Automatic Skin Enhancement with Visible and Near-Infrared Image Fusion

Sabine Susstrunk
School of Computer and
Communication Sciences (IC)
Ecole Polytechnique Fédérale
de Lausanne (EPFL)
Lausanne, Switzerland
sabine.susstrunk@epfl.ch

Clément Fredembach *
School of Computer and
Communication Sciences (IC)
Ecole Polytechnique Fédérale
de Lausanne (EPFL)
Lausanne, Switzerland
clement.fredembach@epfl.ch

Daniel Tamburrino
School of Computer and
Communication Sciences (IC)
Ecole Polytechnique Fédérale
de Lausanne (EPFL)
Lausanne, Switzerland
daniel.tamburrino@epfl.ch

ABSTRACT

Skin tones, portraits in particular, are of critical importance in photography and video, but a number of factors, such as pigmentation irregularities (e.g., moles, freckles), irritation, roughness, or wrinkles can reduce their appeal. Moreover, such "defects" are oftentimes enhanced by scene lighting conditions.

Starting with the observations that melanin and hemoglobin, the key components of skin color, have little absorption in the near-infrared (NIR) part of the spectrum, and that the depth of light penetration in the epidermis is proportional to the incident light's wavelength, we show that near-infrared images provide information that can be used to automatically smooth skin tones in a physically realistic manner.

Specifically, we developed a prototype camera system that 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) with the fast bilateral filter. Smooth and realistic output images are obtained by fusing the base layer of the visible image with the near-infrared detail layer. The proposed method delivers consistently good results across various skin types.

The prototype system is currently in use at the Swiss Camera Museum in Vevey, Switzerland, where the visitors can take their pictures and e-mail themselves the results. In the process, we are collecting the users' preference for either the "original" (visible) image or the "enhanced" (visible and NIR fused) image. The system has been deployed for three months. Preliminary statistics indicate that a large majority (79%) prefers the enhanced image.

---

*Now with Canon Information Systems Research, Sydney, Australia

Categories and Subject Descriptors

I.4 [**Image Processing and Computer Vision**]: Applications

General Terms

Human Factors

Keywords

Skin smoothing, Skin rendering, Near-infrared, Bilateral Filter, Portrait photography, Video Conferencing, HD TV

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 visible (RGB) and NIR wavebands result in markedly dissimilar images that, when compared, offer powerful clues about the imaged scene. The inherent differences in RGB and NIR image intensities and frequencies allow much better disambiguating scene elements, which can be exploited in image enhancement [1, 3, 4, 10], material classification [8] and image segmentation [9]. Additionally, mathematically ill determined problems, such as multiple illuminant and shadow detection [5], become tractable.

Here, we present a specific application where near-infrared information can help us *enhance* images, specifically skin tones. Significant efforts are usually undertaken to make people look "good" in photographs, video, and television. Diffuse lighting and make-up are employed before and during the capturing process, and image manipulation programs are used after capture to remove perceived blemishes, such as pores, wrinkles, freckles, and spots.

NIR images are mostly free of these artifacts; skin appears much smoother (see Fig. 1b). 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
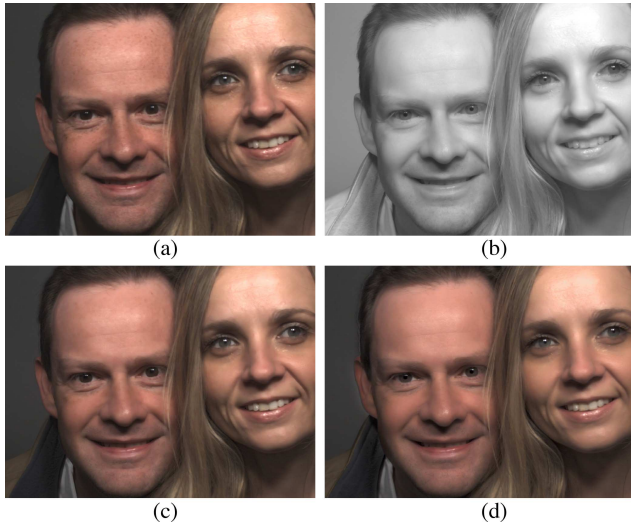
Figure 1: (a) Visible (RGB) image, (b) Near-infrared image, (c) Fused visible and NIR image with $\sigma_s = 4$ and $\sigma_r = 0.05$, (d) Fused visible and NIR image with $\sigma_s = 16$ and $\sigma_r = 0.1$.

part of the spectrum; the depth of light penetration in the skin is approximately proportional to the incident radiation's wavelength [6]. 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 [3] consists of capturing a pair of visible/near-infrared images and separating both of them into base and detail layers using the fast bilateral filter (see Section 2). A smooth and realistic output image can be obtained by fusing the base layer of the visible image with the near-infrared detail layer.

