Assessing Human Skin Color from Uncalibrated Images

J. Marguier,[1] N. Bhatti,[2] H. Baker,[2] M. Harville,[2] S. Süsstrunk[1]

[1] School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[2] Hewlett-Packard Laboratories, Palo Alto, CA

ABSTRACT: Images of a scene captured with multiple cameras will have different color values because of variations in color rendering across devices. We present a method to accurately retrieve color information from uncalibrated images taken under uncontrolled lighting conditions with an unknown device and no access to raw data, but with a limited number of reference colors in the scene. The method is used to assess skin tones. A subject is imaged with a calibration target. The target is extracted and its color values are used to compute a color correction transform that is applied to the entire image. We establish that the best mapping is done using a target consisting of skin colored patches representing the whole range of human skin colors. We show that color information extracted from images is well correlated with color data derived from spectral measurements of skin. We also show that skin color can be consistently measured across cameras with different color rendering and resolutions ranging from 0.1 to 4.0 megapixels. © 2007 Wiley Periodicals, Inc. Int J Imaging Syst Technol, 17, 143–151, 2007; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ima.20114

Key words: digital photography; color assessment; color calibration target; uncalibrated imaging systems

1. INTRODUCTION

Digital images have become a part of our everyday life and imaging devices are everywhere, from higher end digital cameras to video cameras integrated in laptops or cell phones. Taking pictures has never been so easy, but objective color assessment remains an issue, especially with the growth of low-quality devices such as cell phone integrated cameras. Because of imperfect illuminant compensation and to variable camera characteristics, the same scene imaged with different devices can have important variations, which makes the objective assessment of scene colors difficult. This is a problem for online shopping, for example. It is impossible to accurately assign a color from an uncalibrated image displayed on an uncalibrated display, but using calibrated devices is neither practical nor feasible. The method we propose only requires that a calibrated target be present in the scene, which is by far the cheapest calibrated "hardware" to distribute.

The object of interest is imaged together with a reference target, which allows color correcting the image independently of the imaging device and illuminant. There are potentially many applications, particularly in the assessment of skin tones. In fashion, allowing people to upload a portrait taken with any capture device would enable automated suggestion of personal appearance products, such as makeup or clothing that complement skin tone. Our method could also be applied in computer graphics for skin color preservation in video conferencing or for the rendering of people's avatars in the next generation of online gaming.

The color properties of an object can be fully characterized by its reflectance spectra, i.e. the percentage of light reflected by the object's surface at each wavelength and incident angle. The reflectance spectra can be retrieved by direct spectrometric measurement or by using a hyperspectral device. Adding constraints on the reflectance smoothness allows defining the reflectance as a sum of a small number of basis functions and hence reduce the number of unknowns. In this framework, multichannel systems have been proposed (Tominaga, 1999). However, for many applications it is sufficient to retrieve only tristimulus values, which can be achieved using an RGB camera. Several approaches using calibrated trichromatic imaging systems have been presented (Farell et al., 1994; Wu et al., 2000).

In this article, we investigate the possibility of retrieving color information from a single digital picture taken with an unknown, casually posed consumer camera and under unknown lighting conditions, using solely a calibrated reference target present in the scene. The calibration target is imaged with the object of interest. The extracted target values allow computing a color correction matrix that is scene and camera dependent.

This transform is applied to the entire image. After color correction, the image is in a reference space. Face pixels are then extracted, averaged, and the skin color classified. The color correction is limited to a small color gamut, in our example skin tones. The system relies on the assumption that any camera output color image encoding is *sRGB* (IEC 61966-2-1, 1999).

We show that using an appropriate target representative of skin reflectances, skin colors can be mapped using a 3×4 linear transform with an accuracy in terms of *CIELAB* color difference of

---

Correspondence to: J. Marguier; e-mail: joanna.marguier@epfl.ch

$\Delta E^*_{ab} < 1$. These extracted skin color values have high correlation with skin colors derived from in vivo spectral measurements. We also show that skin color can be consistently predicted across a wide range of cameras from different manufacturers with different quality levels. With our method, we can use any uncalibrated camera as a colorimeter and determine skin tones accurately. No calibration of the device, control of the illuminant, or access to sensor raw data is needed.

The article is structured as follows: In Section 2 we discuss methods to retrieve reflectance spectra and tristimulus values using imaging devices, and present a brief review of color measurements applied to skin. Section 3 presents our approach. Section 4 describes the experiments and reports the main results. Section 5 concludes the article.

2. STATE OF THE ART

The radiation falling on a sensor is the product of $E(\mathbf{x}, \lambda)$, the spectral power distribution of the illuminant with $S(\mathbf{x}, \lambda)$, the reflectance spectra of the object. The camera response $\rho_i(\mathbf{x})$ of the ith sensor $R_i(\lambda)$ at a spatial position $\mathbf{x} = (x, y)$ is given by the image formation model

$$\rho_i(\mathbf{x}) = \mathbf{s}(\mathbf{x}, \lambda)^{\mathrm{T}} \cdot diag(\mathbf{e}(\mathbf{x}, \lambda)) \cdot \mathbf{r}_i(\lambda), \qquad i = 1 : n \quad (1)$$

where the vectors $\mathbf{s}(\mathbf{x}, \lambda)$, $\mathbf{r}_i(\lambda)$, and $\mathbf{e}(\mathbf{x}, \lambda)$ are, respectively $S(\mathbf{x}, \lambda)$, $R_i(\lambda)$, and $E(\mathbf{x}, \lambda)$ represented by m samples (usually 31) taken over the visible spectral range (Smith et al., 1992). $diag(\mathbf{e}(\mathbf{x}, \lambda))$ is an $m \times m$ matrix with $\mathbf{e}(\mathbf{x}, \lambda)$ on its diagonal and n is the number of channels of the imaging device.

