Coloring night vision imagery for depth perception

Xiaojing Gu (谷小婧)*, **Shaoyuan Sun (孙韶媛)****, **and Jian'an Fang (方建安)*****

Automation Department, College of Information Science and Technology, Dong Hua University, Shanghai 200051, China

**E-mail: gxj@mail.dhu.edu.cn; **e-mail: shysun@dhu.edu.cn; ***e-mail: jafang@dhu.edu.cn*

Received August 4, 2008

Depth perception for night vision (NV) imagery could largely improve scene comprehension. We present a novel scheme to give fused multi-band NV imagery smoothly natural color appearance as well as depth sense from color. Our approach is based on simulating color cues by varying saturation values of each object in the color NV image, in correspondence with the ratio between the infrared and low-light-level sensor outputs which in practice is the depth feature for same materials. We render the NV image segment-by-segment by taking advantage of image segmentation, dominant color transfer, saturation variation, and image fusion. Experiments have shown that the proposed scheme can achieve satisfying results.

OCIS codes: 100.2980, 040.3780, 040.3060, 350. 2660.

doi: 10.3788/COL20090705.0396.

Night vision (NV) technology enables human beings to operate at night. Infrared (IR) imaging and low-light-level (LLL) imaging are two technologies in NV field, which display either emitted IR radiation or reflected LLL light, and thus provide complementary information of the inspected scene[1,2]. However, the images obtained through these NV imaging systems (or their gray-scale fused representations) are monochromatic and always lack the sense of depth due to their unordinary imaging mechanism. These hinder observers from interpreting the scene well on situational awareness and target detection.

To improve the visual perception on NV imagery, most work has focused on displaying them in color, such as "color space mapping" technology[3−6], which maps multiple spectral bands of imagery into a three-dimensional (3D) color space like RGB or HSV, and "natural color night vision" technology[7−13], which may give the NV imagery a natural color appearance by transferring the characteristics of natural daylight color imagery to the NV imagery based on a color transfer method. Experiments have shown that the "natural color night vision" technology can get a better result than "color space mapping" on visual perception. Though these techniques could give color appearance to NV imagery, very few of them consider the depth perception for NV imagery. Furthermore, Zheng's "natural color night vision" method[13] even weakened the depth perception for NV imagery by rendering it segment-by-segment, since the color transfer procedure employed a set of very different daylight color images from the NV image on structure, although this method can produce colored NV images which appear more like realistic daylight imagery than previous methods. Obviously, beyond natural color appearance, the depth perception for NV imagery could largely improve scene comprehension, which could help in tasks such as understanding the spatial layout of a scene, finding walkable areas in a scene, detecting objects, etc.

Enhancing depth sense for NV imagery is a challenging problem, since it refers to numerous visual cues for depth perception and human beings combine these cues to understand the 3D structure of the world[14]. The reason why NV imagery lacks the depth sense is that there are not sufficient clear depth cues on it. Enhancing details of NV imagery by extending gray range or contrast between different gray levels can generate an indirect benefit on depth perception, since it also makes depth cues such as texture variations and gradients become clear. In fact, it is depth cues but not details that give the observers direct and reliable depth sense on imagery. Therefore, instead of making the existing cues clearer, we consider adding extra cues on the NV imagery for depth perception improvement, which is consistent with the existing ones. Particularly, regarding color NV imagery, it is possible to enhance the depth sense by adding color cues on the chromatic channels, which can express useful and important depth information that is absent in the original monochromatic NV image. Notice that color cues are not equal to color or natural color appearance, they are the variety of color contrasts in distance based on the natural color appearance.

Human beings use numerous visual cues to perceive depth. Such cues are typically classified into binocular cues that require input from both eyes and monocular cues that require the input from just one eye. For single monocular images (most NV images are monocular), depth is judged only from some of monocular cues, such as texture variations and gradients, occlusion, known object size, defocus, color, etc[15]. Among those, color cues are only in color image and the other cues are achromatic that both in color image and gray-scale image.

