Analyzing Near-infrared Images for Utility Assessment

Neda Salamati, Zahra Sadeghipoor and Sabine Süsstrunk
School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
{neda.salamati,zahra.sadeghipoor,sabine.susstrunk}@epfl.ch

ABSTRACT

Visual cognition is of significant importance in certain imaging applications, such as security and surveillance. In these applications, an important issue is to determine the cognition threshold, which is the maximum distortion level that can be applied to the images while still ensuring that enough information is conveyed to recognize the scene. The cognition task is usually studied with images that represent the scene in the visible part of the spectrum. In this paper, our goal is to evaluate the usefulness of another scene representation. To this end, we study the performance of near-infrared (NIR) images in cognition. Since surface reflections in the NIR part of the spectrum is material dependent, an object made of a specific material is more probable to have uniform response in the NIR images. Consequently, edges in the NIR images are likely to correspond to the physical boundaries of the objects, which are considered to be the most useful information for cognition. This feature of the NIR images leads to the hypothesis that NIR is better than a visible scene representation to be used in cognition tasks. To test this hypothesis, we compared the cognition thresholds of NIR and visible images performing a subjective study on 11 scenes. The images were compressed with different compression factors using JPEG2000 compression. The results of this subjective test show that recognizing 8 out of the 11 scenes is significantly easier based on the NIR images when compared to their visible counterparts.

Keywords: V isual cognition, Near-infrared (NIR), Utility assessment, JPEG2000.

1. INTRODUCTION

In rapid-response scenarios, such as security and surveillance, imaging is often used to help human observers make immediate decisions.[1] Since these applications generally deal with large data through-put and in some cases limited storage capacity, it is desirable that the same accurate decisions can be made using compressed images as opposed to full resolution data. In sensor network systems, for instance, images used for detection and recognition should be stored in sensors and transmitted only upon request. Since the memory size and energy of each sensor is limited, it is essential to perform cognition using compressed images, which are easier to store and transfer.[2] Thus, an important question in cognition tasks is to define the cognition threshold, which is the maximum distortion level where observers are still able to correctly identify the scene.

To evaluate the usefulness of distorted images in cognitive tasks, Rouse *et al.*[3] introduced a *utility assessment* framework that specifies the largest distortion level that can be applied to images while still providing observers with enough information to identify the scene. In[3], the goal is to compare the cognition threshold for signal-based and structural-based representations of a scene, based on visible information. To generate different images in the signal-based sequence, the authors remove high frequency components of the image. The sequence of structural-based representations is formed by sequentially removing the low frequency information but keeping the edges of the image.

Similar to[3], utility assessment is usually studied with images that represent the scene in the visible part of the spectrum.[4,5] In this paper, our goal is to explore whether or not using representations of a scene other than visible (spectrum) images can facilitate the cognition task. To this end, we study how well near-infrared (NIR) images perform in utility assessment.

NIR has already been used in some recognition and detection frameworks, such as face recognition,[6] night vision in military systems[7] and pedestrian detection.[8] In face recognition, NIR information is incorporated to construct a facial feature detection system based on the fact that multi-band spectral measurements of facial skin differ significantly from person to person.[6] In pedestrian detection and night vision enhancement systems, the main purpose is to make the detection and recognition of pedestrians, cyclists, and animals easier at night.[9] These systems take advantage of the different reflection in the NIR and visible parts of the spectrum.[8,10]

Due to the proximity of NIR to visible radiation, NIR images share many properties with visible images, in particular the shapes of objects in the scene are preserved. However, surface reflection in NIR is material dependent, consequently these images reveal different characteristics of the scene.[11,12] Objects made of a specific material mostly have the same intensity in NIR images, thus it is more probable that edges in NIR images represent the physical shape of the object rather than changes in intensity or color within the object (see Fig. 1). According to[4], visual cognition occurs mostly based on the shape properties of the objects rather than the color or patterns. Considering the essential information for cognition along with the mentioned NIR image features, we propose that NIR is an appropriate representation to be used in scene cognition tasks.

In our experiment, we compare the cognition threshold of visible and NIR representations of a scene. To this end, we acquired a database of 11 visible/NIR image pairs and compressed them using the JPEG2000 standard with different compression ratios. A subjective test was conducted to measure the cognition threshold in NIR and visible images for each scene. The cognition threshold for each scene is defined as the most compressed image that still conveys enough information for the observer to be able to correctly describe the scene. The subjective test results show that in most cases, the mean cognition threshold of NIR images is lower than the visible counterparts.