It is not a trivial task to retrieve reflectance values from camera responses, especially when $n \ll 31$. For many applications, however, it is sufficient to retrieve colorimetric values instead of the entire reflectance spectra. The human visual system is indeed unable to recover spectral information, and two objects having the same appearance under a given illuminant can have slightly different reflectance spectra $S(\lambda)$. A three-channel camera is then sufficient to retrieve tristimulus values. Several approaches using RGB cameras as colorimeters have been proposed. Wu et al. (2000) used a calibrated camera to compute calibration matrices from camera *RGB* to *CIEXYZ* by either minimizing a cost function in *CIELAB* space or by minimizing the mean square error in *CIEXYZ* color space under several selected illuminant conditions. The application was the colorimetry of human teeth. Hubel et al. (1997) present a method to compute 3×3 color transform matrices intended for camera calibration in digital photography by simple least squares regression, white point preserving least squares regression, and weighted white point preserving least squares regression. This type of approach allows using a calibrated camera as a colorimeter under known illuminant conditions.

Such calibration methods require the access to the raw data of the sensors and the resulting color transform is camera dependent. The transform is usually applied prior to the image rendering implemented in the camera. In our method, we apply the transform *after* color rendering. As such, we need no information about sensor characteristics and in-camera processing, but we need to calculate a transform for every single image. Our transforms are scene and camera dependent.

Camera calibration is generally performed using a standard calibration target, such as the widely used Macbeth ColorChecker (Barnard and Funt, 2002). We propose to color correct skin tones using target embedded in the image, consisting primarily of color patches

characteristic of human skin. Cai (2002) presented a similar approach for the determination of tongue color using a Macbeth ColorChecker.

A. Skin Color Measurement. Skin appearance has been studied in various domains: in computer graphics (for skin rendering), in computer vision (for detection and tracking of faces), in medicine (for diagnostic purposes), and in cosmetology (for makeup and skin care). In particular, it has been shown that lightness varies more than chromaticity across the range of possible human skin tones (Störring et al., 2001). The relative constancy of the skin chromaticity is often used as a cue in face detection algorithms (Hsu et al., 2002).

Skin color is most often measured with reflectance spectrometry (for a review see Igarashi et al., 2005). Perceived color of skin depends on the pigmentation, the blood microcirculation, roughness, sebum, and perspiration (Barel et al., 2001). The multiple layers of the skin make accurate color measurement difficult as the reflection is direction dependent (Igarashi et al., 2005). Even though they are not optimal for skin measurements, traditional spectrometers are inexpensive and simple to use and hence still widely employed for skin colorimetry (Clarys et al., 2000). Narrow band spectrometers were developed specifically for skin measurement after observing that the color of skin has two main components, the melanin and hemoglobin, which selectively absorb wavelengths (Taylor et al., 2006). The use of spectrometers has important drawbacks: the area measured is about 0.05 cm^2 while the skin is not homogeneous (Barel et al., 2001). The pressure of the probe on the skin can be an important source of bias (Piérard, 1998).

A proprietary device composed of an integrating sphere, a spectrometer, and a tri-CCD camera was developed by Caisey et al. (2006), allowing noncontact spectroscopy of different parts of the face and simultaneous imaging for estimation of the skin unevenness.

The melanin and hemoglobin components have been modeled as independent and extracted using in vivo measurements by Nakai et al. (1998). Tsumura et al. (1999) applied the independent component analysis method to the pixel values of color images of faces and also extracted two independent color components corresponding to melanin and hemoglobin. These are indeed located in two different layers of the skin. The results were used for skin color reproduction and rendering, but not skin color classification as we address here.

3. OUR APPROACH

We propose a method to retrieve skin color information from digital images taken with a single, casually posed consumer camera under unknown illuminants. We have no access to the raw data of the sensors and no additional information on the automatic in-camera processing. However, the scene does contain reference color values in the form of a color target. We want to color correct images such that we can retrieve *sRGB* values for the face pixels. Figure 1 shows the entire pipeline (Harville et al., 2005).

The method is targeted toward consumer applications. The user is assumed to use an unknown camera in fully automatic mode and under uncontrolled illuminant. We suppose that the automatic camera mode performs white-balancing and encodes images in *sRGB*, which has a defined illuminant of D65.