We believe that this method of automatic but still natural looking skin smoothing has large potential for photography, video conferencing, and television, especially considering the wide adoption of high definition (HD) displays. However, we still need to evaluate how the general public reacts to these augmented pictures. We have build a prototype camera system that automatically enhances skin tones (Section 3). We use a two-sensor camera with a beam splitter that can capture visible and NIR images concurrently. Based on the provided SDK, we wrote an application that allows automatic smoothing of the images. We have deployed the system at the Swiss Camera Museum[1] in Vevey, Switzerland. Visitors can take their picture and e-mail themselves the results. In the process, they have to indicate if they prefer the original or smoothed version of their portrait. We are collecting the results. The exhibit opened April 23, 2010. Preliminary statistics indicate that a large majority (79%) prefers the smoothed version (Section 4).

2. VISIBLE AND NIR IMAGE FUSION

Flushed skin, visible capillaries, rash, and acne are all features that are present on the skin surface. A naive approach to skin tone enhancement could be to low-pass filter the image such that these small scale features are smoothed out. However, faces also contain other high-frequency featu-

---

res, such as the distinction between skin, eye, iris, and pupil

res, such as the distinction between skin, eye, iris, and pupil that need to be retained. Additionally, hair-based features (e.g., hair, eyelashes, beard) also have to remain as sharp as in the original image.

Deeper light penetration combined with the relative transparency of hemoglobin and melanin to NIR result in an attenuation of the unwanted skin surface features in the NIR image, while the desired high-frequency information is still present (Fig. 1b). Given their size, most, if not all, of undesirable skin features will be located in the detail layer of the visible luminance image. Decomposing the NIR image will, on the other hand, yield a detail layer that contains all high-frequency information except for these undesirable features. A simple method to smooth the image is therefore to fuse the detail layer of the NIR image with the base layer of the visible luminance. Chrominance information is then added, and the image is transformed back into RGB. The entire procedure is illustrated in Fig. 2. The sample results, shown in Fig. 1c and 1d, exhibit natural-looking smooth images.

There are different image fusion algorithms that can be employed, we found that fusion based on bilateral filtering works best [3]. The bilateral filter, proposed by Tomasi and Manduchi [11], is an edge-aware spatial filter that can decompose an image into base and detail layers:

$$BF[I]_{\mathbf{p}} = \frac{1}{W_{\mathbf{p}}} \sum_{\mathbf{q} \in S} G_{\sigma_s}(\|\mathbf{p} - \mathbf{q}\|) G_{\sigma_r}(|I_{\mathbf{p}} - I_{\mathbf{q}}|) I_{\mathbf{q}}$$

where the intensity I at pixel position $\mathbf{p}$ is the weighted sum of two Gaussians: G_{σ_s} that controls the spatial weights, and G_{σ_r} that controls the range weights and ensures that the Gaussian filter does not average over strong edges, which can result in halo artifacts. The bilateral filter thus takes into account the difference in intensity between the different pixel values.

The base layer comprises mostly low frequency information with a small edge content, while the detail layer is primarily high-frequency information. Related to portrait images, the base layer is responsible for large, smooth, regions and the detail one for the stark and small-scale changes across these large regions (see Fig. 2).

The "amount" of smoothing, and thus the visual effect, is dependent on scale and can be controlled by adjusting the two parameters σ_s and σ_r. If they are large, more smoothing will be observed, but with the risk of obtaining halo artifacts. In the specific case of faces, the iris also becomes brighter (see Fig. 3). If the weights are small, the smoothing and iris brightening effect is less noticeable (compare Fig. 1c and 1d)[2]. In Fig. 3, we illustrate the effect of different σ_s and σ_r. Given that for our current prototype, the face to camera distance, the focal length of the camera, and its resolution is constant, we assume a fixed scale and have so far used a single set of parameters: $\sigma_s = 16$ and $\sigma_r = 0.1$ (Fig. 1d and Fig. 3, middle image).

A major drawback of the bilateral filter decomposition is its speed, we thus use here the fast bilateral filter proposed by Durand and Dorsey [2] with no significant decrease in image quality. In our software implementation, processing takes 0.5 seconds. To achieve frame rate necessary for video and HD applications, the fast bilateral filter can be implemented in a DSP.

