S. Süsstrunk and C. Fredembach, SID International Symposium Digest, 2010.

8-1: *Invited Paper:* Enhancing the Visible with the Invisible: Exploiting Near-Infrared to Advance Computational Photography and Computer Vision

Sabine Süsstrunk and Clément Fredembach
School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland

Abstract

Silicon-based digital camera sensors exhibit significant sensitivity beyond the visible spectrum (400-700nm). They are able to capture wavelengths up to 1100 nm, i.e., they are sensitive to near-infrared (NIR) radiation. This additional information is conventionally treated as noise and is absorbed by a NIR-blocking filter affixed to the sensor.

We show that retaining instead of removing NIR information can significantly improve certain computational photography and computer vision tasks. Indeed, intrinsic properties of the NIR wavelength band guarantee that images can be sharper, less affected by man-made colorants, and more resilient to changing light conditions. The benefits of using NIR images in conjunction with standard color images in applications such as haze removal, skin smoothing, single and multiple illuminant detection, shadow detection, and material classification is discussed.

1. Introduction

A digital camera sensor (either CCD or CMOS) is made of silicon, a semi-conductor whose photosensitivity ranges from roughly 200nm to 1100nm. As current digital cameras' primary goal is to capture and reproduce the visible spectrum, a near infrared (NIR) blocking filter, also known as hot-mirror, is placed in front of the sensor. Ultraviolet light (200-400nm) is filtered out by the optical elements of the camera.

Thus, most camera sensor sensitivities are limited to the visible spectrum only, i.e., a wavelength range of 400-700nm. This is suboptimal because the physical differences in the RGB and NIR wavebands result in markedly dissimilar images that, when compared, offer powerful clues about the real world. The inherent differences in RGB and NIR image intensities and frequencies allow us much better disambiguating scene elements, which can be exploited in image enhancement, image segmentation, object recognition, and image classification. Additionally, mathematically ill-determined problems, such as multiple illuminant and shadow detection, become manageable.

In this paper, we present an overview of our current research [1-7] on combining RGB and NIR images to *enhance* photography or to *extract* more accurate information about the visible scene. While RGB and NIR-only imaging has been popular for many years in scientific, commercial, and consumer imaging [1], the combination of the two to improve computational photography and computer vision tasks has only recently been introduced [1-10].

The article is organized as follows. In Section 2, we discuss the physical properties of NIR that result in different image content. In Section 3, we show two applications where NIR images are used to enhance image quality of color photographs, namely in haze removal and in realistic skin smoothing. Section 4 introduces applications where we use NIR images to extract more information about the original scene: illuminant, multiple illuminant and shadow detection as well as material classification. Section 5 concludes the paper.

2. NIR Properties

Visible and NIR images of the same scene are, at first regard, very similar. We can easily recognize that they are taken from the same viewpoint and contain the same scene elements (see Figure 1). On second inspection, however, we do notice fundamental differences. In the NIR image, the trees in the foreground, for example, are much lighter and the mountains in the background are much clearer, i.e. they have more contrast.

Figure 1. Visible (RGB) and NIR image of the same scene.

Some of the physical phenomena resulting in fundamental differences between color and NIR images' intensity, frequency, and contrast that we exploit in our research are discussed in the subsection below.

2.1. Light Scattering

Small particles scatter incident light, which can alter the light intensity at specific wavelengths. When the particles' size is very small (< λ/10), radiation behaves according to Rayleigh scattering, which states that:

$$\frac{I_S}{I_O} = \frac{\text{constant}}{\lambda^4}$$

i.e., the intensity of the scattered light I_S is related to that of the incident light I_O by the inverse of the fourth power of the wavelength λ. As a result, sky appears blue because it is the most scattered color due to its relative short wavelength. In comparison, NIR at 1000 nm is 40 times less scattered than blue at 400nm. As shown in Fig. 1, NIR intensity thus tends to be dramatically lower than its RGB counterpart in sky and its reflected components, such as water.

Another effect of Rayleigh scattering can be observed when atmospheric haze or pollution is present. NIR images appear sharper and contain more details in distant objects than the corresponding color photograph. The effect is clearly visible by looking at the mountains in Figure 1. We exploit this fact in our haze removal algorithm (see Section 3.1).

With increasing particle size, the physical interactions between light and particles change from Rayleigh-type to Mie-type scattering. When the particles are large (> λ/10), spherical, and diverse, the dependency between scattering intensity and wavelength disappears and all wavelengths are equally scattered.

Thus, clouds will look similar in RGB and NIR images, but the contrast between sky and clouds is enhanced in the NIR image.

2.2. Molecular Structure

The interaction of light and particles in the atmosphere is not the sole reason why the intensity content of NIR images differs from those of RGB ones. In general, different molecular structures and complex absorption spectra of natural and man-made materials will result in different intensities in RGB and NIR images.