Color cues (also known as aerial perspective cues) will change due to light scattering by the atmosphere, objects that are a great distance away have lower color saturation and appear hazier[16]. The foreground has high color saturation, while the background has low color saturation. Observers could perceive objects differing only in their color saturation with a background appearing to be at different depths. Color cues are widely used by artists to express space in painting. Unfortunately, images obtained through the NV imaging systems always lack of the depth cues, either color cues or achromatic cues, since they are monochromatic and with low contrast, blurry details, and narrow gray range. It is even worse for a gray-scale fused representation.

We present a novel depth-coloring method in this letter. Our approach is based on simulating color cues by varying saturation values of each object in color NV image, in correspondence with the ratio between the IR and LLL sensor outputs. In practice the ratio is the depth feature for same materials (see Fig. 1). We render the NV image segment-by-segment by taking advantage of image segmentation, dominant color transfer, saturation variation, and image fusion. Dominant color is the mean values of hue and saturation of object in daylight. A look-up-table (LUT) is employed to store dominant colors and reference depth features, which are grouped by their sense contents such as thermal targets, green plants, earth/roads, buildings, sky, etc. Our approach gives NV imagery smoothly natural color appearance as well as sense of depth, and thus improve the situational awareness and target detection.

Our method is schematically shown in Fig. 2. The major points for this method are as follows. 1) The NV image is rendered segment-by-segment. A mean-shift segmentation[17] is applied on the false color image to obtain the image segments by its color properties (corresponding to its scene contents). 2) Depth feature of each segment is obtained by computing the ratio between the average output intensities of IR and LLL sensors (i.e., the ratio between the mean values of R channel and G channel on each false color image segment in RGB color space), since it is reflected by the low-frequency information in the sensor outputs. 3) The color transfer procedure is to replace the values of hue and saturation of every segment with those of corresponding dominant colors, and the values of luminance are replaced by gray-scale fused image. Dominant color is the mean values of hue and saturation of object in daylight. A set of natural color images have been analyzed so that a sequence of dominant colors can be computed. 4) In order to simulate the color cues for depth perception, the saturation values of each segment vary proportionally with the ratio between its depth feature and reference depth feature. The reference depth feature is the depth feature for a known object in middle distance (about 40 − 60 m), which has been computed from a set of multi-band NV images. 5) As color source properties, dominant colors and reference depth features are stored in a LUT and grouped by their sense contents. 6) The mapping between gray-scale segments and dominant colors can

Fig. 1. Two false color images formed by assigning IR image to R channel, LLL image to G channel, and zero value to B channel. The ratio between the IR and LLL sensor outputs approximately reflects the depth relationships for same materials. Regarding the same kind plants, near ones appear redder (higher color saturation) than the distant ones (i.e., the ratio is higher).

be done automatically by a pattern recognition process based on texture and radiation/reflection similarity measure, which is still under development. Thereby the pattern recognition portion of this algorithm is carried out manually in the following experiments in this letter.

Mean shift segmentation[17] is applied on the false color image to obtain image segments. Mean shift is a non-parametric feature space analysis technique and one of the most statistic methods for image segmentation, which actually is an iterative procedure for locating stationary points of a density function given discrete data sampled from that function. Mean shift segmentation contains mean shift filtering procedure and concatenation procedure.

To obtain a meaningful segmentation, the false color image is pre-transformed from RGB color space to $L^*u^*v^*$ color space, in which perceived color differences