Because of imperfect illuminant compensation, different sensor responses, variations in image processing, and in quality across devices, uniform color rendering is never achieved. The resulting image color values of a given scene imaged with different
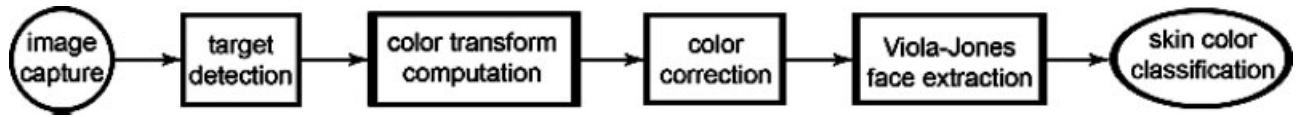
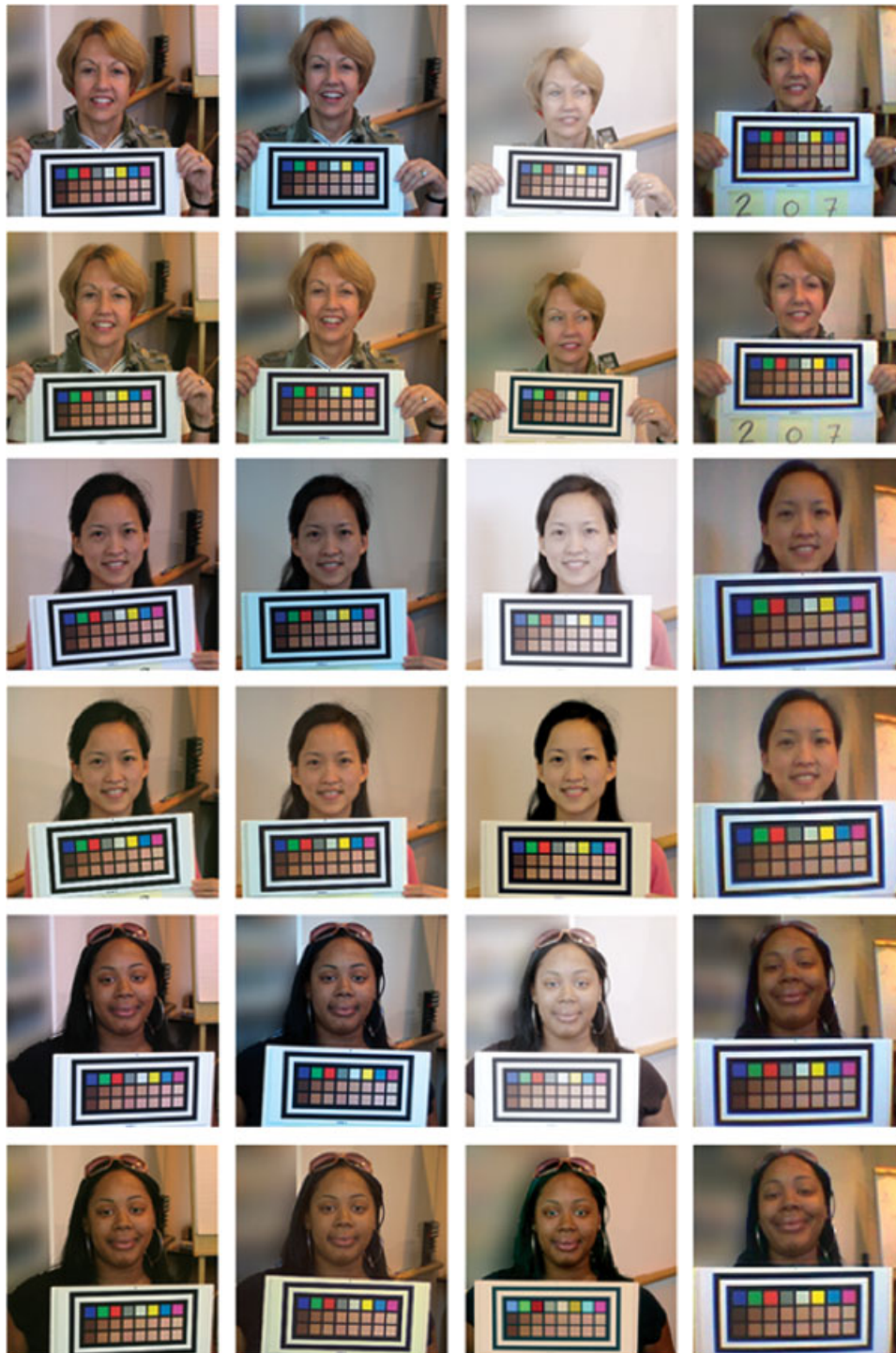**Figure 1.** Overview of image processing pipeline for skin color measurement and classification.

Figure 2. Uncorrected images (odd rows) and corrected images (even rows) for cameras (from left to right) Canon S400, HP850, Nikon D1, and Nokia 6820.
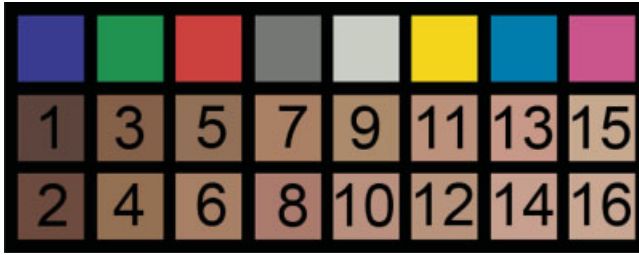
Figure 3. Calibration target, the skin patches are numbered by increasing lightness. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

uncalibrated cameras will have important variations. Figure 2 shows examples of possible variations in color across devices and the results after color correction.

To be able to classify colors consistently, we need the same object to have equal *sRGB* pixel values independently of the illuminant and the camera. We thus compute a color transform using known reference values, present in the scene in the form of a calibration target. The patches are extracted, their color values are averaged over each patch, and their mean values are compared with reference values of the target given by *sRGB* triplets. A 3×4 color transform **A** maps the target patches mean color values **M** extracted from the image onto reference target values **T**.

$$\mathbf{T}_{\{3 \times n\}} = \mathbf{A}_{\{3 \times 4\}} \cdot \mathbf{M}_{\{4 \times n\}} \qquad (2)$$

where **T** is a matrix whose ith column contains the ith value of the n reference patches $\mathbf{t}_i = (t_i^{red}, t_i^{green}, t_i^{blue})^T$ and **M** is a matrix whose ith column contains the ith value of the n mean camera patch color $\mathbf{m}_i = (m_i^{red}, m_i^{green}, m_i^{blue}, 1)^T$.

We want to find **A** minimizing $\|\mathbf{T} - \mathbf{AM}\|_2$, i.e. minimizing the least mean square error in *sRGB* color space. **A** is computed using the Moore–Penrose pseudo-inverse denoted $^+$. Right-multiplying Eq. (2) by $\mathbf{M}^+ = \mathbf{M}^T(\mathbf{MM}^T)^{-1}$ gives $\mathbf{TM}^+ = \mathbf{AMM}^+ = \mathbf{A}$. Finally, we obtain

$$\mathbf{A} = \mathbf{TM}^+ \qquad (3)$$

The pseudo-inverse of **M** is computed by singular value decomposition. **A** provides a 3×3 color transform plus a per-component offset. The linear transform **A** is recomputed for each image and will thus differ depending on the camera characteristics and lighting conditions.

Note that it is not possible to correctly map all colors with a simple linear transform. Minimizing the least mean square error in *sRGB* is computationally fast and simple, but does not give any control on the error repartition. Consequently, we apply the method only to a limited gamut of colors. In this framework, skin tones can be accurately assigned from uncalibrated images. Colors outside the considered gamut may not be correctly mapped and the overall color appearance of the image may be poor (see Fig. 2).