This paper is structured as follows. In Section 2, we discuss NIR image properties in more detail and explain how they can be useful in cognition tasks. We also explain the effects of compression on NIR and visible images and how it affects the scene cognition. Section 3 describes the method used for measuring the cognition threshold. We analyze the result of the subjective test in Section 4, and Section 5 concludes the article.

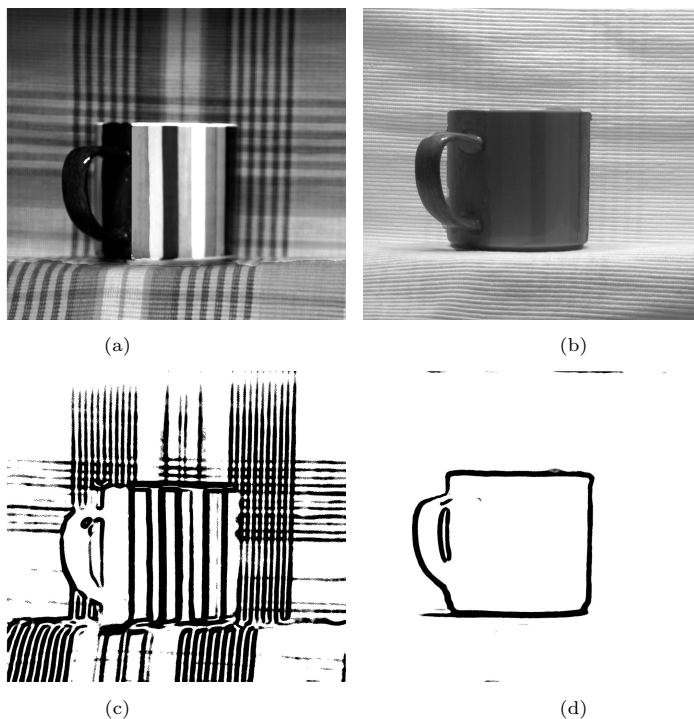(a)　　　　　　　　　　　　　(b)

(c)　　　　　　　　　　　　　(d)

Figure 1. (a) Visible image, (b) NIR image. (c) Edge map of the smoothed visible image. (d) Edge map of the smoothed NIR image. The edge maps were produced by using difference of Gaussians on the low-pass filtered image.
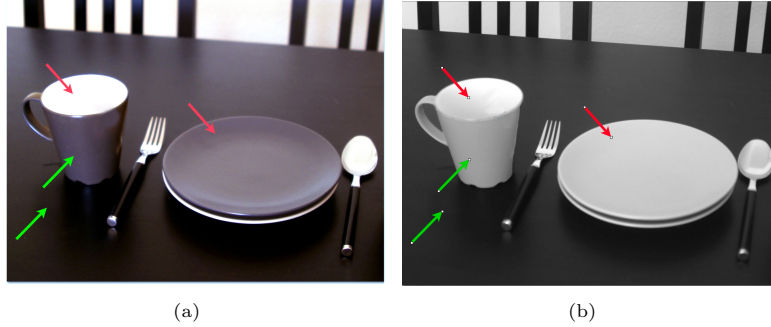
Figure 2. (a) Visible image, (b) NIR image. See text for explanation.

2. MOTIVATION

The sensors of digital cameras are made of silicon that is inherently sensitive to the NIR part of the radiation spectrum (700-1100 nm) as well as the visible band (400-700 nm). Capturing NIR images can thus easily be done by removing the hot mirror of the camera, which is used to absorb the NIR radiation, and then mounting a visible cut-off filter in front of the lens.[13]

The appearance of NIR and visible images of the same scene is similar and the general shape of objects are preserved in NIR images (see Fig. 2 for illustration). However some differences can be observed in the NIR and visible representations of a scene. The reflection in the NIR band is more material dependent compared to the reflection in the visible part. It means that objects made of a specific material mostly look the same in the NIR images.[11,14] For instance, the white inside of the mug and the black plate in Fig. 2 (red arrows), made of the same material, appear the same in the NIR image in spite of having different intensities in the visible image. However, the black outside of the mug and the black table (Fig. 2, green arrows), made of different materials, have the same appearance in the visible image but different intensities in the NIR representation. This feature makes the edges in the NIR image to mostly correspond to the shape boundaries of the physical objects. This is illustrated in the second row of Fig. 1, which shows the edge images of the smoothed NIR and visible representations of a scene.

