Objective evaluation of target detectability in night vision color fusion images

Yihui Yuan (袁轶慧)*, Junju Zhang (张俊举), Benkang Chang (常本康), Hui Xu (许 辉), and Yiyong Han (韩亦勇)

School of Electronic Engineering and Optoelectronic Technology, Nanjing University of Science and Technology, Nanjing 210094, China

**Corresponding author: xiaoyuan1101yyh@yahoo.cn*

Received July 7, 2010; accepted August 16; posted online January 1, 2011

An evaluation for objectively assessing the target detectability in night vision color fusion images is proposed. On the assumption that target detectability could be modeled as the perceptual color variation between the target and its optimal sensitive background region, we propose an objective target detectability metric in CIELAB color space defined by four color information features: target luminance, region perceptual luminance variation in human vision system, region hue difference, and region chroma difference. Experimental results show that this proposed metric is perceptually meaningful because it corresponds well with subjective evaluation.

OCIS codes: 110.3000, 330.1720, 350.2660.
doi: 10.3788/COL201109.011101.

Night vision color through fusion of images derived from visible and infrared detectors has been increasingly used in applications. A great deal of color image fusion algorithms have been developed[1−5]; thus, the demand for objective evaluation of color fusion performance has risen. The evaluation of the performance of image fusion reported for grayscale images usually employs edge or mutual information[6−7]. However, these methods cannot be extended into color image. A number of studies on perceptual evaluation have been conducted[8−10]. However, subjective evaluation is time-consuming and expensive. Objective evaluation on color fusion image remains almost untouched, leading to an increasing need for objective measurements. Shi *et al.* defined the subband contrast index over a sub-band to measure details and target detectability[11]. Tsagaris *et al.* proposed a measurement to objectively assess the performance of color fusion image by considering the amount of common information, as well as the distribution of color information in the final image[12].

Since color fusion is mainly applied to improve target recognition in night vision, we consider target detectability as one of the most important factors to measure the quality of color fusion images. In this letter, we propose an objective evaluation to predict the target detectability, in which the detectability metric is defined based on the perceptual color variation between target and background region. We consider CIELAB as a suitable color space to measure color characteristics. In the CIELAB system[13], a color is represented by three channels. The first channel is the luminance L^*, which defines the lightness or darkness of a color. Variables a* and b* are chromatic yellow-blue and red-green opponent channels, which tell us about the hues.

The transformation equations from the CIEXYZ system to the CIELAB space are addressed in Ref. [13]

Two important magnitudes, chroma C^* and hue h^*, are defined from the coordinates a^* and b^* as follows:

$$\begin{array}{l} C^* = (a^{*2} + b^{*2})^{1/2} \\ h^* = \arctan\ (b^*/a^*) \end{array} . \quad (1)$$

According to the research of human vision system, human eyes are more sensitive to local contrast[14−15]. The local region with large local contrast will attract more attention[16].

During the procedure of observing a color image, the viewing field of observation can be divided into four areas: "stimulus," "proximal field," "background," and "surround" as shown in Fig. 1.

"Stimulus" and its neighborhood within 2° viewing angle named "proximal field" are the areas between which human vision is most sensitive to color difference. If a given target is chosen in the color fusion image, we regard the target as a stimulus. Its proximal field should be selected as background area. "Stimulus" and "proximal field" regions compose the optimum sensitive area named "Region Ω." In a color fusion image, we regard Region Ω as a rectangle instead of circular for convenience. In Region Ω, the target is assumed to be the center, from which the vertical and horizontal width r can be expressed as

$$r = D \tan\ (\theta/2), \quad (2)$$

$$N = rX/l. \quad (3)$$

where D denotes the distance between screen and viewer, θ denotes the viewing angle chosen as 2° in this case , r denotes the width of Ω region (unit: cm), X denotes the horizontal resolution of the display, l denotes the

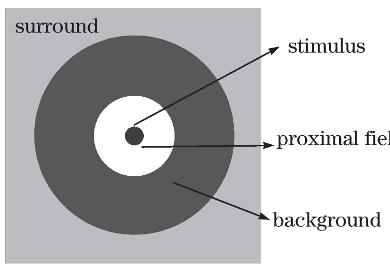

Fig. 1. Configuration of the viewing field.

horizontal size of the screen (unit: cm), and N denotes the width of Ω region expressed in pixels of the image. This way, the rectangle region Ω is selected in the color fusion image.