---

[1]www.cameramuseum.ch

[1]www.cameramuseum.ch

[2]Note that the described effects are more visible if the images are viewed on a screen.
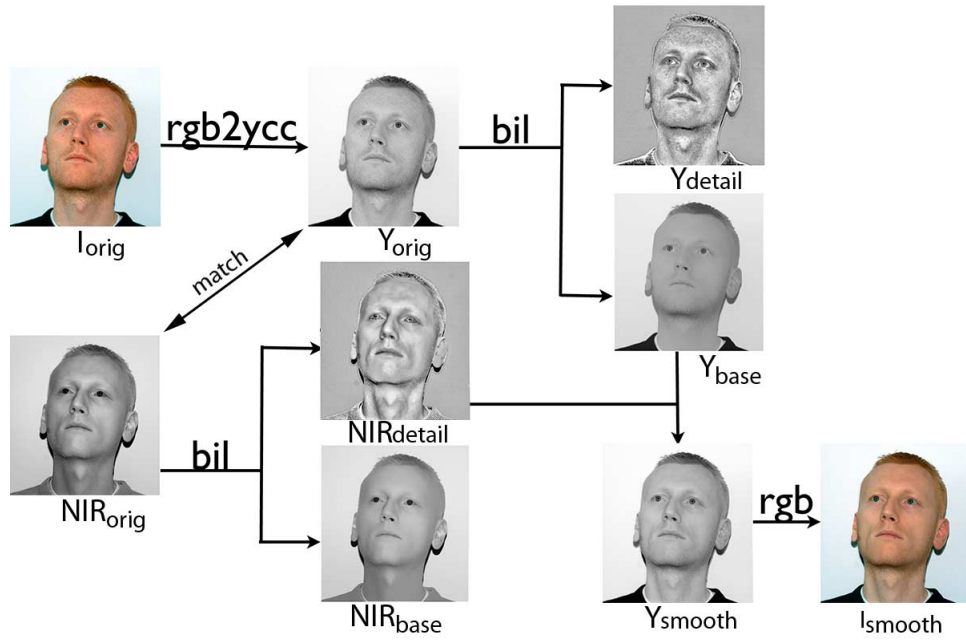
Figure 2: Flowchart detailing the bilateral filter based image fusion of visible and NIR images.
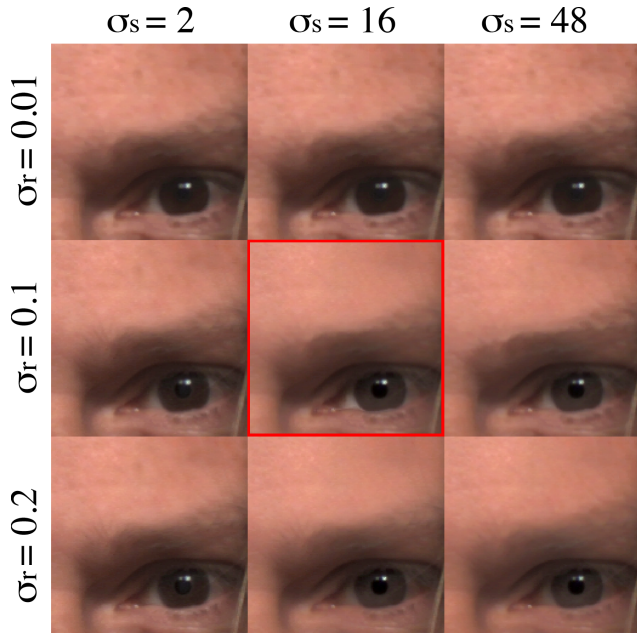

Figure 3: σ_s = **[2 16 48] from left to right and** σ_r = **[0.01 0.1 0.2] from top to bottom. The middle image (in red) is rendered with our current settings.**

3. PROTOTYPE AND DEMO

The face images created with the above method are thus enhanced without the need of "manual" intervention, such as make-up or post-processing. The wanted details, such as hair and eyelashes, are preserved, while skin irregularities are attenuated or removed altogether. The result is a "natural" looking enhanced image that should be preferred by the majority of users.

To test that hypothesis, we built a prototype of the system and deployed it in a public setting, namely as an installation in a newly opened exhibit entitled *Digital Revolution* at the Swiss Camera Museum in Vevey, Switzerland. To acquire visible and NIR images simultaneously, we use a JAI AD-080GE camera[3]. This 2CCD multi-spectral camera can simultaneously captures both color and NIR images in one camera housing and through a single lens. It uses a multi-faceted prism in the optical path with bandpass filters on each spectral axis to capture RGB through a Bayer Filter Array on one sensor and NIR on the other. The camera, primarily intended for industrial quality control, has a resolution of 1024 x 768 pixels and is able to capture 30 frames per second in full frame mode. An SDK is provided that we programmed to allow simultaneous RGB and NIR capture.

The installation[4] is set-up to resemble a photo booth, which was and is still used to take passport and other ID photos. Once the pictures are taken, both the visible and smoothed image are displayed on a screen, and the visitor can choose to have them sent via e-mail. However, before they are enabled to do so, they need to indicate if they prefer the original or the fused image. This extra step allows us collect valuable data about user preference by age and gender.