Since the general shape of the objects does not change in the NIR images, edges that correspond to the object boundaries are present in these images, as can be seen in Fig. 1. On the other hand, reflection in the NIR part of the spectrum mostly does not depend on color, so some of the patterns and fine details in the real scene disappear from the NIR images while they are perceivable in the visible images. For instance, in Fig. 1 patterns on the cup and the background are attenuated or disappeared altogether in the NIR image. As visual cognition occurs mostly based on the shape properties of the objects rather than the objects' patterns and the fine details in the scene[4,15] the lack of fine details in NIR images compared to visible should not considerably affect the cognition threshold. Moreover, in some cases the presence of surface patterns in the visible image may create more difficulties in recognizing the scene. For example in Fig. 1, the surface patterns of the cup and the background in the visible image are distracting, while in the NIR image the shape of the cup can easily be perceived.

The material dependency of reflection in the NIR band results in preserving the object boundaries, which are the essential information in object recognition, while suppressing surface pattern due to coloration. Thus, our hypothesis is that cognition using compressed NIR images will achieve superior results compared to compressed visible images. The experimental setup used to test the above hypothesis is explained in the next section.

3. METHODOLOGY

The goal of this paper is to compare the cognition threshold for the NIR and visible representations of a scene. We applied the signal-based representation framework proposed in[3] to generate the sequence of distorted images for visible and NIR representations, and conduct a subjective test to measure the human cognition threshold for these representations.
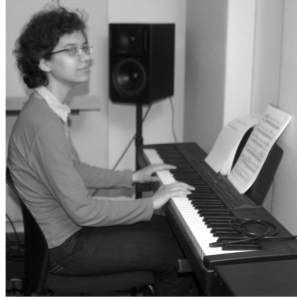
(a) girl in a field

(b) girl playing keyboard

(c) guitar player

(d) girl in a library

(e) train

(f) cyclist

(g) car

(h) soccer player

Figure 3. The visible and NIR representations of the scenes used in the test.
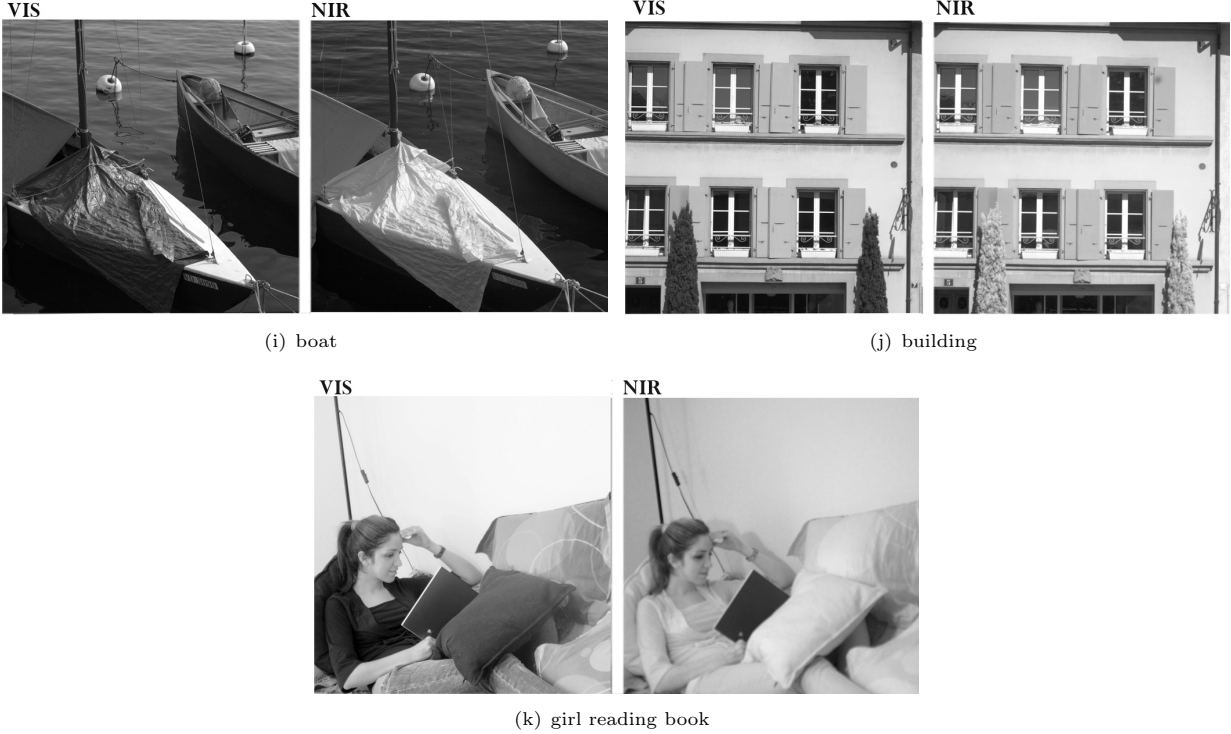
(i) boat

