

Converting Near Infrared Facial Images to Visible Light Images using Skin Pigment Model

Kimshing Goh, Tetsu Matsukawa, Takahiro Okabe, Yoichi Sato
 Institute of Industrial Science, the University of Tokyo
 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan
 {kimshing, te2, takahiro, ysato}@iis.u-tokyo.ac.jp

Abstract

In this paper, we propose a physics-based method to synthesize facial images in visible wavelengths from multi-band near infrared (NIR) images. The study on photometric properties of human skin shows that melanin and hemoglobin components are dominant factors that affect the skin appearance under different light spectrum. Specifically, a set of intensities observed at a certain surface point with varying wavelength is represented by a linear combination of both the pigment basis vectors, which describe absorbance due to both the pigments, from multispectral image dataset by using Independent Component Analysis (ICA). Then, our method estimates the coefficients, which are pixel-wise densities of both the pigments, from a multiband NIR image, and finally converts it to a visible light (VIS) image. We demonstrate that our proposed method works well for real facial images even though only a small dataset is available for learning basis vectors.

1 Introduction

One of the major challenges in face recognition is how to deal with images taken under varying illumination or low-lit images [1].

To alleviate the problem, Li et al. [2] proposed a system using active near infrared flash to acquire illumination invariant facial images. However, their system is based on the strong assumption that NIR images are available as gallery images as well as probe images. This limits the applicability of the system because NIR images are often unavailable for gallery images, e.g., photos on a passport or a driving license. In such cases, probe NIR images are matched against VIS gallery images, resulting in poor recognition accuracy because the appearance of a face changes significantly between NIR and VIS images.

Several methods for matching NIR images to VIS images have been proposed. One approach focuses on extracting features invariant across both NIR and VIS spectra [3, 4]. Another approach try to synthesize a VIS image from a NIR image [5–8]. The advantage of the latter approach is that existing face recognition systems can be used with no modification. For this reason, we focus on the latter approach and, in particular, we propose a method for converting NIR images into VIS images based on the photometric properties of human skin.

Chen et al. [5] proposed a method for NIR to VIS image conversion based on local linear embedding. Each image patch is approximated by a weighted sum of their k nearest neighbors of training NIR patches

using local binary pattern (LBP) similarities. Then a VIS image is synthesized by using the same weights and corresponding VIS patches. Zhang et al. [6] extended this idea by using sparse representation. Shao et al. [7] learned the relationship of VIS and NIR images by using a multifactor analysis. Similarly, Zhang et al. [8] learned the relationship of NIR and quotient images. All of these methods share the same problem that a large number of patch pairs are necessary, e.g., patches collected from more than 100 individuals, to produce satisfactory results. This is because these methods try to convert NIR patches to VIS patches without considering the underlying physical phenomenon.

Human skin is a multi-layered structure with various pigments. Melanin and hemoglobin pigment are dominant pigments that affect skin appearance [9, 10]. Tsumura et al. [11] proposed a technique to extract melanin and hemoglobin bases and densities from a RGB spectral by using ICA [12]. They synthesized various skin colors such as tanning and alcohol consumption by changing the extracted pigment densities. However, their analysis was limited to RGB color channels, and no discussion was made for skin appearances in the NIR spectrum.

In this paper, we propose a method for converting NIR images to a VIS images on the basis of the skin pigment model. Our method estimates the coefficients, which are pixel-wise densities of two dominant pigments, from a multiband NIR image, so at least two images under different spectra in NIR¹ is needed. From the estimated pigment density and spectral bases, we can synthesis the images of VIS spectral.

As far as we know, this work is the first time attempt to extract skin pigment density from multispectral NIR images and synthesize to VIS images by using skin pigment model. Through the experiment, we confirm that our proposed pixel-wise density based prediction is able to synthesize VIS images with little training samples.

The rest of this paper is organized as follows. In Section 2, the skin pigment model is briefly introduced. In Section 3, our proposed framework for synthesizing a VIS image from multi-band NIR images is described. In Section 4, results of our experiments are shown and discussed, followed by a conclusion in Section 5.