B. Reference Target. The target contains three rows of eight patches set against a black background and surrounded by a frame used for its automatic detection. The first row contains primary and secondary colors and two shades of gray. The two last rows contain 16 patches characteristic of the range of human skin ordered by uni-

formly increasing lightness alternating on two rows. Figure 3 shows the target.

We printed the target on photopaper medium and measured the reflectance spectrum of each patch. Figure 4 shows the desired patch spectra and the actual printed target spectra. These measures allow computing first *CIEXYZ* values under illuminant D65 and then *sRGB* values. With reference to Eq. (1), $\mathbf{s}(\lambda)$ are the target reflectances, $\mathbf{e}(\lambda)$ is the standard CIE illuminant D65, and $\mathbf{r}_i(\lambda)$ are the 1931 $CIE\,\overline{x}, \overline{y}, \overline{z}$ color matching functions. The *CIEXYZ* to *sRGB* transform is specified in (IEC 61966-2-1, 1999). For each image, the target patches are extracted and the color transform **A** (3) is computed. This transform is then applied to the entire image prior to face pixels extraction and color classification (Harville et al., 2005).

4. EXPERIMENT

Fifty-three people were imaged holding a copy of the calibration target, captured with four different RGB cameras. The cameras were an HP850 (3.9 megapixels), a Nikon D1 (2.7 megapixels), a Canon S400 (4.0 megapixels), and a Nokia 6820 cell phone camera (0.1 megapixels). The images contain mixed illuminants: a side daylight illumination through a window and an overhead fluorescent lighting. The images are white-balanced and color rendered in automatic mode, and the output color encoding is *sRGB*.

Figure 2 shows three examples of before and after color correction computed using the 16 skin tones. Rows 1, 3, and 5 show the uncorrected images having resolutions from 0.1 to 4.0 megapixels and different color qualities, and rows 2, 4, and 6 the same images corrected using the method described in Section 3. Table I shows the corresponding mean face color values extracted from the color corrected images and the equivalent color correction magnitude in normalized color coordinates.

To investigate the accuracy of our method, we first study the colorimetry of the target color patches to determine which color patches should be used to compute the color transform **A**. We then
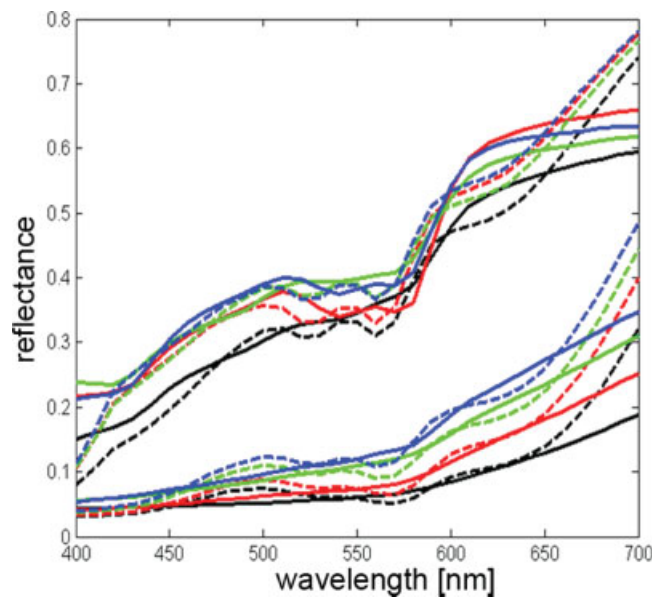

Figure 4. Desired (solid line) and achieved (dotted line) target patch spectra. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table I. Mean face color values after correction in normalized color coordinates for each of the three examples of Figure 2.

	Canon S400			HP 850			Nikon Dl			Nokia 6820		
	Y	r	g	Y	r	g	Y	r	g	Y	r	g
Face 1												
Resulting mean skin color	0.483	0.446	0.315	0.466	0.442	0.311	0.522	0.429	0.316	0.495	0.437	0.308
Color correction $\Delta(Y, r, g)$	−0.005	0.019	0.013	0.017	0.048	0.009	−0.200	0.062	−0.002	0.037	0.021	−0.006
Face 2												
Resulting mean skin color	0.466	0.422	0.311	0.497	0.438	0.318	0.533	0.423	0.329	0.486	0.438	0.315
Color correction $\Delta(Y, r, g)$	0.028	0.015	0.017	0.083	0.037	0.006	−0.218	0.061	0.005	0.049	0.002	0.001
Face 3												
Resulting mean skin color	0.362	0.490	0.320	0.397	0.464	0.322	0.383	0.469	0.325	0.379	0.455	0.313
Color correction $\Delta(Y, r, g)$	−0.031	0.033	0.019	−0.002	0.047	0.012	−0.283	0.089	0.006	0.026	0.004	0.000

$Y = (R + G + B)/3$, $r = R/(R + G + B)$, and $g = G/(R + G + B)$.

compare the image extracted mean face colors with face color values derived from measured skin spectra. We finally investigate the consistency of skin color assessment across several cameras by studying the correlation between the different resulting skin tones extracted from images of the same people captured with four different devices.

C. Which Reference Values to Use in the Color Transform?

Previous work (Harville et al., 2005) showed that not considering the primaries and secondaries in the computation of the color transform gave better results in terms of skin color classification, despite a poorer visual appearance of the whole image. It was not yet clear whether or not the neutral patches should be taken into account. The black and white target background could also be used in the color correction, but those colors are more likely to be clipped in the images and were thus not considered.

The target is the only well-defined element in the image; we thus use it to estimate the color transform performance. Several subsets of eight skin patches, with and without the gray patches, are used to compute **A** using (3). The error is then computed on the remaining skin patches. To allow better comparison between the different transforms, we also computed the error on the entire set of skin patches.

