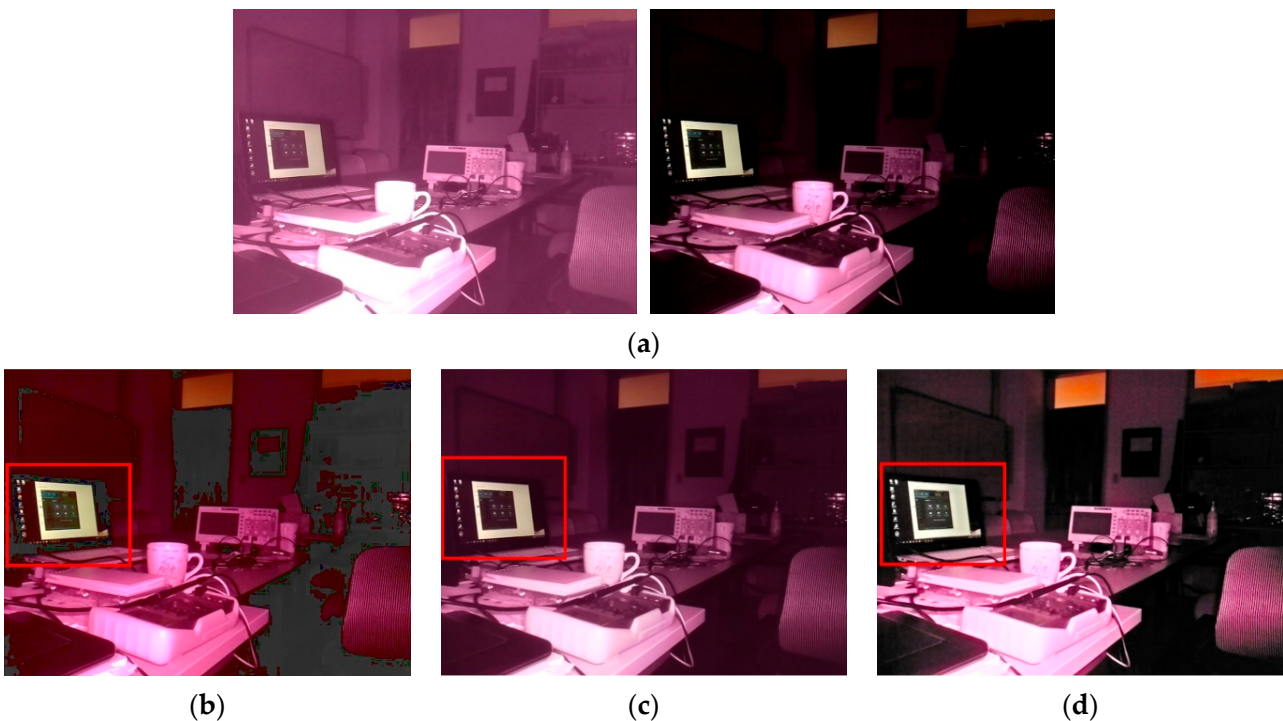

Figure 13. Input image pair #8 and fused results: (**a**) high-intensity IR (**left**) and low-intensity IR images (**right**); (**b**) Vanmali; (**c**) Li; (**d**) proposed method.

In Figure 6, Vanmali's method shows strong noise and color blocking for the overall image, and also over-saturation is shown in the book and the color checker. Li's method shows the desaturation and low local contrast in the results. In Figure 7, Vanmali's method

shows color noise on the desk at right bottom and over-saturation in the color checker, and the Li's method shows white balance shift and de-saturation. In Figure 8, Vanmali's method shows strong color noise in the sharpness test chart and the flower area, whereas Li's method shows a desaturation effect in the flower and flowerpot regions. In Figure 9, Vanmali's method shows color noise and over-saturation in the color checker and the front of the desk and represents low-level noise in the dark region. The proposed method reduces noise and color blocking, providing a higher local contrast than that of other methods.