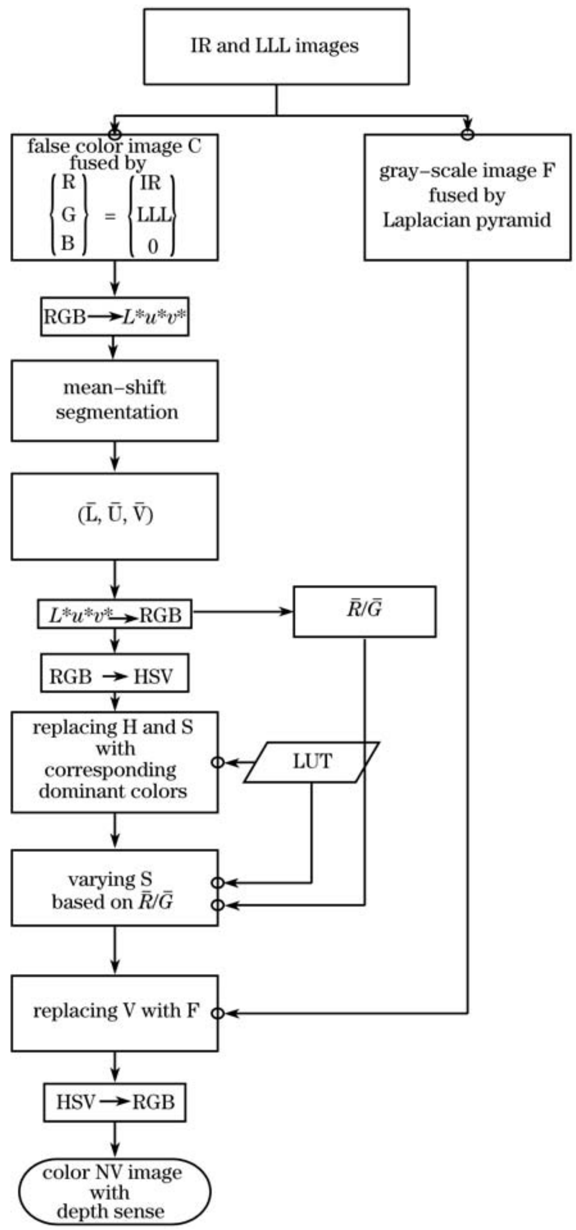

Fig. 2. Diagram of the proposed depth-coloring method. The small circles denote the inputs expected by that procedure.

correspond to Euclidean distances. In the mean shift filtering procedure, the general iteration expression is

$$y_{j+1} = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}, \qquad j = 1, 2, \cdots, \tag{1}$$

where x is the center of the current position of the kernel $g(x)$, y_{j+1} is the next iteration point, h is the kernel bandwidth which is positive, and n is the number of sampling data. The iteration is stopped when $\|y_{j+1} - y_j\| < \varepsilon$. For image filtering, a multivariate kernel is defined as

$$K_{h_\mathrm{s},h_\mathrm{r}}(x) = \frac{C}{h_\mathrm{s}^2 h_\mathrm{r}^p} k\left(\left\|\frac{x^\mathrm{s}}{h_\mathrm{s}}\right\|^2\right) k\left(\left\|\frac{x^\mathrm{r}}{h_\mathrm{r}}\right\|^2\right), \tag{2}$$

where x^s is the spatial part, x^r is the range part of a feature vector, $k(x)$ is the common profile used in both domains, h_s and h_r are the employed kernel bandwidths, p is the dimension of pixels, p =1 for gray-level images and p =3 for color images, and C is the corresponding normalization constant. In practice, an Epanechnikov or a normal kernel always provides satisfactory performance.

In the following concatenation procedure, the clusters are delineated in the joint domain by grouping together all convergence points which are closer than h_s in the spatial domain and h_r in the range domain. In order to compute the depth feature of each segment, the mean values of each channel in $L^*u^*v^*$ color space $\bar{L}$, $\bar{U}$, and $\bar{V}$, are computed with regard to the mean shift filtered false color image first, and then transformed to RGB color space to get $\bar{R}$ and $\bar{G}$ (B channel should always be zero in this algorithm), where $\bar{R}$ and $\bar{G}$ are the average output intensities of IR and LLL sensors. Depth feature of the ith segment D_i is defined as

$$D_i = \bar{R}_i \left(\bar{G}_i\right)^{-1}. \tag{3}$$

In order to simulate the color cues for depth perception, the saturation value of the ith segment S_i is changed after the dominant color transfer procedure to a new saturation value S_i^*,