The five tested transforms were computed using the following patches (see Fig. 3):

1. The first row of skin tones (odd numbers) and the two middle gray patches.
2. The first row of skin tones (odd numbers) only.
3. The second row of skin tones (even numbers) and the two middle gray patches.
4. The second row of skin tones (even numbers) only.
5. Eight patches (1, 2, 5, 8, 9, 10, 12, and 15) forming a convex hull of all skin colored patches in *CIELAB*, including two patches in the center.

These five transforms were tested on the entire set of 53 images taken with the HP850 camera. The error was estimated by computing color differences in *CIELAB*, i.e. the Euclidian distances between the values extracted from images and the target reference values

$$\Delta E_{ab}^* = \sqrt{(L_{ref}^* - L_{image}^*)^2 + (a_{ref}^* - a_{image}^*)^2 + (b_{ref}^* - b_{image}^*)^2} \quad (4)$$

$\Delta E_{ab}^* = 1$ is usually considered to be the minimum distance between two distinguishable color stimuli.

Table II shows the values of ΔE_{ab}^* for the five transforms. Comparing the results for the transforms using the gray patches (transforms 1 and 3) with the transforms using skin colored patches only (transforms 2, 4, and 5), we see that leaving out the gray patches allows a color prediction of $\Delta E_{ab}^* \approx 0.8$ as opposed to $\Delta E_{ab}^* \approx 1$ with the middle grays. The skin tones are better mapped if the color transform is computed using skin tones only. There is no significative difference among the ΔE_{ab}^* values obtained for the three transforms using solely skin patches and we cannot determine which of the three transforms performs the best. However, it is important to cover the whole range of lightness in the choice of skin patches for the color transform computation. Figure 5 shows the same image corrected with transforms, from left to right, computed using all 16 skin patches, eight skin patches covering the whole range of lightness, and the eight lighter patches, respectively.

The number and size of the patches have an influence on the quality of the color correction. To compute Eq. (2), $\mathbf{T}\{3 \times n\}$ and $\mathbf{M}_{\{4 \times n\}}$ must have at least four columns, and thus the minimal set contains four patches. The mean face color values obtained using four patches (1, 8, 10, and 12) forming a convex hull in *CIELAB* space and the mean face color values obtained using the 16 skin colored patches are close. Table III shows the difference between the color corrected patch values, using 4 and 16 patches for the color correction, averaged over all 53 images.

Reducing the number of patches allows increasing their size. Larger color patches will be better detected, especially for low-resolution cameras. In the images taken with the Nokia 6820 cell phone, patches can be as small as 10×10 pixels. At this scale, JPEG artifacts can introduce important errors in the estimation of patch color values. Even though skin tones can be color corrected using only four patches, the color correction will be less robust. If, for example, shadows are projected on just one patch, the resulting color transform becomes inconsistent. Thus, a higher number of patches

Table II. Error in terms of color differences for each transform, averaged over the 53 images.

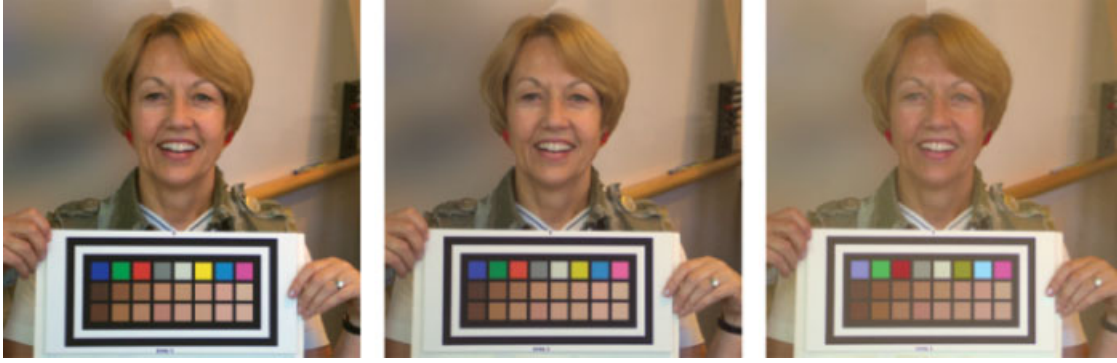|	ΔE Estimated on 8 Patches		ΔE Estimated on 16 Patches	
Transform	Mean ΔE	Var ΔE	Mean ΔE	Var ΔE
1. Skins 1 + gray	1.19	0.12	1.08	0.11
2. Skins 1	1.05	0.05	0.82	0.04
3. Skins 2 + gray	1.45	0.05	1.17	0.11
4. Skins 2	1.18	0.26	0.79	0.20
5. Skins in lab	0.81	0.03	0.82	0.04

Figure 5. The same image corrected with transforms computed using all 16 skin patches, eight skin patches covering the whole range of lightness, and the eight lighter patches, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

allows a finer sampling of skin tones and in general more accurate transforms. The trade-off depends mainly on the camera resolution.

D. Face Color Values Derived from Spectral Measurements.

The faces are extracted using the Viola–Jones face detection (Viola and Jones, 2001; Harville et al., 2005) and only pixels having a lightness between 10 and 90% (computed as $Y = (R + G + B)/3$) are considered to remove outliers due to hair, eyebrows, eyes, and lips. The mean color values of the remaining pixels after color correction are then used as the skin color estimate. The skin reflectance of each face was also measured on a uniform area of the cheek with a portable Microflash spectrometer using $0°/45°$ measurement geometry.