(j) building

(k) girl reading book

Figure 3. The visible and NIR representations of the scenes used in the test (cont.).

22 images of size 412×412, which represent 11 different natural scenes in the visible and NIR parts of the spectrum was acquired. Figure 3 shows the images and a description for each image. The scenes were chosen to be similar to the images used in[3] . The sequence of distorted images was generated with JPEG2000 image compression,[16] which is fairly consistent with human perception. JPEG2000 mostly removes highly textured regions in the image[17] , which are not considered useful information for cognition[4] .

The bitrates (BR) of compressed images was first chosen to be logarithmically equally spaced between 0.017 to 0.765 bpp to generate 16 images (as has been proposed in[3]). After conducting a preliminary test, we found that using logarithmically equally spaced bitrates resulted in the same cognition thresholds for different observers. An explanation could be that applying the chosen set of bitrates, the perceived amount of visual information is changing highly non-uniformly in the subsequent compressed images, which could decrease the accuracy of measuring the cognition threshold. Thus, we generated 31 images with the following bitrates in the interval of 0.017 to 0.765 bpp (Fig. 4 shows some images in that sequence):

BR $\in$ {0.0170, 0.0201, 0.0210, 0.0222, 0.0229, 0.0234, 0.0239, 0.0246, 0.0278, 0.0307, 0.0327, 0.0344, 0.0393, 0.0457, 0.0550, 0.0644, 0.0774, 0.0920, 0.1078, 0.1255, 0.1511, 0.1815, 0.2041, 0.2396, 0.2864 0.3411, 0.4078, 0.4863, 0.5481, 0.6480, 0.7600}

A subjective test was conducted to identify the observers' cognition threshold for both types of representations. Since there are two representations of each scene (NIR and visible), it is important to make sure that each subject observes only one type of representation for each scene. Thus, observers who viewed the NIR image of a specific scene did not judge the corresponding visible image and vice versa. Each observer viewed a set of 31 compressed versions of each scene, which evolved from very low quality to the highest quality and was asked to provide a general description of the scene. No time limit was imposed in the test. The bitrate of the images from which the observer was able to describe the scene correctly was recorded as the cognition threshold for each scene and observer. The average cognition bitrate for each scene and all the observers was computed as the cognition threshold for that scene.

30 observers (11 female, 19 male) with normal or corrected-to-normal acuity participated in the test, with ages from 24 to 39. Thus, each representation of the scene was viewed by 15 observers. A 24" display (Apple LED cinema display) with a resolution of 1920×1200 pixels was used to display the images, and normal office lighting was used in the test environment. The observers were seated at a distance of 40 cm from the display. The 412×412 images were displayed on a gray background ($R = G = B = 128$).

