Color correction of uncalibrated images for the classification of human skin color

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

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

[2]***Hewlett-Packard Laboratories, Palo Alto, USA***

Abstract

Images of a scene captured with multiple cameras have different color values due to variations in capture and 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 for cosmetics recommendations. A subject is imaged with a calibration target in the scene. This 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 a 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. Skin color can be consistently assessed across cameras with different color rendering and resolutions ranging from 0.1 Mpixels to 4.0 Mpixels.

Introduction

Our goal is to retrieve accurate color information from uncalibrated images taken with unknown cameras. Due to incomplete illuminant compensation and to the different characteristics of available cameras, consistent color rendering is not achieved and the same object captured with different cameras have different pixel values in the resulting images. Consequently, the retrieval of accurate color information requires either the pre-calibration of the imaging devices and the control of the illuminant, or additional scene information.

We are interested in retrieving *colorimetric* as opposed to spectral information about the scene. Thus, to use any consumer camera as a colorimeter, we need known color information present in the scene in the form of a calibration target, whose pixel values are used as reference for the color correction of images. The method we present is developed for the color correction and classification of skin tones, but can be generalized to other colors.

The appearance of skin has been studied mostly for rendering purposes in computer graphics, for face detection and tracking in computer vision, for diagnostic purposes in dermatology, and for makeup and skin care in cosmetics. However, there is little research on how to assess skin tones accurately from digital images. Perceived color is the most discriminative of skin attributes and depends on its pigmentation, blood microcirculation, roughness, sebum, and perspiration [2]. Its objective measurement has been made mostly by traditional reflectance spectrometry (for a review see [5]) and using narrow band spectrometers developed specifically for dermatology [8]. Spectrometry of skin has two main drawbacks: the area measured is about 0.05 cm^2, but skin is not homogeneous [2]. Additionally, the pressure of the probe on the skin can be an important source of bias [7]. Still, traditional spectrometers are inexpensive and simple to use and thus widely employed.

Figure 1. *uncorrected images (top row) and corrected images (bottom row) for cameras (from left to right) Canon S400, HP850, Nikon D1, Nokia 6820*

To overcome the problem of the probe pressure on skin, a proprietary device composed of an integrating sphere, a spectrometer, and a tri-CCD camera was developed [3], allowing non contact spectroscopy of different parts of the face and simultaneous imaging for estimation of the skin color inhomogeneity.

Due to uneven tan, blemish, or shine, the color and appearance of skin are usually not uniform across a subject's face. Moreover, because of its volume, there are also important shadows and specularities across the face, making the estimation of skin color from images more difficult.

We present a simple and inexpensive method to assess skin color from digital images for applications such as online shopping, for which the use of calibrated devices is not feasible, or for automated suggestion of personal appearance products, such as makeup or clothing that complement skin tone. Uncalibrated images taken under unknown illuminants are color corrected by mapping selected pixel values onto reference values present in the scene in the form of a target. This target is extracted and a 3×4 linear color transform is computed by least mean square error between the extracted target color values and pre-computed color values. This color correction transform is then applied to the entire image and face pixel values extracted. Figure 1 shows the results of our method applied to images of the same subject captured with four different cameras.

We show that our method allows color correcting skin tones with an accuracy in terms of *CIELAB* color difference of $\Delta E^*_{ab} <$ 1. Face color values extracted from color corrected images show

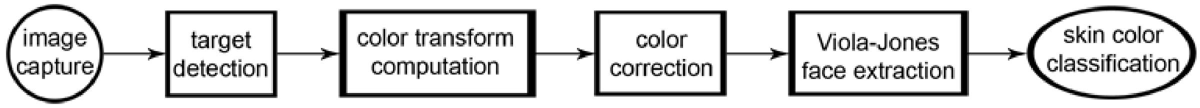

Figure 2. *The pipeline consists mainly in image color correction, face pixel extraction, and classification of skin tones.*

high correlation with skin color derived from in vivo spectral measurements, and also show high correlation across a wide range of cameras from different manufacturers and different quality levels.