2 Skin Pigment Model

The color of human skin changes with respect to subsurface scattering of dermal and epidermal layers

¹The use of special equipment for capturing multispectral NIR images could be a limitations of our proposed method from a practical point of view. Recently, however, such equipment is getting more popular in the field of multi-spectral imaging [13, 14], and could be used for our purpose.

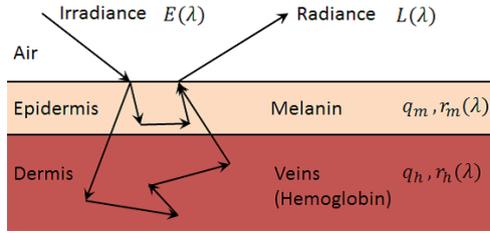


Figure 1. Light transport model of epidermal and dermal layers.

[9, 10]. Fig. 1 shows a light transport model of epidermal and dermal layers, where subsurface scatter of both layers are well-considered. Melanin and hemoglobin pigments are predominantly contained in these layers. The subsurface reflectance of these layers are modeled by modified Lambert-Beer law [11] as,

$$L(\lambda_n) = \exp\{-q_m r_m(\lambda_n) - q_h r_h(\lambda_n)\} E(\lambda_n), \quad (1)$$

where λ_n is the n -th wavelength of incoming/outgoing light, $E(\lambda)$ and $L(\lambda)$ are the spectral distributions of incoming irradiance and outgoing radiance, respectively, and q_m , q_h , $r_m(\lambda)$, $r_h(\lambda)$ are the pigment densities and absorbance coefficients of melanin and hemoglobin, respectively.

Taking the logarithm of Eq. (1), we obtain the following additive form,

$$c(\lambda_n) = q_m r_m(\lambda_n) + q_h r_h(\lambda_n) + b(\lambda_n), \quad (2)$$

where $c(\lambda) = -\log(L(\lambda))$, and $b(\lambda) = -\log(E(\lambda))$. This equation shows the linear relationship of irradiance on logarithmic domain and melanin and hemoglobin pigments. In other words, if we observe a certain surface point of a skin with N spectral wavelength, the N -dimensional multi-spectral observation of logarithmic radiance $[c(\lambda_1) - b(\lambda_1), c(\lambda_2) - b(\lambda_2), \dots, c(\lambda_N) - b(\lambda_N)]$ lies in two dimensional subspace spanned by $[r_m(\lambda_1), r_m(\lambda_2), \dots, r_m(\lambda_N)]$ and $[r_h(\lambda_1), r_h(\lambda_2), \dots, r_h(\lambda_N)]$.

3 Skin Pigment-based Image Synthesis

In this section, we describe the proposed synthesis method based on skin color model introduced in Section 2, which is applicable in both VIS and NIR spectra. The conceptual diagram of the proposed method is shown in Fig. 2. We use multispectral images which cover VIS and NIR spectrum. We denote logarithmic of N pixel values obtained along wavelength axis at position (i, j) as a column vector, $\mathbf{c} = [c(\lambda_1), c(\lambda_2), \dots, c(\lambda_N)]^t$. Then, we rewrite Eq. (2) by vector and matrix formulation as follows:

$$\mathbf{c}(i, j) = \mathbf{R}\mathbf{q}(i, j) + \mathbf{b}, \quad (3)$$

where $\mathbf{R} = [\mathbf{r}_m, \mathbf{r}_h]$, $\mathbf{r}_m = [r_m(\lambda_1), r_m(\lambda_2), \dots, r_m(\lambda_N)]^t$, $\mathbf{r}_h = [r_h(\lambda_1), r_h(\lambda_2), \dots, r_h(\lambda_N)]^t$, $\mathbf{q} = [q_m, q_h]^t$, and $\mathbf{b} = [b(\lambda_1), b(\lambda_2), \dots, b(\lambda_N)]^t$ is a constant bias vector caused by irradiance. In this work, we assume constant illumination condition throughout our experiments, so, b can be calculated as $b(\lambda_n) = \min_{i,j}(\log(c(i, j, \lambda_n)))$ for each wavelength.