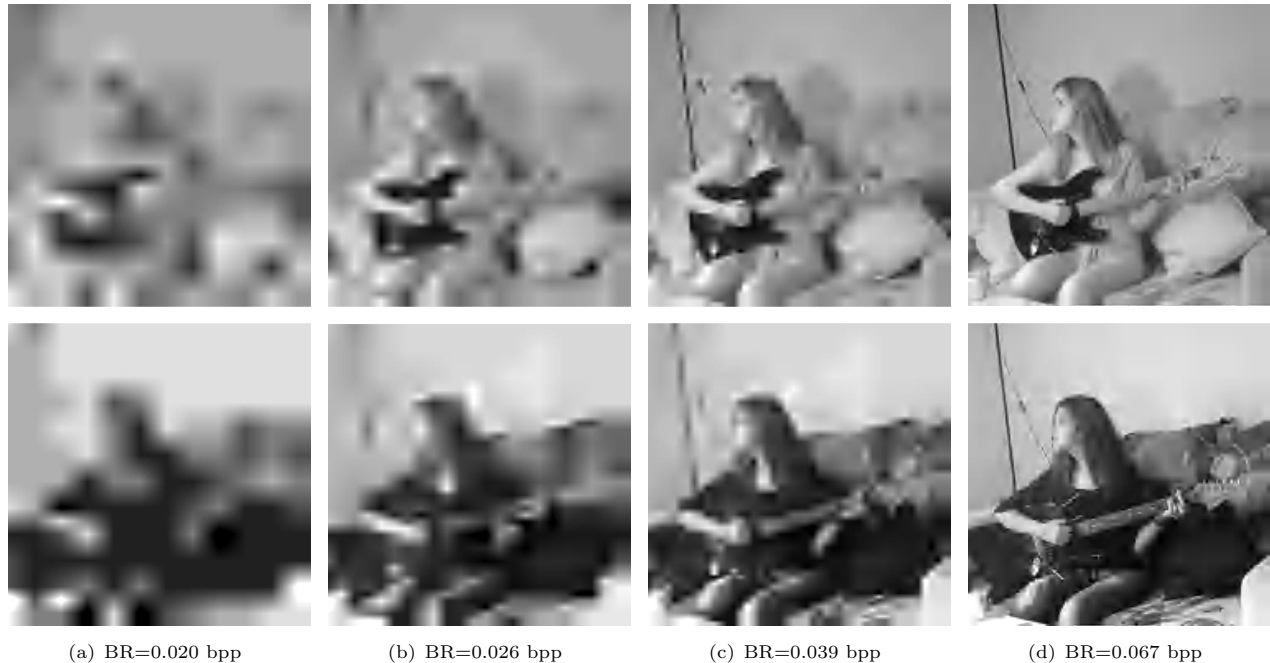(a) BR=0.020 bpp   (b) BR=0.026 bpp   (c) BR=0.039 bpp   (d) BR=0.067 bpp

Figure 4. Four of 31 images of the scene "guitar palyer" with increased bitrate (BR) from 0.020 to 0.067 bpp (First row:NIR images, second row: visible images).

4. RESULTS AND DISCUSSION

To study the difference between the cognition threshold in NIR and visible images, we first analyzed the mean cognition bitrate (M_i) for both the NIR and visible representations of all scenes in the database.

$$M_k = \frac{\sum_{i=1}^{N} m_{k,i}}{N} \qquad k \in \{NIR, VIS\} \tag{1}$$

where N is the number of valid observations for evaluating k representation (in our experiment $N = 15$). $m_{k,i}$ is the cognition threshold for evaluation of the k representation by the ith observer. Fig. 5 shows M_k for all the images in the database. The confidence interval for each image, as shown in Fig. 5, corresponds to 95% confidence level. For 10 images, the NIR representation has a cognition threshold at a lower bitrate than visible images. We applied the *analysis of variance* ($ANOVA$) test[18] in order to find out if the difference of the mean cognition bitrate is statistically significant. The smaller the p, the more significant the difference between the cognition bitrate of NIR and visible images. A typical value for rejecting the hypothesis that the scores for different representations come from the same population is $p < 0.05$. As noted in Table 1, for 8 out of 11 images in the database, observers were able to accurately describe the scene in the NIR representation at a statistically significantly smaller bitrate than the visible images. For the "soccer player" (h) and "boat" (i) images, although the cognition bitrates of the NIR images are lower than the cognition threshold for the visible images, the difference is not significant. We noticed that for the "soccer player" image, the observers were mainly focusing on the central part of the images and could not recognize the ball and the pose of the soccer player's foot, thus they could not come up with the exact description of "soccer player" at low bitrates for the NIR or
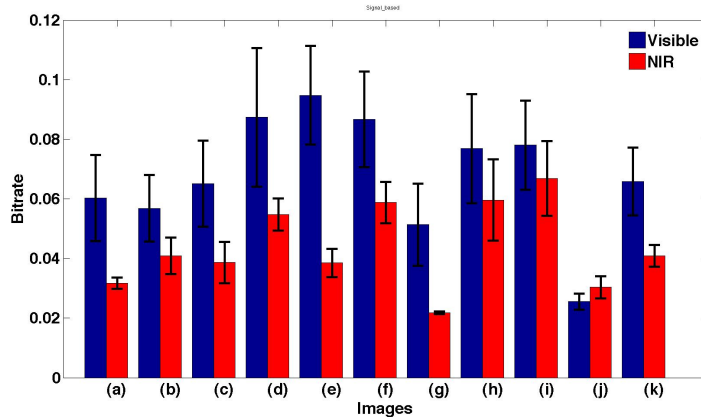
Figure 5. Mean cognition bitrate for both visible and NIR representation of all the images in the database.