We computed *sRGB* skin values for each of the 53 people using the skin reflectance spectra. *sRGB* values are converted into normalized color coordinates $Y = (R + G + B)/3$, $r = R/(R + G + B)$, and $g = G/(R + G + B)$. These values were compared with face colors extracted from the HP850 images color corrected with the transform computed using all skin patches. The correlation between extracted (Y_{image}, r_{image}, g_{image}) face color values and spectrally derived ($Y_{spectra}$, $r_{spectra}$, $g_{spectra}$) values is high (see Fig. 6). However, spectrally derived values have systematically smaller r components and larger Y components than values extracted from the images. These differences can have several causes: direct measurement of skin reflectance by spectrometry is not optimal as there are often important differences of shades within one face, the pressure of the probe can be a source of error, and the skin is not lambertian, flat, and un-textured. Moreover, there is still a significant amount of shadow in the estimated skin pixels despite the lightness bounds applied to remove nonskin pixels.

E. Color Correction Consistency Across Cameras.

Skin color values extracted from color corrected images cannot be directly matched with *sRGB* values derived from spectra. Nevertheless, the systematic discrepancies indicate that the color correction of skin tones is coherent across cameras. Figures 7 and 8 show the lightness values Y and normalized chromaticity coordinates (r, g) extracted from images taken with the four cameras, displayed for each pair of cameras, respectively. The black dotted line indicates the linear relation.

The correlation is high across all pairs of cameras and the relation is linear. Table IV reports the correlation coefficients for the corresponding (Y, r, g) values for each pair of cameras and the equivalent *CIELAB* ΔE_{ab}^* color difference averaged over all images. Correlation coefficients range from 0.90 to 0.98 for Y, from 0.57 to 0.96 for r, and from 0.82 to 0.91 for g. ΔE_{ab}^* ranges from 1.8 to 5. The quality of the estimation depends clearly on the quality of the camera. The data extracted using the Nokia cell phone is noisier, but still correlates well with data extracted using higher quality cameras. The resulting error $\Delta E_{ab}^* \approx 5$ is quite large, but the extracted face color still allows for a good classification of skin tones. An accuracy of $\Delta E_{ab}^* = 1$, considered as the distance between two distinguishable color stimuli, may not be required for all applications.

Note that these results were obtained using four cameras from four different manufacturers and without bypassing any of the image processing implemented in the camera.

5. CONCLUSION

Accurate color cannot be retrieved from uncalibrated images taken with uncalibrated cameras. However, a limited range of colors can be estimated by using appropriate color information in the form of a

Table III. Difference between the patch color values color corrected using 4 and 16 patches.

Color differences for Y	Mean $|\Delta Y|$	Max $|\Delta Y|$	Mean $|\Delta Y_{norm}|$	Max $|\Delta Y_{norm}|$
	0.03	0.11	1.2%	6.1%
Color differences for r	Mean $|\Delta r|$	Max $|\Delta r|$	Mean $|\Delta r_{norm}|$	Max $|\Delta r_{norm}|$
	0.01	0.05	0.7%	4.9%
Color differences for g	Mean $|\Delta g|$	Max $|\Delta g|$	Mean $|\Delta g_{norm}|$	Max $|\Delta g_{norm}|$
	0.01	0.02	0.9%	2.7%

The symbol $\Delta (\cdot)_{norm}$ Indicates the difference between (Y, r, g) values of the two sets normalized by their sum. The values are averaged over all 53 images.
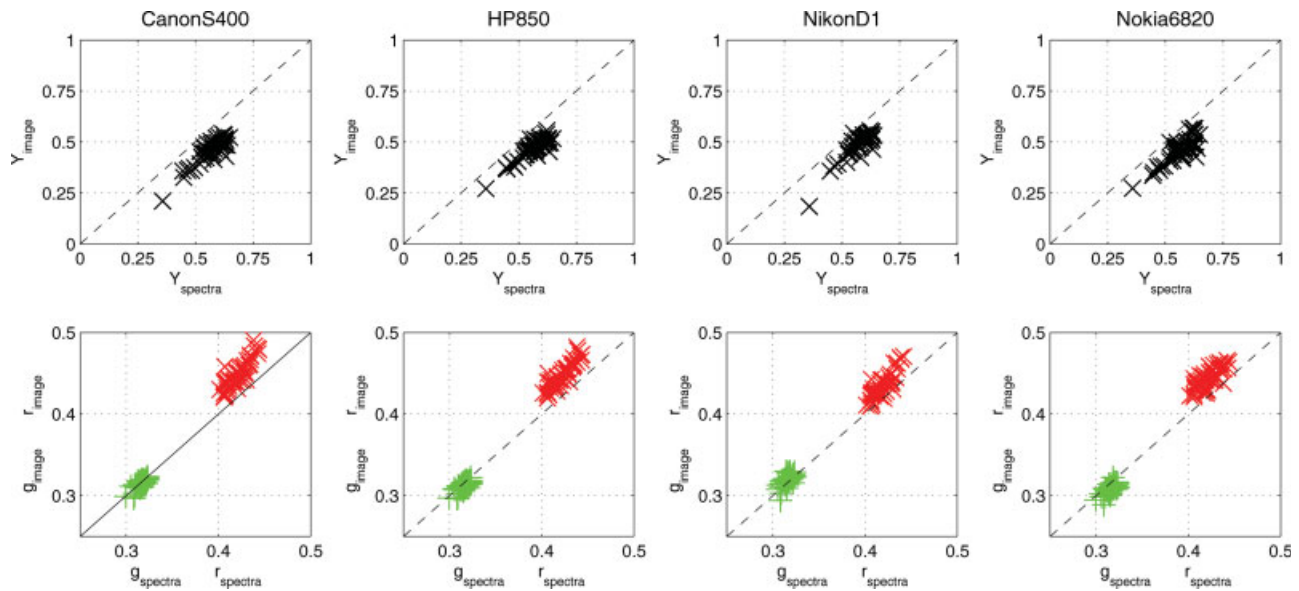
Figure 6. Spectrally derived values (*x*-axis) vs. image extracted values (*y*-axis) for Canon S400, HP850, Nikon D1, and Nokia 6820. The top row shows lightness values $Y = (R + G + B)/3$. The bottom row shows normalized color coordinates $r = R/(R + G + B)$ and $g = G/(R + G + B)$. The black dotted line indicates the linear relation. [Color figure can be viewed in the online issue, which is available at www.interscience. wiley.com.]

target present in the scene. A color transform mapping the scene target color values onto pre-computed target reference values is computed by least mean square estimation in *sRGB* and applied to the entire image.