Fig. 3 shows the framework of our proposed method composed of 3 main steps; these are spectral basis learning, pigment densities estimation, and VIS images synthesis.

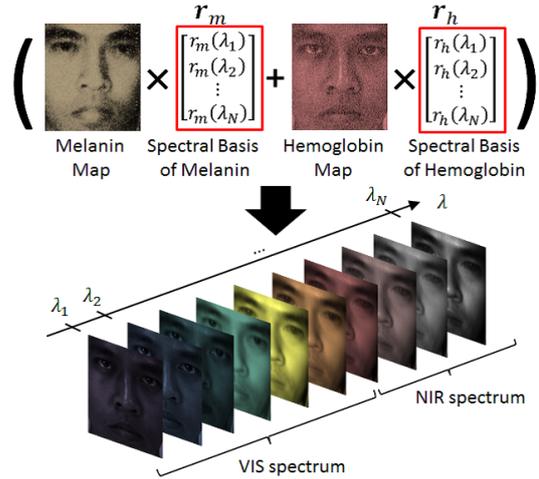


Figure 2. Conceptual diagram of the proposed method.

3.1 Learning spectral basis

A dataset composed of N spectral images of various subjects are first acquired from VIS and NIR spectra. Skin patches with size of $W \times H$ are clipped from K subjects to form the training set. For each spectral, patch is expressed as a $P(=WH)$ dimensional column vector. Then patches from all subjects are concatenated to form a N -by- PK data matrix \mathbf{C} . Fig. 4 shows the flow chart of the decomposition of this data matrix into two matrices which are corresponding to pigment densities and absorbance coefficients.

To extract two-dimensional subspace of the dominant pigments, principal component analysis (PCA) is firstly applied [11]. Fig. 5 shows the cumulative contribution ratio of 6 principal components obtained from the 6 spectral data used in the experiment. This figure shows that two principal components are sufficient to describe the values of 6 spectra with accuracy as high as 96.8%.

Then, the input data is projected onto the two-dimensional subspace spanned by eigenvectors corresponding to first and second principal components.

We then perform ICA [12] on projected data to decompose the data matrix into 2 matrices, which are mixing matrix \mathbf{R} and densities matrix \mathbf{q} . Note that mixing matrix \mathbf{R} is composed of spectral basis of melanin \mathbf{r}_m and hemoglobin \mathbf{r}_h , that describe absorbance coefficient of both pigments at specific wavelength.

Each spectral basis vector is divided into two subvectors based on NIR and VIS spectrum region. Subvector on NIR region \mathbf{R}_{nir} is used for estimating density map of melanin and hemoglobin. On the other hand, subvector on VIS region \mathbf{R}_{vis} is used in synthesis step. Similarly, the bias vector obtained from the training set is divided into subvectors and used for following steps.

3.2 Estimation of density maps

Given at least two NIR images taken under different wavelengths, density map of melanin and hemoglobin densities can be estimated using least square estimation as follows:

$$\hat{\mathbf{q}}(i, j) = \underset{\mathbf{q}(i, j)}{\operatorname{argmin}} \|\mathbf{c}_{nir}(i, j) - \mathbf{R}_{nir}\mathbf{q}(i, j) - \mathbf{b}_{nir}\|^2, \quad (4)$$