The following steps should be followed in segmenting Ω region into target and background region in the color fusion image. Since the color fusion image has similar gray distribution with source image, we segment Ω region in the source infrared (IR) image, and then map both the target and the background region to the color fusion image to simplify the process. The thresholding method is an effective tool to segment target and background. Otsu threshold segmentation algorithm is one of the most referenced because it is the most simple and standard for automatic threshold selection. A discriminant criterion is proposed in Otsu method to discover the optimum threshold as

$$\sigma_{\mathrm{B}}^2=\frac{[\mu_{\mathrm{T}}\omega(k)-\mu(k)]^2}{\omega(k)\,[1-\omega(k)]}, \tag{4}$$

where $\omega(k)=\sum_{i=1}^{k}P_i$ is the 0th cumulative moment of the histogram up to the kth level, $\mu(k)=\sum_{i=1}^{k}iP_i$ is the 1st cumulative moment, and $\mu_{\mathrm{T}}=\sum_{i=1}^{L}i\cdot P_i$ is the total mean level. Variable L denotes the levels of gray, k denotes the threshold, and P_i is the probability distribution. The optimal threshold k^* is obtained by maximizing σ_{B}^2[17]. In region Ω, the pixel whose value is less than k^* is represented by gray-level 0, that is, black for image. All the pixels are regarded as background region, and the brighter region is regarded as target. They are directly mapped to the final color fusion image.

According to perceptual experiments, we discover four features that could determine target detectability in its entirety: target's own luminance value L^*, which describes the degree of light or dark (luminance denotes the color channel in CIELAB space, not the gray intensity); perceptual luminance variation between target and background region, which indicates perceptual contrast; region chroma difference; and region hue difference between target and background. We generalize them to objectively evaluate target detectability.

In human visual system, Weber's law gives the ratio of just-noticeable-difference and intensity, which guides the research on relationship of perceptual variation and intensity variation. An approximately quantitative relation between the gray level variation ΔI and perceptual variation ΔI_{p} was measured according to Weber's law[18−19]. In our case, we assume that the luminance possesses similar characteristic as gray level to human vision, since both describe the degree of bright or dark. The single problem is the luminance value, being different in range to gray level. This leads to the transformation of luminance value to [0,255] range, that is

$$L^*(i,j)=255\cdot\frac{L^*(i,j)-\min L^*(i,j)}{\max L^*(i,j)-\min L^*(i,j)}\quad (i,j)\in\Omega, \tag{5}$$

where $\min L^*(i,j)$ and $\max L^*(i,j)$ are the minimum and maximum luminance values in region Ω, respectively. Then the perceptual luminance variation ΔL_{p}^* could be defined as

$$\frac{\Delta L_{\mathrm{p}}^*}{\Delta L^*}=c(L_{\mathrm{b}}^*)\approx\begin{cases}\dfrac{1}{(0.575-0.009L_{\mathrm{b}}^*)(L_{\mathrm{b}}^*+1)} & \text{if}\quad 0\le L_b^*<60\\ \dfrac{1}{0.035(L_{\mathrm{b}}^*+1)} & \text{if}\quad 60\le L_{\mathrm{b}}^*\le 200\\ \dfrac{1}{[0.035+0.001(L_{\mathrm{b}}^*-200)](L_{\mathrm{b}}^*+1)} & \text{if}\quad 200<L_{\mathrm{b}}^*\le 255\end{cases}, \tag{6}$$

where ΔL^* denotes the luminance difference between target and background region, defined as standard deviation of target luminance against background luminance.

$$\Delta I=\Big[\frac{1}{n_{\mathrm{t}}}\sum_{i,j\in\Omega_{\mathrm{t}}}\Big(L^*(i,j)-L_{\mathrm{b}}^*\Big)^2\Big]^{\frac{1}{2}}, \tag{7}$$

where $L^*(i,j)$ denotes each pixel's luminance in target region Ω_{t} and L_{b}^* denotes the mean luminance value of background region Ω_{b}.

$$L_{\mathrm{b}}^*=\frac{1}{n_{\mathrm{b}}}\sum_{i,j\in\Omega_{\mathrm{b}}}L^*(i,j), \tag{8}$$

where n_{t} and n_{b} are the number of pixels in the target and background region, respectively.

In Eq. (1), h^* is calculated in the form of a radian and should be converted to degree measure varying from 0 to 360°. The definition of hue difference between two regions is given by

$$\Delta h=\begin{cases}Rad\,\langle Std\rangle\,, & |Std|\le 180^\circ\\ 2\pi-Rad\,\langle Std\rangle\,, & |Std|>180^\circ\end{cases},\quad Std=\Big\{\frac{1}{n_{\mathrm{t}}}\sum_{i,j\in\Omega_{\mathrm{t}}}\Big[h(i,j)-h_{\mathrm{b}}\Big]^2\Big\}^{\frac{1}{2}}, \tag{9}$$

where $Rad\,\langle\cdot\rangle$ represents the conversion of degree to radian measure, and h_{b} denotes the mean hue value of the background region.