Using this method, skin tones can be consistently classified using uncalibrated images taken with a large variety of cameras and color corrected using a target consisting of patches characteristic of the range of human skin tones. With this mapping, skin tones can
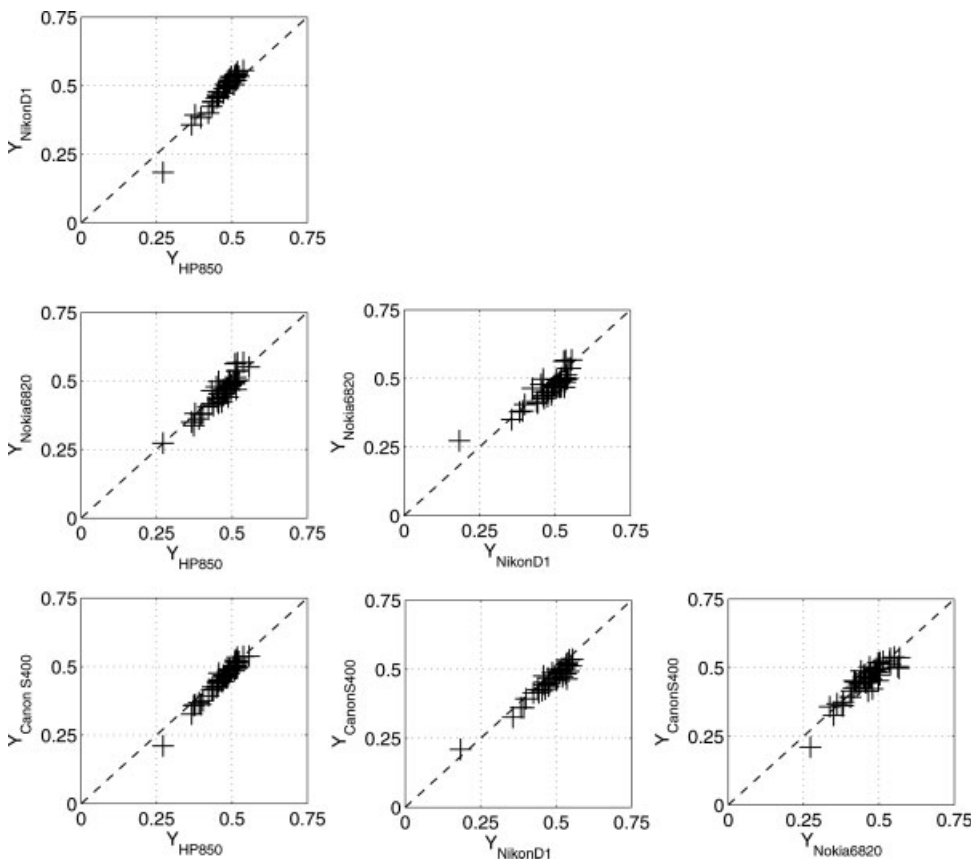

Figure 7. Lightness $Y = (R + G + B)/3$ values compared for each pair of cameras. The *x*-axis shows (from left to right) cameras HP850, Nikon D1, and Nokia 6820 and the *y*-axis shows (from top to bottom) cameras Nikon D1, Nokia 6820, and Canon S400. The black dotted line indicates the linear relation.

Figure 8. Normalized color coordinates $r = R/(R + G + B)$ and $g = G/(R + G + B)$ compared for each pair of cameras. The x-axis shows (from left to right) cameras HP850, Nikon D1, and Nokia 6820 and the y-axis shows (from top to bottom) cameras Nikon D1, Nokia 6820, and Canon S400. Th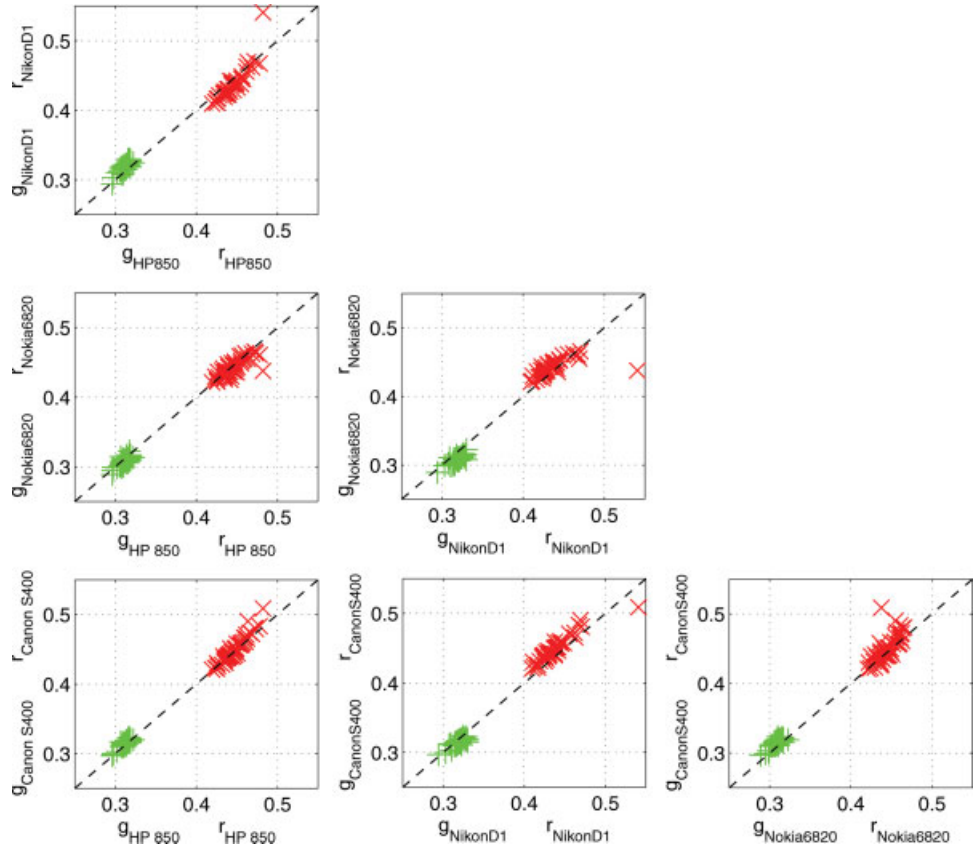e black dotted line indicates the linear relation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