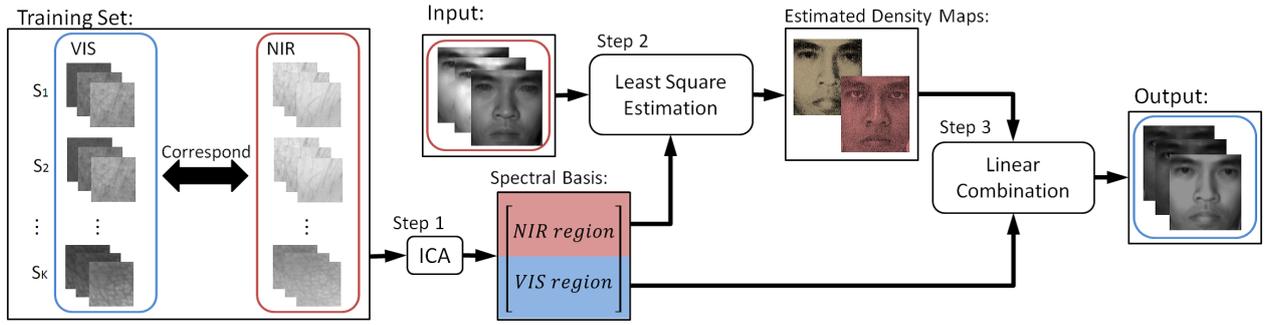
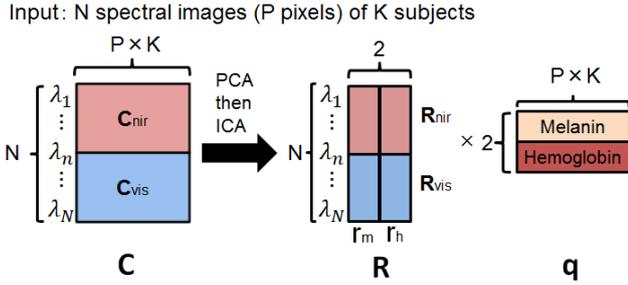


Figure 3. Schematic flow of the proposed method.



Output: Spectral basis of melanin & hemoglobin ($2 \times N$)

Figure 4. Decomposition of data matrix into melanin and hemoglobin factors.

where $\hat{q}(i, j)$ is the estimated pigment densities at position (i, j) on image, c_{nir} is logarithmic values of input NIR images, and b_{nir} is the subvector of bias vector in NIR region.

3.3 Synthesizing VIS images

For synthesis step, we use the estimated \hat{q} and R_{vis} in Section 3.1 to synthesize images in VIS region. The synthesis process is indicated as

$$c_{vis}(i, j) = R_{vis}\hat{q}(i, j) + b_{vis}, \quad (5)$$

where c_{vis} is logarithm values of output VIS images, and b_{vis} is the subvector of bias vector in VIS region. We then find the inverse log of c_{vis} and reshape it back to image size.

Since we estimate and synthesis in a pixel-wise manner using common spectral basis of melanin and hemoglobin over a face, registration of face patch is not needed. Therefore, our method can be also applicable to facial images with different expressions and orientations to the training set.

4 Experiments

In this work, we set $N=6$, where we use 3 NIR images to synthesize 3 VIS images (RGB-channels) to confirm the validity of our proposed method. Three NIR images are used instead of two to reduce the density estimation error. We implemented a multispectral imaging system composed of a camera (Chameleon CMLN-13S2M-CS) and a filter wheel with 6 band-pass filters attached to it. Band-pass filters used are of narrowband range with center wavelength at 450nm, 532nm, 610nm, 766nm, 880nm, and 960nm, respectively. Facial images of 8 subjects are acquired at 6 wavelength mentioned above. Face region of images

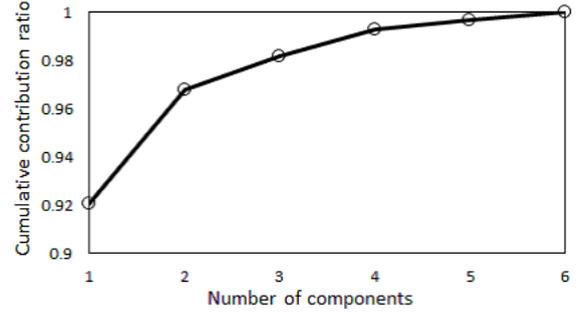


Figure 5. Relationship between the number of components and the cumulative contribution ratio in skin image set of 6 spectra.