In Figures 10–13, the red boxes represent the low contrast, color distortion, and color noise regions. In Figure 10, Vanmali's method shows color distortion in the cup, while Li's method shows low contrast and desaturation effects in the same region. In Figure 11, Li's method shows reflected light from the bright IR light. In Figure 12, Vanmali's method shows low-level noise in the red box at left bottom and color shift in the red box at right top. In contrast, the proposed algorithm demonstrates little color distortion and high contrast. Figure 13 is taken in a dark scene. The low-intensity IR image contains little color information because the dark scene uses only an IR lamp, which provides the camera with insufficient color information. However, it offers detailed information about areas that are not displayed owing to saturation in the high-intensity IR image. In the synthesized image, Vanmali's method results in image distortion in the background, and the contrast is low in the monitor area around the red box. In contrast, the proposed algorithm shows a higher contrast than the conventional methods, and the image inside the monitor appears well. These facts confirm that the proposed method has less color distortion and noise and better local contrast than the images produced using the conventional methods.

Subsequently, we performed objective evaluations for the naturalness image quality evaluator (NIQE) [33] and the chroma difference. The NIQE is a supervised method based on a natural scene statistics (NSS) model. The NSS model is trained on statistical features derived from a corpus of natural and undistorted images. The NIQE measures the distance between the NSS model and the extracted features from the fused image. Lower NIQE scores reflect better perceptual quality of an image. The result of NIQE is shown in Figure 14. The results of Figures 7, 10 and 13 are in Figure 14; the proposed method is visually superior, but the proposed method does not have the lowest score compared with the conventional methods. NIQE is based on an "opinion unaware" model that does not require training on databases of human judgments of distorted images. It often does not match human perceived quality. In particular, NIQE is evaluated using a gray scale image. As a result, the results of NIQE are somewhat insensitive to the color noise and the desaturation. To evaluate the color distortion, we added chroma difference using color patches of a color checker.

For 24 color points of the color checker in Figure 11, the relative chroma difference scores were calculated from the Euclidean distance between the result image and the reference image on the chrominance channel in the LAB color space. The reference image was captured with the proposed image acquisition system in a normal environment without the IR lamp. The image in the right top of Figure 15 shows the color checker of the reference image. The chroma difference equation is as follows:
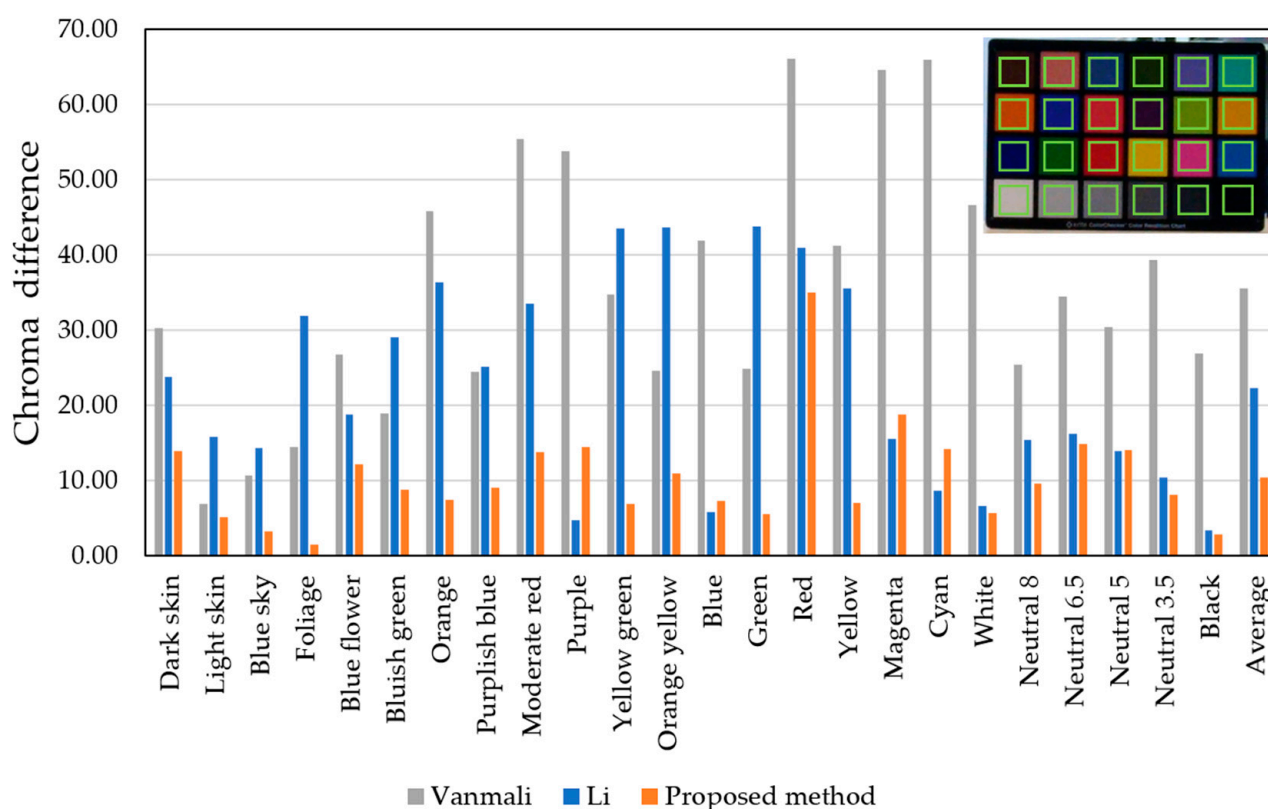
$$\Delta ab = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2}, \tag{20}$$