Figure 2. *Left:* RGB image. *Right:* NIR image. See text for explanation.

A valuable use of NIR images is to distinguish between surfaces that appear identical to the human eye. Indeed, NIR imaging is "transparent" to a number of colorants or paints and as such can see through that first layer to reveal the surface underneath. There are thus surfaces that can be distinguished in the visible spectrum but appear the same in the NIR. The converse is also true: some surfaces that have the same color in the visible images show different intensities in the NIR images, due to different material characteristics. We can exploit this "NIR-metamerism" for image segmentation and material classification (see Section 4.2). An illustration of everyday objects exhibiting this behavior is provided in Fig. 2. Note that the red and white scarf pattern (middle of the image, green arrow) disappears in the NIR. On the other hand, the two black objects (left side of the image, yellow arrows) have a very different NIR response. The same applies to the turquoise colored scarf and strap in the right side of the image.

3. Exploiting NIR to *Enhance* Images

As discussed above, color and NIR images have different intensity and frequency content, given by light and material interactions. We can take advantage of these differences in two ways. First, by selective registration and fusion of color and NIR images, we can enhance certain regions of the color image to improve visual quality. Two such applications, haze removal and skin smoothing, are introduced below.

3.1. Haze Removal

In landscape photography, distant objects often appear blurred with a blue colorcast, a degradation caused by atmospheric haze. The goal of haze removal, or *dehazing*, an image is thus to enhance its contrast, pleasantness, and information content.

Current automatic haze removal methods using a single image make use of the haze imaging equation, which is the sum of two terms, *direct attenuation* and *airlight*. Direct attenuation describes the scene radiance and its decay in the medium, airlight is the ambient light reflected into the line of sight by atmospheric particles [12]. As the problem is under-constrained, correct assumptions need to be made in order to obtain good results.

Tan [13] observes that haze-free images have larger local contrast and that the airlight is smooth. His results, after maximizing local contrast, tend to be oversaturated and can yield halo artifacts. Fattal [14] obtains, after assuming that the transmission and surface shading are uncorrelated, physically correct dehazed images, but his assumption might fail in cases of very dense haze. He et al. [12] introduce the dark channel prior, based on the observation that very dark pixels exist in every part of natural scenes. The additive airlight brightens these dark pixels, increasing with distance. A depth map can thus be obtained, which is then used to recover the scene radiance.

Figure 3. Comparison of the original image, our visible/NIR fusion, Fattal's dehazing approach (Courtesy of R. Fattal), and He et al.'s haze removal (Courtesy of K. He).

The disadvantage of these proposed based techniques is their complexity. Our method, which exploits Raleigh scattering and thus the fact that haze is much less present in NIR images, is fast and does not need heuristics for haze and airlight detection.

To fuse visible and NIR images, we first transform the visible RGB image into a luminance-chrominance color space. We obtain a one-channel NIR image containing intensity data [1]. We apply an edge-preserving multiresolution decomposition based on the Weighted Least Squares (WLS) optimization framework [15] to both the visible luminance and the NIR intensity images. At each resolution, a pixel level fusion criterion that maximizes contrast is applied, with the exception of the lowest frequency approximation image where the visible luminance image is retained to take into account the luminance perception of a human observer [3].

In Figure 3, we compare our algorithm to He et al.'s [12] and Fattal's [14] single color image haze removers. Our method improves any scattering degradation following Rayleigh's law, as NIR image are intrinsically haze-free. As the high frequency content of NIR images is otherwise very similar to the visible, our algorithm can be applied to any image, regardless of the actual presence of haze, without introducing artifacts. Moreover, we do not need to generate a correct depth map or to detect the airlight, tasks that can be error prone.

3.2. Realistic Skin Smoothing

Another photographic enhancement application where NIR images can provide essential information is in portrait photography. Large efforts are often undertaken to enhance such images. While some artifacts are induced by the capturing process, a number of unwanted details are "intrinsic" to the photographed person, e.g., pores, wrinkles, freckles, and spots. As a result, models and photographers employ many techniques to mask or correct these less appealing features, such as make-up, image editing software, or print airbrushing.

Figure 4. RGB image, NIR image, and fused image.

NIR images are mostly free of these artifacts; skin appears much smoother (see Figure 4). This is due to the absorption behavior of melanin and hemoglobin, the key components of skin color. They have little absorption in the near-infrared part of the spectrum; the depth of light penetration in the skin is approximately proportional to the incident radiation's wavelength. As most of the unwanted skin artifacts are on or near the skin's surface, they are to a large extend attenuated in the NIR image.