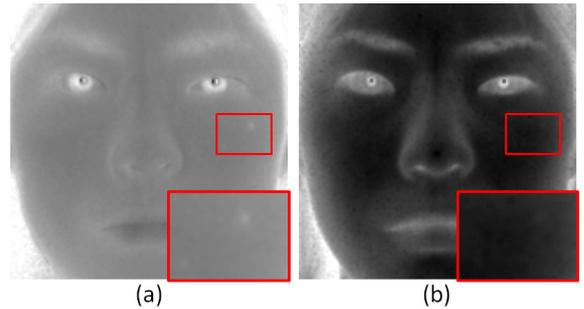


Figure 6. Estimated pigment densities corresponding to two independent components. Lower intensity indicates lower pigment density.

are clipped into the size of 200×200 . Training set is composed of patches clipped from face region of frontal facial images with neutral expression. For evaluation, facial images of the same subjects include different orientations and expressions to training set are acquired, and we learn spectral bases from different subjects to input NIR images.

Fig. 6 shows the result of pigment densities estimation based on two learned spectral bases of melanin and hemoglobin. Fig. 6(a) shows lower concentration of pigment at the lip region and higher concentration of pigment at mole spot on the right cheek. This estimation result agrees well with the physiological facts of melanin [9]. Similarly, Fig. 6(b) shows higher concentration of hemoglobin pigment on lip region. However, because the pigment model holds only for skin region, pigment densities at the region of eyes and eyebrows are not estimated correctly.

Fig. 7(a) shows synthesized VIS images by using two pigment densities shown in Fig. 6. The synthesized images look fine with details such as mole on the right cheek preserved. Synthesized results for probe images of side view and with expression are shown in Fig. 7(b)

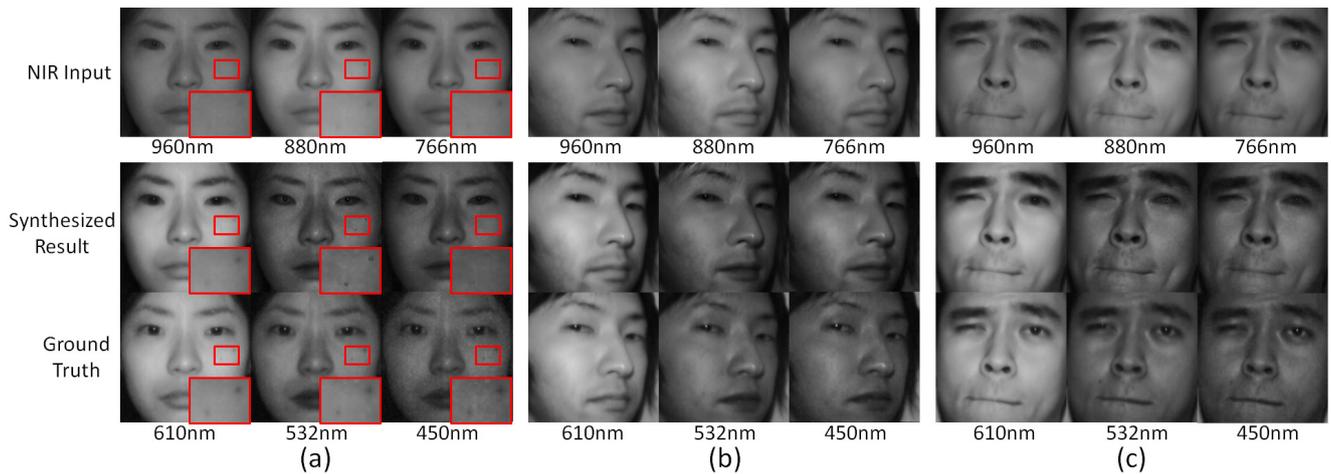


Figure 7. Synthesized VIS facial images (a)of frontal view, (b)of side view, (c)with expression.

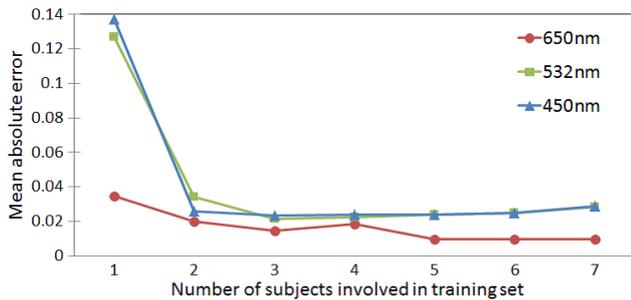


Figure 8. Mean absolute error of synthesized images in different number of subjects involved in training set.

and Fig. 7(c), respectively. Note that although our training set contains only skin patches of frontal view and neutral expression, our proposed method is able to convert probe NIR images of different orientations and expressions to RGB images. However, throughout our experiments, the appearance of eyes region is not synthesized to be what expected under VIS spectrum. This occurred because our pigment model holds only for skin region. It can be refined by separating eyes region from facial images and applying different synthesis algorithm for that region.