$$S_i^* \propto K^{-1} S_i, \tag{4}$$

$$K = D_i \left(D_{\mathrm{ref}_i}\right)^{-1}, \tag{5}$$

where $D_{\mathrm{ref}i}$ is the reference depth feature corresponding to D_i and is stored in LUT. If $K \approx 1$, the scene content of this segment is in a middle distance (about $40 - 60$ m) like the reference scene, it should have the same saturation as the reference in order to appear at the middle ground in terms of visual perception. If $K < 1$, the scene content of this segment is in the background, it should have a low color saturation, and the fewer this value is, the lower the saturation is, and the further away to the observers it appears to be. If $K > 1$, the scene content of this segment is in the foreground, it should have a high color saturation, and the greater this value is, the higher the saturation is, and the nearer to the observers it appears to be. In practice, S^* is proportional to KS or $K^{-1}S$ depending on NV imaging condition, such as induced band range and direction of shadow, which should be determined before using the scheme under a new imaging condition.

In our depth-coloring method, hues and saturations are replaced by those of corresponding dominant colors, and the saturations vary corresponding to the ratios between depth features and reference depth features. Dominant colors and reference depth features are grouped by their sense contents, as can be seen from Table 1.

Table 1. LUT for Dominant Colors (H,S) and Reference Depth Features ($\bar{R}/\bar{G}$)

Sense Content	H	S	$\bar{R}/\bar{G}$
Thermal Targets	0.1667	1.0000	5.556
Green Plants	0.3110	0.4512	1.242
Earth/Roads	0.2669	0.3063	0.616
Buildings/Rocks	0.6389	0.0392	0.852
Sky	0.5333	0.0236	0.682
Others	0.2669	0.3063	1.242

In the follow experiments, the false color images are formed by assigning IR images to R channel, LLL images to G channel, and zero value to B channel. Gray-scale fused images are computed by normal Laplacian pyramid. The segmented images are obtained with the mean shift segmentation algorithm by employing a normal kernel as $k(x)$ and setting the parameters C, p, h_s, and h_r to be 1, 3, 16, and 20, respectively. In the saturation varying procedure, S^* is proportional to KS.

We analyzed a pair of IR and LLL NV images taken outdoors, as shown in Figs. 3(a) and (b). The gray-scale fused image was computed by normal Laplacian pyramid, as shown in Fig. 3(c). The false color image is shown in Fig. 3(d). The segmented image with averaged R channel and G channel (Fig. 3(e)) was obtained based on mean shift procedure. Dominant color rendered images without saturation varying and its final result which was produced by replacing the values of luminance with gray-scale fused images are shown in Figs. 3(f) and (h). Dominant color rendered images with a saturation varying procedure and its final result are shown in Figs. 3(g) and (i).

From the visual examination of the resultant images, it can be seen that the rendered images with a saturation varying procedure yield a stronger depth sense than the dominant color transfer alone, especially in the spatial layout of the green plants. Meanwhile, the dominant color transfer technology gives NV imagery a smoothly natural color appearance, since the uniform colors filter the noisy points on image visually. From the comparison given in Figs. 4(a) and (b), it is obvious that the color NV image rendered by Zheng's method is not as natural and rich of depth sense as the image product by our depth-coloring method, since in Zheng's method the depth cues are disturbed by the color characteristics from a set of natural daylight images with very different structures.