---

[3]http://www.jai.com
[4]http://ivrg.epfl.ch/research/topics/ACMdemo.html

Table 1: The number and percentage of times the enhanced image was preferred by *male* visitors.

Male	Total #	# Enh.	% Enh.
child (< 10 yrs.)	29	24	83%
adolescent (10-20 yrs.)	15	14	93%
young adult (21-40 yrs.)	40	30	75%
middle aged (41-60 yrs.)	107	76	71%
elderly (> 60 yrs.)	50	41	82%
Total	**241**	**185**	**77%**

Table 2: The number and percentage of times the enhanced image was preferred by *female* visitors.

Female	Total #	# Enh.	% Enh.
child (< 10 yrs.)	31	17	55%
adolescent (10-20 yrs.)	42	37	88%
young adult (21-40 yrs.)	77	66	86%
middle aged (41-60 yrs.)	64	53	83%
elderly (> 60 yrs.)	39	34	87%
Total	**253**	**207**	**82%**

4. EVALUATION

The exhibit opened April 23, 2010. Out of 1145 images taken so far, 1023 are included in this evaluation. The others (122) were pictures of something else than a face, groups of people making faces, or wrongly exposed pictures. Out of the 1023 usable images, 507 were duplicates (50%). That is an average of 2 pictures taken per person. In practice, groups of adolescents tend to take many pictures.

Out of the 516 non-duplicate portraits, 409 (79%) prefer the enhanced image. Note that the visitor is made aware which of the images is the original RGB image and which one is enhanced. Table 1 and 2 provides a summary of the results by gender and age. As we do not collect the age of the participants, the categories are somewhat arbitrary and based on manual judgment. 19 images could not be categorized, as they contain different gender. Of these, 16 (84%) also prefer the enhanced image.

Interestingly, there is not a large difference between male and female preferences, although women in general tend to prefer the enhanced image slightly more. For girls' portraits, the number is predictably close to chance (50%), as children usually do not yet have many skin artifacts. Yet, the same trend cannot be observed for boys.

The exhibit will continue until December 2011. In a next step, we will adapt the spatial σ_s and range σ_r of the bilateral filter to collect statistics when the smoothing effect is slightly lower than at its current setting.

5. CONCLUSION

We present an application for visible and NIR image fusion, namely automatic skin enhancement. NIR images contain less skin artifacts, they appear smoother while still retaining the wanted high frequency details, such as hair, beard, and eyebrows. The enhanced image is a fusion of the base layer of the visible image and the detail layer of the NIR image, obtained by bilateral filtering.

We believe that such an application is of interest to photography, video conferencing, and high-definition television. We have thus built a prototype based on a 2CCD camera that can simultaneously capture RGB and NIR on two diffe-

rent sensors. We have deployed the system in a local museum and are collecting the visitor's preference. Preliminary findings support our hypothesis that a large majority prefers the "enhanced" images.

Widespread adoption of the presented technique will depend on the development of a camera that can concurrently capture RGB and NIR images on a single sensor. Silicon, the light sensitive material of current digital camera sensors, is inherently capable of capturing both visible and NIR radiation. We have thus started to investigate [7] how such a camera system needs to be designed.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] E. Bennett, J. Mason, and L. McMillan. Multispectral bilateral video fusion. *IEEE Transaction on Image Processing*, 16:1185–1194, 2007.

[2] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *Proc. ACM SIGGRAPH*, 2002.

[3] C. Fredembach, N. Barbusica, and S. Süsstrunk. Combining visible and near-infrared images for realistic skin smoothing. In *Proc. IS&T/SID 17th Color Imaging Conference*, 2009.

[4] C. Fredembach and S. Süsstrunk. Colouring the near infrared. In *Proc. IS&T/SID 16th Color Imaging Conference*, 2008.

[5] C. Fredembach and S. Süsstrunk. Illuminant estimation and detection using near infrared. In *Proc. IS&T/SPIE Electronic Imaging, Digital Photography V*, volume 7250, 2009.

[6] T. Igarashi, K. Nishino, and S. K. Nayar. The appearance of human skin: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3:1–95, 2007.

[7] Y. M. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk. Designing color filter arrays for the joint capture of visible and near-infrared images. In *Proc. IEEE International Conference on Image Processing*, 2009.

[8] N. Salamati, C. Fredembach, and S. Süsstrunk. Material classification using color and NIR images. In *Proc. IS&T/SID 17th Color Imaging Conference*, 2009.

[9] N. Salamati and S. Süsstrunk. Material-based object segmentation using NIR information. In *Proc. IS&T/SID 18th Color Imaging Conference*, 2010.

[10] L. Schaul, C. Fredembach, and S. Süsstrunk. Color image dehazing using the near-infrared. In *Proc. IEEE International Conference on Image Processing*, 2009.

[11] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. IEEE 6th International Conference on Computer Vision*, 1998.