where Δab is the chroma difference. a and b represent mean values of chrominance channels of each color patch. The subscript 1 represents the reference image and 2 represents the result image of each method.
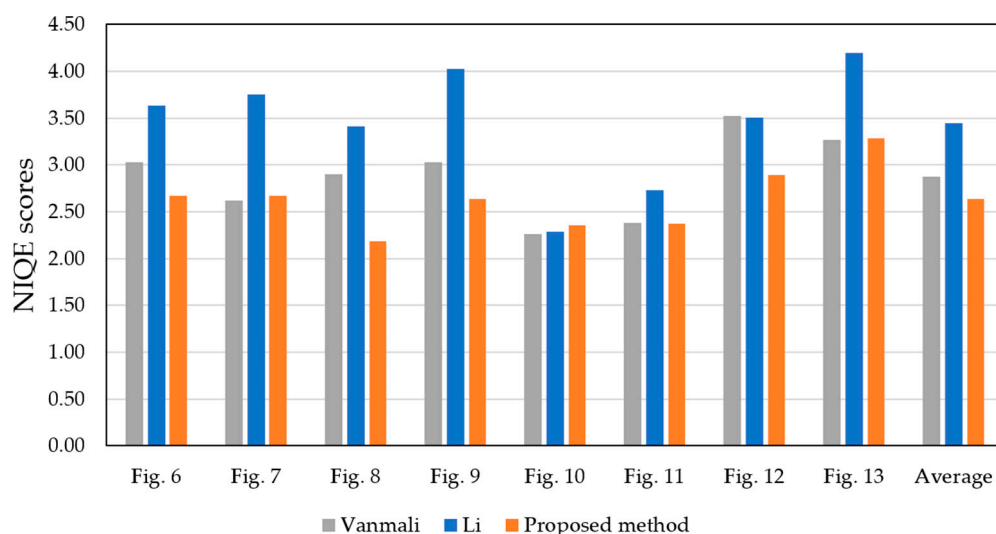
Figure 14. Naturalness image quality evaluator (NIQE) scores for each method (lower scores reflect better image quality).

Figure 15. Chroma difference for each method (lower scores reflect better image quality).

The scores show how well those methods preserve chroma information after image fusion. The accuracy of color reproduction for the reference image was found to be the best in the proposed method on average. Figures 14 and 15 show the scores resulting from using NIQE and chroma difference, respectively.