In conclusion, a novel "depth-coloring" method is presented in this letter. It gives fused multi-band NV imagery smoothly natural color appearance as well as depth sense by rendering it with dominant colors segment-by-

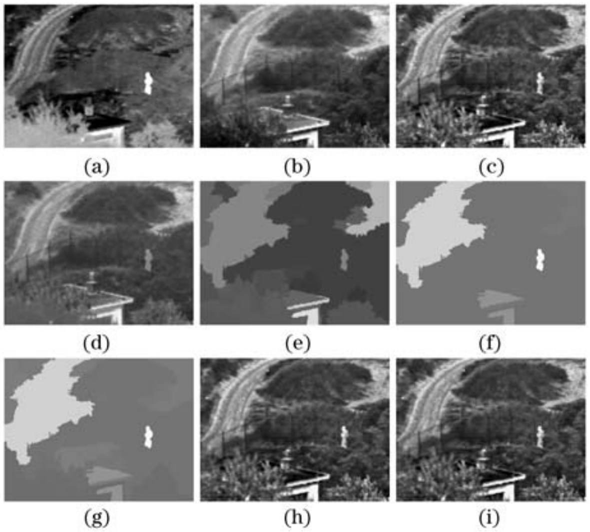

Fig. 3. (a) IR and (b) LLL NV images; (c) gray-scale fused image with (a) and (b); (d) false color image by using (a) and (b); (e) segmented image with averaged R channel and G channel; (f) colored without saturation varying; (g) colored with saturation varying; (h) final colored image corresponding to (f) with "V" replaced by (c); (i) final colored image corresponding to (g) with "V" replaced by (c).

Fig. 4. NV image pair in Fig. 1 rendered by (a) Zheng's "natural color night vision" method and (b) the proposed depth-coloring method.

segment. We add visual cues for depth perception based on varying saturation values of each object's dominant color. Experimental results indicate that the scheme can achieve satisfying results. We plan to develop a classifier that can recognize the segments in NV images so that the mapping between NV segments and color schemes can be done automatically. The proposed method may be applied to the single-band NV images based on image analysis in the future.

We thank Dr. Lex Toet of TNO Human Factors for providing the NV images. This work was supported by the National Natural Science Foundation of China (No. 60502042) and the Shanghai Rising-Star Program (No. 06QA14003).

References

1. B. Guo, Q. Zhang, and Y. Hou, Chin. Opt. Lett. **6,** 338 (2008).
2. C. Liu, Z. Jing, G. Xiao, and B. Yang, Chin. Opt. Lett. **5,** 274 (2007).
3. A. Toet and J. Walraven, Opt. Eng. **35,** 650 (1996).
4. A. M. Waxman, A. N. Gove, D. A. Fay, J. P. Racamato, J. E. Carrick, M. C. Seibert, and E. D. Savoye, Neural Netw. **10,** 1 (1997).
5. D. A. Fay, A. M. Waxman, M. Aguilar, D. B. Ireland, J. P. Racamato, W. D. Ross, W. W. Streilein, and M. I. Braun, Proc. SPIE **4023,** 106 (2000).
6. G. Ni, M. Xiao, Q. Qin, and G. Huang, Acta Opt. Sin. (in Chinese) **27,** 2101 (2007).
7. A. Toet, Information Fusion **4,** 155 (2003).
8. A. Toet, Displays **26,** 15 (2005).
9. S. Sun, Z. Jing, G. Liu, and Z. Li, Chin. Opt. Lett. **3,** 448 (2005).
10. V. Tsagiris and V. Anastassopoulos, Displays **26,** 191 (2005).
11. S. Shi, L. Wang, W. Jin, and Y. Zhao, Proc. SPIE **6623,** 66230B (2008).
12. M. A. Hogervorst and A. Toet, in *Proceedings of 10th International Conference on Information Fusion* 1 (2007).
13. Y. Zheng and E. A. Essock, Information Fusion **9,** 186 (2008).
14. A. E. Welchman, A. Deubelius, V. Conrad, H. H. Bülthoff, and Z. Kourtzi, Nature Neurosci. **8,** 820 (2005).
15. A. Saxena, S. H. Chung, and A. Y. Ng, Int. J. Comput. Vis. **76,** 53 (2008).
16. R. P. O'Shea, S. G. Blackburn, and H. Ono, Vision Res. **34,** 1595 (1994).
17. D. Comaniciu and P. Meer, IEEE Trans. Pattern Anal. Machine Intell. **24,** 603 (2002).