theoretically be estimated with an error of $\Delta E_{ab}^{*} < 1$. There is noise in the measurement of face colors using real images, due to the estimation of the face pixels and the unevenness of skin. However, results correlate well with spectral data and across a variety of devices, with resolution as low as 0.1 megapixels and poor image quality.

Table IV. Correlation coefficients between the (Y, r, g) values of extracted skin color and corresponding ΔE_{ab}^{*} for each pair of cameras.

	HP	Nokia	Nikon	Canon
Y				
HP	1			
Nokia	0.93	1		
Nikon	0.98	0.90	1	
Canon	0.98	0.91	0.96	1
r				
HP	1			
Nokia	0.80	1		
Nikon	0.89	0.57	1	
Canon	0.96	0.71	0.94	1
g				
HP	1			
Nokia	0.80	1		
Nikon	0.86	0.82	1	
Canon	0.91	0.84	0.85	1
av. ΔE				
HP	0			
Nokia	3.59	0		
Nikon	3.28	5.32	0	
Canon	1.90	4.00	3.40	0

The high correlation across devices indicates that with a solid ground truth, the system can be correctly trained to give consistent results. The assessment of color is not perfect though, but this method does not require any expensive calibrated imaging devices or controlled illuminant and can be performed with any consumer camera.

The least mean square estimation of the color correction matrix in *sRGB* allows a fast and computationally low color correction, which can be used in a variety of applications.

The method assumes a uniform illuminant across the image, shadows and mixed illuminants can be an important source of error. A comparison between the two halves of the face allows eliminating some failing cases. Also, the number and size of the patches have an influence. A higher number of patches ideally give a finer sampling of skin tones and more robustness. Reducing the number of patches allows increasing their size and thus improves the extraction and estimation of the target color values, especially for low-resolution cameras. The trade-off depends on the application and on the type of cameras used.

REFERENCES

A.O. Barel, P. Clarys, K. Alewaeters, C. Duez, J.-L. Hubinon, and M. Mommaerts, The Visi-Chroma VC-100A: A new imaging colorimeter for dermatocosmetic research, Skin Res Technol 7 (2001), 24–31.

K. Barnard and B. Funt, Camera characterization for color research, Color Res Appl 27 (2002), 153–164.

Y. Cai, A novel imaging system for tongue inspection, Proc IEEE Instrum Meas Technol Conf, (2002), 159–163.

L. Caisey, F. Grangeat, A. Lemasson, J. Talabot, and A. Voirin, Skin color and makeup strategies of women from different ethnic groups, Int J Cosmet Sci, 28 (2006), 427.

P. Clarys, K. Alewaeters, R. Lambrecht, and A.O. Barel, Skin color measurements: Comparison between three instruments: The Chromameter, the DermaSpectrometer, and the Mexameter, Skin Res Technol 6 (2000), 230–238.

J. Farell, D. Sherman, and B. Wandel, "How to turn your scanner into a colorimeter," Proceedings of IS&T 10th international congress on advances in non-impact printing technologies IS&T, 1994, pp. 579–581.

M. Harville, H. Baker, N. Bhatti, and S. Süsstrunk, Image-based measurement and classification of skin color, Proc IEEE Int Conf Image Process 2 (2005), 374–377.

R.-L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, Face detection in color images, IEEE Trans Pattern Anal Machine Intell 40 (2002), 696–706.

P.M. Hubel, J. Holm, G.D. Finlayson, and M.S. Drew, "Matrix calculations for digital photography," Proceedings of 5th IS&T color imaging conference, 1997, pp. 105–111.

IEC 61966-2-1, Multimedia systems and equipment, Colour measurement and management, Part 2-1. Colour management, Default RGB colour space, *sRGB*, 1999.

T. Igarashi, K. Nishino, and S.K. Nayar, The appearance of human skin, Technical Report CUCS02405, 2005.

H. Nakai, Y. Manabe, and S. Inokuchi, Simulation and analysis of spectral distributions of human skin, Proc 14th Int Conf Pattern Recognit 2 (1998), 1065–1067.

G.E. Piériard, EEMCO guidance for the assessment of skin colour, J Eur Acad Dermatol Venerol 10 (1998), 1–11.

B. Smith, C. Spiekermann, and R. Sember, Numerical methods for colorimetric calculations: Sampling density requirements, Color Res Appl 17 (1992), 394–401.

M. Störring, H. J. Andersen, and E. Granum, Physics-based modelling of human skin colour under mixed illuminants, J Robot Autonomous Syst 35 (2001), 131–142.

S. Taylor, W. Westerhof, S. Im, and J. Lim, Noninvasive techniques for the evaluation of skin color, J Am Acad Dermatol 54 (2006), S282–S290.

S. Tominaga, Spectral imaging by a multi-channel camera, J Electron Imaging 8 (1999), 332–341.

N. Tsumura, H. Haneishi, and Y. Miyake, Independent-component analysis of skin color image, J Opt Soc Am A 16 (1999), 2169–2176.

P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, IEEE Comput Soc Conf Comput Vision Pattern Recognit 1 (2001), 511–518.

W. Wu, J. P. Allebach, and M. Analoui, Imaging colorimetry using a digital camera, J Imaging Sci Technol 44 (2000), 267–279.