Fig. 8 shows the mean absolute errors of synthesized images in different number of subjects to be involved in training set. Errors are evaluated by averaging the absolute difference between synthesized image and ground truth image of all pixels. Error of the synthesized images are reduced as to increase the number of subjects involved in training set. Because spectral bases are interpersonally invariant, the use of skin patches clipped from 2-3 subjects are sufficient for our learning process.

5 Conclusion

We proposed a framework for converting facial images from multi-band NIR images to VIS images based on photometric properties of human skin. Experimental results showed that our proposed method works well for real facial images with different orientations and expressions even when only a small dataset is available for learning.

For future work, the following issues still remain to be addressed: applications to different illumination

conditions of spectral distribution or direction, and the use of camera with different spectral sensitivity.

References

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. "Face recognition: A literature survey," *ACM-CSUR*, Vol. 35, No. 4, pp. 399–458, 2003.
- [2] S. Z. Li, R. F. Chu, S. C. Liao, and L. Zhang. "Illumination invariant face recognition using near-infrared images," *IEEE Trans. PAMI*, Vol. 29, No. 4, pp. 627–639, 2007.
- [3] D. Yi, R. Liu, R. F. Chu, Z. Lei, and S. Li. "Face matching between near infrared and visible light images," *Advances in Biometrics*, pp. 523–530, 2007.
- [4] B. Klare and A. K. Jain. "Heterogeneous face recognition: Matching NIR to visible light images," In *Proc. ICPR*, pp. 1515–1516, 2010.
- [5] J. Chen, D. Yi, J. Yang, G. Zhao, S. Z. Li, and M. Pietikainen. "Learning mappings for face synthesis from near infrared to visual light images," In *Proc. CVPR*, pp. 156–163, 2009.
- [6] Z. Zhang, Y. Wang, and Z. Zhang. "Face synthesis from near-infrared to visual light via sparse representation," In *Proc. IJCB*, pp. 1–6, 2011.
- [7] M. Shao, Y. Wang, and Y. Wang. "A super-resolution based method to synthesize visual images from near infrared," In *Proc. ICIP*, Vol. 2, pp. 2453–2456, 2009.
- [8] Z. Zhang, Y. Wang, Z. Zhang, and G. Zhang. "Face synthesis from near-infrared to visual light spectrum using quotient image and kernel-based multifactor analysis," In *Proc. ICME*, pp. 1–4, 2011.
- [9] S. B. Kusse. "Spectral imaging and analysis of human skin," *Master Thesis, University of Eastern Finland*, 2010.
- [10] E. Angelopoulou. "Understanding the color of human skin," In *Proc. SPIE Conf. Human Vision and Electronic Imaging VI*, Vol. 4299, pp. 243–251, 2001.
- [11] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake. "Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin," *ACM Trans. Graph.*, Vol. 22, No. 3, pp. 770–779, 2003.
- [12] A. Hyvärinen and E. Oja. "Independent component analysis: algorithms and applications," *Neural Networks*, Vol. 13, No. 4–5, pp. 411–430, 2000.
- [13] Y. Suzuki, K. Yamamoto, K. Kato, M. Andoh, and S. Kojima. "Skin detection by near infrared multi-band for driver support system," In *Proc. ACCV*, pp. 722–731, 2006.
- [14] K. Sakashita, Y. Yagi, R. Sagawa, R. Furukawa, and H. Kawasaki. "A system for capturing textured 3D shapes based on one-shot grid pattern with multi-band camera and infrared projector," In *Proc. 3DIMPVT*, pp. 49–56, 2011.