Table 1 lists the computational time for each method. All methods run on the same environment (a single CPU i7-6700 3.4 GHz, Ram 8 G, GPU Nvidia RTX 2070 super, image resolution 640 (W) × 480 (H), WITHROBOT, Seoul, Korea). The proposed method and Li's method were processed using a Python code and Vanmali's method was run based on

MATLAB. The results show the computational speed of the proposed method is faster than the conventional methods. From the overall assessment, we conclude that the proposed method performs better than the conventional methods.

Table 1. Computational speed for each method.

Image Resolution	Vanmali	Li	Proposed
640 (W) × 480 (H)	0.363 s	0.802 s	0.335 s

5. Conclusions

This paper suggested a high- and low-intensity IR image fusion method to enhance local contrast and reduce noise and color distortion for use in a surveillance system under low-light conditions. Images were taken with an image acquisition system that utilizes an alternating IR lamp. The proposed method consists of luminance blending, local tone mapping, and CsC. Luminance blending combines the high- and low-intensity IR images using weighting maps based on chroma, intensity, the difference in depth of base intensity, and details. In local tone mapping, we used the multi-scale method based on the LAT to balance local details and global rendition. The CsC stage was used to preserve the chrominance components of the low-intensity IR image and reduce the noise and color distortion caused by blending and local tone mapping. The performance of the proposed method was compared to conventional methods. Results showed that the proposed method outperformed other methods in terms of high quality of object details, overall tone rendering, and color visibility compared to the captured images. The proposed image acquisition system assumes an indoor scene containing less outdoor IR light and less ambient IR effect on the cameras. In future research, we plan to capture and fuse images in various environments through hardware improvements. Additionally, we will apply the proposed method to object detection to improve the detection performance in various scenes.

Author Contributions: Conceptualization, S.-H.L.; Investigation, H.-J.K.; Methodology, S.-H.L.; Project administration, S.-H.L.; Software, H.-J.K.; Supervision, S.-H.L.; Writing—original draft, H.-J.K. Both authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that there is no conflict of interests regarding the publication of this paper.

References

1. Vanmali, A.V.; Gadre, V.M. Visible and NIR image fusion using weight-map-guided Laplacian–Gaussian pyramid for improving scene visibility. *Sādhanā* **2017**, *42*, 1063–1082. [CrossRef]
2. Singh, R.; Vatsa, M.; Noore, A. Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognit.* **2008**, *41*, 880–893. [CrossRef]
3. Zhuo, S.; Zhang, X.; Miao, X.; Sim, T. Enhancing low light images using near infrared flash images. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 2537–2540.
4. Etnews. Available online: https://www.etnews.com/200912240153 (accessed on 20 March 2021).
5. Warren, M.E. Compact multi-zone infrared laser illuminator. U.S. Patent 10,244,181 B2, 26 March 2019.
6. Afrooze, S.; Sun, G. Imaging apparatus with scene adaptive auto exposure compensation. U.S. Patent 9,615,032 B2, 4 April 2017.
7. Li, S.; Yang, B.; Hu, J. Performance comparison of different multi-resolution transforms for image fusion. *Inf. Fusion* **2011**, *12*, 74–84. [CrossRef]