Our framework [6] thus consists of capturing a pair of visible/near-infrared images and separating both of them into base and detail layers (akin to a low/high frequency decomposition) using the fast bilateral filter [11]. A smooth, realistic, output image can be obtained by fusing the base layer of the visible image with the near-infrared detail layer. The results look realistic, as can be observed in Figure 6.

4. Exploiting NIR to *Extract* Information

In the previous section, we have shown two applications where selective fusion of visible and NIR images can enhance a photographic image. In this section, we show applications where the extra information that NIR images provide can give more accurate information about the original scene content.

4.1. Illuminant and Shadow Detection

The color signal captured by an imaging system, eye or camera, is the product of incident light and the imaged surface's reflective properties. The human visual system is, to a certain extent, color constant, i.e. it is able to discount the contribution of the illuminant and retain the object's surface color. This is not the case for a camera. Thus, objects are not captured and rendered properly, e.g., snow illuminated by skylight appears blue in photographs even though we see it white.

In order to make a camera color constant (a process also known as white-balancing), one effectively has to separate the incident color signal into its illuminant and reflectance components. For regular RGB cameras, the problem is ill-posed, and solving for color constancy can only be done using either additional information or by making assumptions about the imaged scene or the world in general [2].

Using NIR information in conjunction with visible images for solving color constancy effectively doubles the wavelength range over which the illuminant can be estimated. Camera RGB filters have peak sensitivities that are only about 100nm apart, thus illuminant detection and estimation is usually performed by comparing fairly correlated information. Because of its greater distance to the visible spectrum (400nm on average), NIR is less correlated and can therefore provide more relevant information for illuminant estimation.

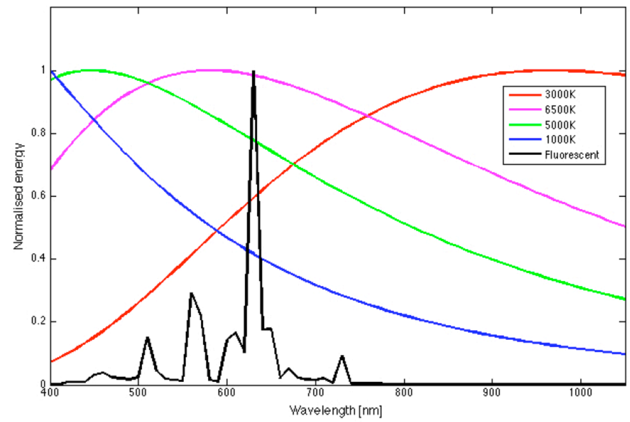

Figure 5. Normalized energy of common light sources. Note that in the NIR, the curves are monotonic and well separated over that part of the spectrum. Fluorescent lights have a very limited output in the near-infrared.

We show that considering R, G, and B to NIR pixel ratios is very robust to most commonly encountered illuminants and able to retrieve the correct solution [2]. The main advantage of using NIR is that for common lighting conditions, it has very large response variation with respect to the type of incident light: incandescent light bulbs have their emission peak in the NIR, while scattered skylight (which is the color of outdoor shadows) and fluorescent lighting have virtually no emission in that part of the spectrum (Figure 5). This actually enables us to distinguish between different lights that have an identical white point, but different metameric properties. Using the same observations as above and extending the ratio comparisons to image regions also allows detecting multiple illuminants in a scene (Figure 6).

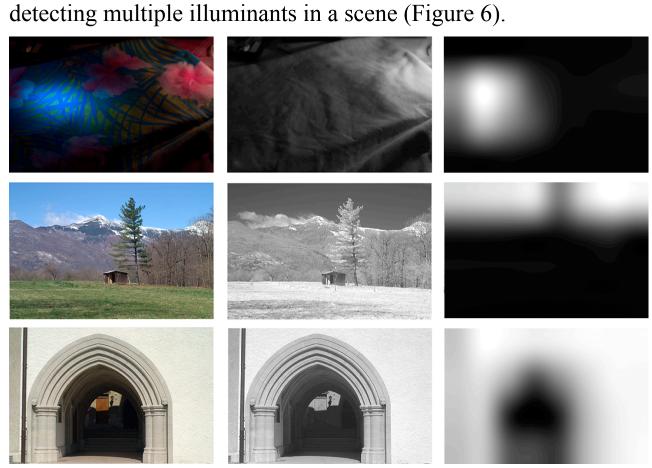

Figure 6. Column-wise: RGB image, NIR image, and illuminant map. *Top row*: a cloth lit by an LED and an incandescent light bulb. Note how the bright and coherent LED light source disappears in the NIR. *Middle row*: a simple outdoor scene where the sky is very well detected from the rest of the sun-illuminated scene. *Bottom row*: out-door shadow detection.

Figure 7. The visible image, binary shadow map, and shadow mask obtained by a matting algorithm [16].