visible representations. In the case of the "boat" image, the visible representation does not contain too many fine details. As a result, the cognition thresholds for the NIR and visible image does not differ significantly. As can be seen in Fig. 5, it is easier to recognize the "building" scene (image (j)) in the visible representation. However, the difference in cognition thresholds in NIR and visible images is not statistically significant.

Image label	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
p value	0.002	0.036	0.007	0.022	0.000	0.009	0.001	0.190	0.314	0.072	0.0011

Table 1. p value for all the image pairs in the database.

These results suggest that in many cases it is easier to recognize scenes from distorted NIR images than visible images. However, when different color (or intensities) within the same material are essential for scene recognition, the NIR representation might fail. For instance, a scene containing a flower surrounded by other vegetation would be recognizable in the visible image even if the image is highly distorted (see Fig. 6). Flower and grass have the same chemical composition and appear the same in the NIR images. Thus, choosing the NIR or the visible representation in cognition tasks depends on the specific application. While in many situations, compressed NIR images are more easily recognized than their visible counterparts, their material dependency needs to be evaluated given the scenario.

5. CONCLUSION

We examined the usefulness of NIR representation in cognition tasks and compared the cognition thresholds of NIR and visible images. The cognition thresholds in NIR and visible images were evaluated using a sequence of compressed images. The results of the subjective test verified the hypothesis that for scene recognition, in many cases the NIR image is a better representation than the visible.

Since one of the basic differences between NIR and visible images is that edges in the NIR image mostly correspond to physical object boundaries, the results of this paper support the hypothesis that the essential features for visual cognition are object boundaries rather than color and texture. Thus, it is interesting to study the human cognition threshold in a scene representation which only contains the edges of the original scene. This "structural-based" representation for scene recognition has been studied in[3] where the images represent the scene in the visible part of the spectrum. As future work, we will thus examine the performance of NIR and visible structural-based representations and analyze the advantages and disadvantages of the structural representations of NIR images compared to the visible images.

(a) Visible image   (b) NIR image

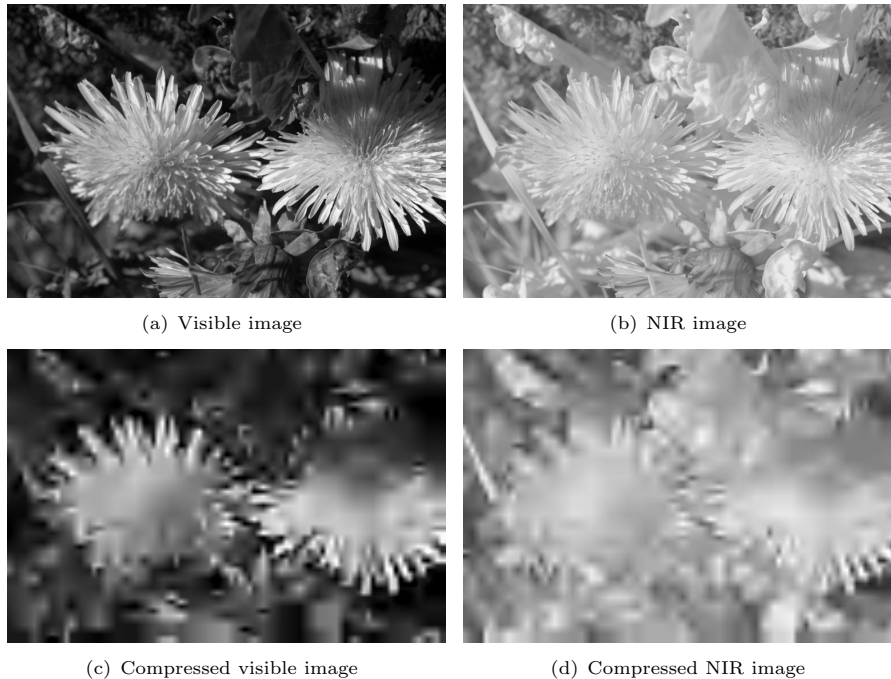(c) Compressed visible image   (d) Compressed NIR image