We first normalize the chroma value in region Ω. The region chroma difference is defined as

$$\Delta C=\Big\{\frac{1}{n_{\mathrm{t}}}\sum_{i,j\in\Omega_{\mathrm{t}}}\Big[C(i,j)-C_{\mathrm{b}}\Big]^2\Big\}^{\frac{1}{2}}, \tag{10}$$

where C_{b} represents the mean chroma value in region Ω_{b}.

In human vision system, human eyes are more sensitive to a variety of luminance than to hue and chroma. This leads to that we regard as ΔL_{p}^*, the principal factor with the largest contribution to target detectability. In addition. A brighter target could enhance perceptual contrast in a nonlinear way. Hence, we characterized ΔL_{p}^* and L_{t}^* in the form of exponent, where ΔL_{p}^* is selected as a bottom number. In this research, we normalized luminance value in region Ω to obtain a relative value range from 0 to 1. Thus, a target detectability metric d_{k} contributed by each factor was defined as

$$d_{\mathrm{k}}=(\Delta L_{\mathrm{p}}^*)^{L_{\mathrm{t}}^*}\cdot[t\times\Delta h+(1-t)\times\Delta C]^{\frac{1}{\gamma}}, \tag{11}$$

where L_t^* denotes the mean luminance value of target region Ω_t and variable t ($0 \leq t \leq 1$) is the adjustment parameter. Variable γ is the modulation coefficient to modulate the contribution of Δh and ΔC to d_k compared to that of ΔL_p^* and L_t^*. The larger the value of γ, the less the contribution of Δh and Δc to the final metric. We select t as 0.5 and γ as 2.0 in this letter. The larger the value of d_k, the better the target detectability.

This proposed evaluation was experimentally used to assess the target detectability in color fusion images. The first dataset was provided by TNO Human Factors Research Institute, as shown in Fig. 2. The detailed description of the data acquisition procedure can be found in Refs. [2] and [9].

The second dataset was obtained by our own fusion system, as shown in Figs. 3(a) and (b). Four methods were applied in fusion images in this dataset. In method 1, the IR and visible images were fused employing color transfer method in $l\alpha\beta$ space[2] using three different target images with nature color, as shown in Figs. 3(c)−(e). In method 2, firstly, the IR and visible images were fused using a fusion scheme based on the Laplacian pyramid with a 4-level decomposition. Secondly, the color image was produced by assigning the IR image to the red channel, the fusion image to the green channel, and the visible image to the blue channel, as shown in Fig. 3(f). In method 3, the IR and visible images were fused employing color transfer method in YUV space[20], as shown in Fig. 3(g). In method 4, the IR and visible images were fused employing TNO false color fusion method[1], as shown in Fig. 3(h).

The optimum sensitive area, Ω, was selected according to Eqs. (2) and (3). The images were displayed on a 19-inch liquid crystal display (LCD) with a resolution of 1440×900 pixels. The subjects viewed the display placed in an office at a regular distance of 50 cm. Hence, region Ω should be a rectangle area whose vertical and horizontal width towards the target are 30 pixels in each image, as shown within a rectangle frame in Fig. (4). Then, the target and background are segmented in region Ω according to the above-mentioned method. The results are shown in Fig. (4).

Finally, the proposed metric was employed to assess target detectability in the two datasets. For the first

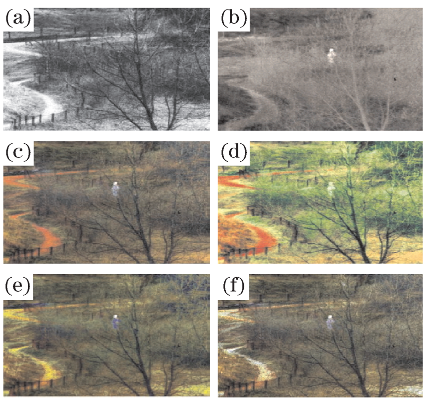

Fig. 2. Source images provided by TNO Human Factors Research Institute (a) visible image; (b) IR image; (c), (e), and (f) color fusion images of color transfer method; (d) color fusion image using method mentioned in Ref. [7].