8. Pajares, G.; de la Cruz, J.M. A wavelet-based image fusion tutorial. *Pattern Recognit.* **2004**, *37*, 1855–1872. [CrossRef]
9. Li, S.; Yin, H.; Fang, L. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3450–3459. [CrossRef] [PubMed]
10. Xiang, T.; Yan, L.; Gao, R. A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain. *Infrared Phys. Technol.* **2015**, *69*, 53–61. [CrossRef]
11. Kong, W.; Zhang, L.; Lei, Y. Novel fusion method for visible light and infrared images based on NSST-SF-PCNN. *Infrared Phys. Technol.* **2014**, *65*, 103–112. [CrossRef]
12. Li, H.; Wu, X.-J. DenseFuse: A Fusion Approach to Infrared and Visible Images. *IEEE Trans. Image Process.* **2019**, *28*, 2614–2623. [CrossRef] [PubMed]
13. Kong, W.; Lei, Y.; Zhao, H. Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization. *Infrared Phys. Technol.* **2014**, *67*, 161–172. [CrossRef]
14. Zhao, J.; Chen, Y.; Feng, H.; Xu, Z.; Li, Q. Infrared image enhancement through saliency feature analysis based on multi-scale decomposition. *Infrared Phys. Technol.* **2014**, *62*, 86–93. [CrossRef]
15. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [CrossRef]
16. Ma, J.; Zhou, Z.; Wang, B.; Zong, H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys. Technol.* **2017**, *82*, 8–17. [CrossRef]
17. Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* **2019**, *45*, 153–178. [CrossRef]
18. Eckhorn, R.; Reitbock, H.J.; Arndt, M.; Dicke, P. A neural network for feature linking via synchronous activity: Results from cat visual cortex and from simulations. *Can. J. Microbiol.* **1989**, *46*, 759–763.
19. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **2018**, *42*, 158–173. [CrossRef]
20. Li, J.; Huo, H.; Liu, K.; Li, C. Infrared and visible image fusion using dual discriminators generative adversarial networks with Wasserstein distance. *Inf. Sci. (NY)* **2020**, *529*, 28–41. [CrossRef]
21. Ma, J.; Liang, P.; Yu, W.; Chen, C.; Guo, X.; Wu, J.; Jiang, J. Infrared and visible image fusion via detail preserving adversarial learning. *Inf. Fusion* **2020**, *54*, 85–98. [CrossRef]
22. Ma, J.; Xu, H.; Jiang, J.; Mei, X.; Zhang, X.P. DDcGAN: A Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion. *IEEE Trans. Image Process.* **2020**, *29*, 4980–4995. [CrossRef] [PubMed]
23. Son, D.M.; Kwon, H.J.; Lee, S.H. Visible and near-infrared image synthesis using pca fusion of multiscale layers. *Appl. Sci.* **2020**, *10*, 8702. [CrossRef]
24. Sugimura, D.; Mikami, T.; Yamashita, H.; Hamamoto, T. Enhancing Color Images of Extremely Low Light Scenes Based on RGB/NIR Images Acquisition With Different Exposure Times. *IEEE Trans. Image Process.* **2015**, *24*, 3586–3597. [CrossRef]
25. Jung, T.; Kwon, H.-J.; Hahn, J.; Lee, S.-H. Enhanced HDR Image Reproduction Using Gamma-Adaptation-Based Tone Compression and Detail-Preserved Blending. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2020**, *103*, 728–732. [CrossRef]
26. Kwon, H.-J.; Lee, S.-H. CAM-based HDR image reproduction using CA–TC decoupled JCh decomposition. *Signal Process. Image Commun.* **2019**, *70*, 1–13. [CrossRef]
27. Kwon, H.-J.; Lee, S.-H.; Lee, G.-Y.; Sohng, K.-I. Luminance adaptation transform based on brightness functions for LDR image reproduction. *Digit. Signal Process.* **2014**, *30*, 74–85. [CrossRef]
28. Stevens, J.C.; Stevens, S.S. Brightness Function: Effects of Adaptation. *J. Opt. Soc. Am.* **1963**, *53*, 375. [CrossRef]
29. Bartleson, C.J.; Breneman, E.J. Brightness Perception in Complex Fields. *J. Opt. Soc. Am.* **1967**, *57*, 953. [CrossRef] [PubMed]
30. Mantiuk, R.K.; Myszkowski, K.; Seidel, H.-P. High Dynamic Range Imaging. In *Wiley Encyclopedia of Electrical and Electronics Engineering*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015; Volume 41, pp. 1–42.
31. Lee, G.-Y.; Kwon, H.-J.; Lee, S.-H. HDR image reproduction based on visual achromatic response. *Opt. Rev.* **2020**, *27*, 361–374. [CrossRef]
32. Durand, F.; Dorsey, J. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.* **2002**, *21*, 257–266. [CrossRef]
33. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [CrossRef]