Figure 6. A typical photograph with vegetation. (a) Visible image, (b) NIR image (c) Compressed visible image (d) Compressed NIR image. Flower and grass have the same chemical characteristics and appear the same in the NIR image.

ACKNOWLEDGMENTS

REFERENCES

[1] Peli, E., [*Vision models for target detection and recognition*], World Scientific Publishing (1995).

[2] Chow, K.-Y., Lui, K.-S., and Lam, E. Y., "Efficient on-demand image transmission in visual sensor networks," *EURASIP Journal on Advances in Signal Processing* **2007** (2007).

[3] Rouse, D. M. and Hemami, S. S., "Quantifying the use of structure in cognitive tasks," in [*Proceedings of IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging*], **6492** (2007).

[4] Ullman, S., [*High-level vision: object recognition and visual cognition*], MIT press, third ed. (2000).

[5] Kosslyn, S. M., "Information representation in visual images," *Cognitive Psychology* **7** (1975).

[6] Li, S. Z., Zhang, L., Liao, C. S., Zhu, X. X., Chu, R. F., Ao, M., and He, R., "A near-infrared image based face recognition system," in [*Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*], (2006).

[7] Meitzler, T., Bryk, D., Sohn, E., Lane, K., Bednarz, D., Jusela, D., Ebenstein, S., Smith, G., Rodin, Y., Rankin, J., and Samman, A., "Noise and contrast comparison of visual and infrared images of hazards as seen inside an automobile," in [*Proceedings of SPIE Photonics West: Enhanced and Synthetic Vision*], **4023** (2000).

[8] Tsimhoni, O., Bärgman, J., Minoda, T., and Flannagan, M. J., "Pedestrian detection with near and far infrared night vision enhancement," Tech. Rep. UMTRI-2004-38, The University of Michigan, Transportation Research Institute (2004).

[9] Rumar, K., "Night vision enhancement systems: what should they do and what more do we need to know?," Tech. Rep. UMTRI-2002-12, The University of Michigan, Transportation Research Institute (2002).

[10] Küpper, L. and Schug, J., "Active night vision systems," Tech. Rep. SAE Technical Paper Series No. 2002-01-0013, Society of Automotive Engineering (2002).

[11] Salamati, N., Fredembach, C., and Süsstrunk, S., "Material classification using NIR and visible images," in [*Proceedings of IS&T/SID 17th Color Imaging Conference*], (2009).

[12] Salamati, N. and Süsstrunk, S., "Material-based object segmentation using NIR information," in [*Proceedings of IS&T/SID 18th Color Imaging Conference*], (2010).

[13] Fredembach, C. and Süsstrunk, S., "Colouring the near-infrared," in [*Proceedings of IS&T/SID 16th Color Imaging Conference*], (2008).

[14] Burns, D. and Ciurczak, E., [*Handbook of near-infrared analysis*], Marceld Ekkerin,Inc. (2001).

[15] Sassi, M., Vancleef, K., Machilsen, B., Panis, S., and Wagemans, J., "Identification of everyday objects on the basis of gaborized outline versions," in [*Proceedings of European Conference on Visual Perception (ECVP)*], (2010).

[16] Taubman, D. S. and Marcellin, M. W., [*JPEG2000: image compression fundamentals, standards, and practice*], Kluwer Academic Publishers (2002).

[17] Sheikh, H., Bovik, A., and Cormack, L., "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Transactions on Image Processing* **14** (2005).

[18] Snedecor, G. W. and Cochran, W. G., [*Statistical methods*], Iowa State University Press (1989).