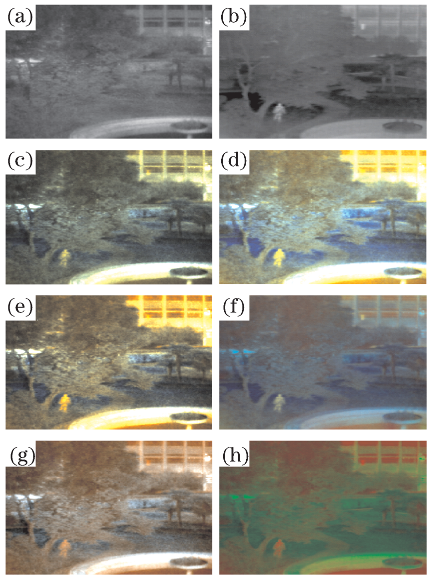

Fig. 3. Source images: (a) visible image; (b) IR image; (c), (d), and (e) color fusion image employing method 1 using different target images; (f) color fusion image employing method 2; (g) color fusion image employing method 3 using the same target image as “Image (e)”; (h) color fusion image employing method 4.

dataset, the results can be found in Table 1. The d_k value points out Figs. 2(c) and (f) outperforming the other two schemes. Although Fig. 2(d) performs best in luminance L_t^*, it still does not give excellent target detectability because of its least color difference between target and background region and its poorest perceptual luminance variation, which is considered the most important factor to recognize targets. In addition, Fig. 2(e) demonstrates poor result for this dataset although it has the biggest hue difference. This is mainly because it was poor in luminance L_t^* and perceptual luminance variation.

In the case of the second dataset, Figs. 3(e) and (d) are proven superior (Table 2). Figure 3(c) indicates comparable performance, whereas Figs. 3(f), (g), and (h) perform poorer. Figures 3(h) and (d) perform best in hue difference because the target and background regions in Fig. 3(h) appear red and green, respectively, which are opponent colors in colorimetry. Similarly, in Fig. 3(d), the target and background regions appear yellow and blue, which are also opponent colors. However, Fig. 3(h) performs worst in the final target detectability because of its poorest luminance and perceptual luminance variation. Comparing Figs. 3(c) and (d), perceptual luminance variation of the latter is much lower than that of the former. However, Fig. 3(d) outperforms Fig. 3(c) in final d_k value due to its excellent performance in luminance.

A qualitative subjective experiment has been conducted. A total of 25 subjects were served in this experiment. In each dataset, they were asked to judge target detectability of images with scores from “1” to “N”·“N” denotes the number of images in the dataset. For example, in the second dataset, “N” is 6, so that the scores are set as “1, 2, 3, 4, 5, 6.” A larger score denotes better target detectability. Then the overall score of every imagine is obtained by summing each person’s score. The subjective versus objective scheme ranking results of two datasets are shown in Figs. 5 and 6, respectively. The results show that the proposed metric d_{k} corresponds well with subjective scores.

In conclusion, an objective evaluation of target detectability has been proposed based on CIELAB color space. Four color information features are considered in establishing the objective target detectability metric.

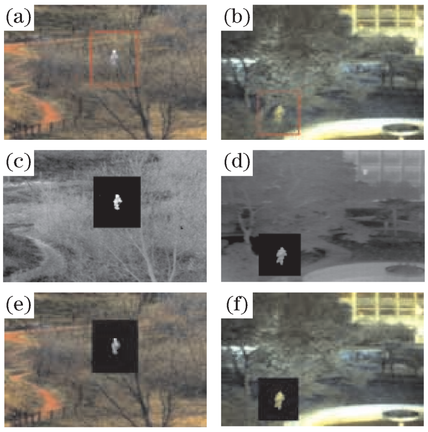

Fig. 4. Optimum sensitive region Ω within rectangle frame and results of target and background segmentation: (a), (b) region Ω in the first and second datasets, respectively; (c), (d) the extracted target in IR image of the first and second datasets, respectively; (e), (f) the extracted target in final color fusion image of the first and second datasets, respectively.