Shadow detection is a special case of multiple illuminant detection. Starting with the observation that in most imaging situations (with the exception of carbon black and water), the object reflectance times the channel's quantum efficiency is much higher in the NIR than any of the visible R, G, or B channels, we can deduct that dark areas in both visible and NIR images are prime shadow candidates [7]. This allows us to create a binary shadow map that can be input into a matting algorithm [16] to create a full shadow mask.

4.2. Material Classification

As mentioned above, NIR imaging is "transparent" to a number of colorants or paints and as such can see through the first layer to reveal the surface underneath (see Figure 2). The NIR intensity better reflects the material's reflective property; it is often not influenced by surface colorants. Additionally, texture patterns resulting from different colorization of an object disappear. NIR images used in conjunction with visible images can thus give us a much better understanding of an object's inherent material composition.

In a first study [5], we have analyzed the classification of four material classes (textiles, tiles, linoleum, and wood) with a total of 51 samples. The features were color, visible lightness versus NIR intensity, and texture. We obtained a classification rate of 98% compared to 60% using features of the visible images only.

5. Conclusion

We showed that combining "invisible" NIR images with color photographs improve several computational photography and computer vision tasks, such as haze removal, skin smoothing, illuminant and shadow detection, and material classification. Silicon, the light sensitive material of current digital camera sensors, is inherently capable of capturing both visible and NIR radiation.

Widespread adoption of the presented techniques will depend on the development of a camera that can concurrently capture RGB and NIR images on a single sensor. We have thus started to investigate [4] how such a camera system needs to be designed. However, by either capturing a scene sequentially and exchanging a hot mirror with a visible light-blocking filter between the two shots [1], or by using a two sensor and beam-splitter set-up [8,9], research in harnessing the power of multispectral image acquisition can continue to progress.

6. Acknowledgements

We would like to thank our co-authors N. Barbuscia, Y.M. Lu, N. Salamati, L. Schaul, and M. Vetterli. Part of the work presented in this paper was supported by the Swiss National Science Foundation under grant number 200021-124796/1.

7. References

[1] C. Fredembach and S. Süsstrunk, "Colouring the Near-Infrared," *Proc. IS&T/SID 16th CIC* (2008).

[2] C. Fredembach and S. Süsstrunk, "Illuminant Estimation and Detection using Near-Infrared," *Proc. IS&T/SPIE EI: Digital Photography V* (2009).

[3] L. Schaul, C. Fredembach, and S. Süsstrunk, "Color Image Dehazing using the Near-Infrared," *Proc. IEEE ICIP* (2009).

[4] Y. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk, "Designing Color Filter Arrays for the Joint Capture of Visible and Near-Infrared Images," *Proc. IEEE ICIP* (2009).

[5] N. Salamati, C. Fredembach, and S. Süsstrunk, "Material Classification Using Color and NIR Images," *Proc. IS&T/SID 17th CIC* (2009).

[6] C. Fredembach, N. Barbuscia, and S. Süsstrunk, "Realistic Skin Smoothing using the Near-Infrared," *Proc. IS&T/SID 17th CIC* (2009).

[7] C. Fredembach and S. Süsstrunk, "Automatic and Accurate Shadow Detection from (potentially) a Single Image using Near-Infrared Information," submitted to *IEEE Trans. on Pattern Anal. and Machine Intel.,* (2010).

[8] E.P. Bennett, J.L. Mason, and L. McMillan, "Multispectral Bilateral Video Fusion," *IEEE Trans. on Image Proc.,* **16**, (2007).

[9] X. Zhang, T. Sim, and X. Miao, "Enhancing Photographs with Near Infrared Images," *Proc. IEEE CVPR* (2008).

[10] D. Krishnan and R. Fergus, "Dark Flash Photography," *ACM Transactions on Graphics,* **28**, (2009).

[11] F. Durand and J. Dorsey, "Fast Bilateral Filtering for the Display of High-Dynamic-Range Images," *ACM Transactions on Graphics,* **21**, (2002).

[12] K. He, J. Sun, and X. Tang, "Single Image Haze Removal using Dark Channel Prior," *Proc. IEEE CVPR* (2009).

[13] R.T. Tan, "Visibility in Bad Weather from a Single Image," *Proc. IEEE CVPR* (2008).

[14] R. Fattal, "Single Image Dehazing," *ACM Transactions on Graphics,* **27**, (2008).

[15] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edgepreserving Decompositions for Multi-Scale Tone and Detail Manipulation," *ACM Transactions on Graphics,* **27**, (2008).

[16] A. Levin, D. Lischinski, and Y. Weiss, "A Closed-Form Solution to Natural Image Matting," *IEEE Trans. on Pattern Anal. and Machine Intel.*, **30**, (2008).