Table 1. Evaluation Results for the First Dataset in Fig. 2

Image	(c)	(d)	(e)	(f)
L_{t}^*	0.6969②	0.9051①	0.5916④	0.6760③
ΔL_{p}^*	5.9481①	3.3129④	3.5748③	5.6128②
ΔC	0.2383①	0.0154④	0.0927③	0.1926②
Δh	0.2257③	0.1723④	0.5244①	0.2686②
d_{k}	1.6686①	0.9059④	1.1803③	1.5412②

Table 2. Evaluation Results for the Second Dataset in Fig. 3

Image	(c)	(d)	(e)	(f)	(g)	(h)
L_{t}^*	0.7251 196	0.8999①	0.8288③	0.8445②	0.7446④	0.6282⑥
ΔL_{p}^*	7.3545②	4.8777③	7.5098①	4.6540④	3.7110⑤	0.9829⑥
ΔC	0.2488④	0.2494③	0.3029②	0.1396⑤	0.0607⑥	0.3850①
Δh	0.2167⑤	0.3141②	0.3003③	0.2917④	0.0926⑥	0.3627①
d_{k}	2.0502③	2.2094②	2.9202①	1.7015④	0.7350⑤	0.6049⑥

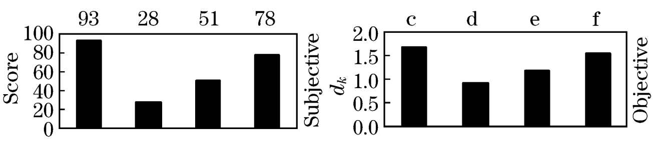

Fig. 5. Subjective versus objective scheme ranking results of Fig. 2. "c" "d" "e" "f" represent Figs. 2(c)−(f).

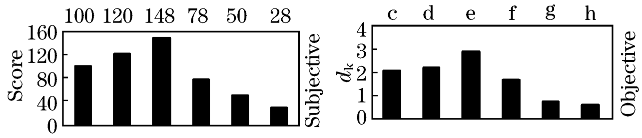

Fig. 6. Subjective versus objective scheme ranking results of Fig. 3. "c" "d" "e" "f" "g" "h" represent Figs. 3(c)−(h).

Among them, perceptual luminance variation is considered as the principle factor influencing the final target detectability metric, which is proposed based on Weber's law in human vision system. The experiment verifies that these four features determine target detectability in color image. In addition, experimental results show this target detectability metric could objectively evaluate target detectability in color fusion image, and corresponds well with subjective evaluation.

This work was supported by the Research Fund for the Doctoral Program of Higher Education (No. 20070288040) and the Scientific Innovation Research of College Graduate in Jiangsu Province (No. CX09B-097Z).

References

1. A. Toet and J. Walraven, Opt. Eng. **35,** 650 (1996).
2. A. Toet, Information Fusion **4,** 155 (2003).
3. S. Sun, Z. Jing, Z. Li, and G. Liu, Chin. Opt. Lett. **3,** 202 (2005).
4. S. Ma, J. Fang, S. Sun, and X. Gu, Acta Opt. Sin. (in Chinese) **29,** 1502 (2009).
5. Y. Zhao, L. Wang, W. Jin, S. Shi, and X. Wang, Acta Opt. Sin. (in Chinese) **29,** 654 (2009).
6. C. Xydeas and V. Petrovic, Electron. Lett. **36,** 308 (2002).
7. V. Tsagaris and V. Anastassopoulos, Opt. Eng. **45,** 026201 (2006).
8. W. Krebs, J. McCarley, T. Kozek, G. Miller, M. Sinai, and F. Werblin, in *Proceedings of the 43 rd Annual Meeting Human Factors and Ergonomics Society* 1333 (1999).
9. A. Toet and E. Franken, Displays **24,** 25 (2003).
10. J.-S. Shi, W.-Q. Jin, and L.-X. Wang, J. Infrared and Millimeter Waves (in Chinese) **24,** 236 (2005).
11. J. Shi, W. Jin, L. Wang, and H. Chen, Proc. SPIE **5640,** 594 (2005).
12. V. Tsagaris, Opt. Eng. **48,** 066201 (2009).
13. D. Malacara, *Color Vision and Colorimetry: Theory and Applications* (SPIE Press, Bellingham, 2002).
14. A. Toet, L. J. Vanruyven, and J. M. Valeton, Opt. Eng. **28,** 789 (1989).
15. S.-J. Wang, X.-H. Ding, Y.-H. Liao, and D.-H. Guo, Acta Elec. Sin. (in Chinese) **36,** 19 (2008).
16. Y. Dou, L. Kong, and L. Wang, Acta Opt. Sin. (in Chinese) **29,** 2511 (2009).
17. N. Otsu, IEEE Trans. Syst. Man Cybern. **9,** 62 (1979).
18. C. Wang and Z.-F. Ye, Acta Automatica Sinica **33,** 132 (2007).
19. Z. Xie and T. Stockham, Jr., IEEE Trans. Syst. Man and Cybern. **19,** 379 (2002).
20. L. Wang, S. Shi, W. Jin, and Y. Zhao, Proc. SPIE **6787,** 67870S (2007).