The article is structured as follows: in the next section we present the color assessment with uncalibrated cameras, then we describe the experiments, report the main results, and conclude the article.

Color assessment with uncalibrated cameras

Color values of an object are never identical when imaged with multiple uncalibrated cameras, even when the lighting condition is the same. This is due to imperfect illuminant compensation, different sensor responses, and to variations in image processing algorithms and in hardware quality. Accurate color assessment thus requires either pre-calibration of the devices or a color correction of the output images.

We propose a method to retrieve skin color information from digital images taken with a single, casually posed consumer camera. The illuminant spectra and sensor responses are unknown but we use a calibrated target present in the scene as reference to compute a linear color correction transform.

A subject is imaged along with the reference target, which is automatically detected and whose pixels are extracted. A 3×4 linear transform mapping the image extracted target values onto pre-computed reference target values is computed by least mean square error and then applied to the entire image. The color transform is computed in *sRGB* [1], i.e. the reference values are *sRGB* values. We work under the assumption that the output images are already encoded in *sRGB*. In other words, the in-camera processing should result in color triplets that are already close to *sRGB* values. The color correction method is explained in detail in [6].

The face is extracted using a Viola-Jones face detector [9] and then its skin tone is classified according to the extracted face pixel values. Figure 1 shows an example of color correction.

The processing pipeline can be divided into three main steps: 1) image color correction 2) face pixel extraction, and 3) classification of skin tones (see Figure 2). In this paper, we will focus on the color transform and on its performance. For a discussion of the whole image pipeline, see [4].

Computation of the color correction transform

The color correction transform is a 3×4 matrix $\underline{A}$ mapping the mean patch camera color values $\underline{M}$ onto the *sRGB* reference target values $\underline{T}$.

$$\underline{T}_{\{3\times n\}} = \underline{A}_{\{3\times 4\}} \cdot \underline{M}_{\{4\times n\}} \tag{1}$$

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

We want to find $\underline{A}$ minimizing $\|\underline{T} - \underline{A}\underline{M}\|$, i.e. minimizing the least mean square error in *sRGB* color encoding. $\underline{A}$ is given by

$$\underline{A}_{\{3\times 4\}} = \underline{T}_{\{3\times n\}} \underline{M}^{+}_{\{n\times 4\}} = \underline{T}_{\{3\times n\}} \underline{M}^{T}_{\{n\times 4\}} (\underline{M}\underline{M}^T)^{-1}_{\{4\times 4\}} \tag{2}$$

where $^+$ denotes the Moore-Penrose pseudo-inverse. $\underline{A}$ provides a 3×3 color transform plus a per-component offset.

The choice of *sRGB* as a reference encoding was made for computational and simplicity reasons. Performing the least mean square computation of the transform in a perceptual space such as *CIELAB* requires more calculation and non-linear transforms of *sRGB* values. The use of *sRGB* as reference encoding rather than linear *sRGB* is motivated by the presence of a non-linearity in *sRGB*, which is perceptually more relevant that linear RGB values.

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 (see Figure 3). The first row contains primary and secondary colors and two shades of grey. The two last rows contain 16 patches characteristic of human skin colors ordered by uniformly increasing lightness alternating on two rows.

We printed the target on photopaper with close to lambertian surface characteristics and measured the reflectance spectrum S^k of each patch k. These measures were used to compute *sRGB* reference values using a simple image formation model

$$(X^k_{D65}, Y^k_{D65}, Z^k_{D65}) = \frac{1}{N} \sum_{\lambda_i} S^k_{\lambda_i} \cdot E^{D65}_{\lambda_i} \cdot (\bar{x}_{\lambda_i}, \bar{y}_{\lambda_i}, \bar{z}_{\lambda_i}) \tag{3}$$

where E^{D65} is the illuminant spectrum, S^k the reflectance spectrum, N a normalization constant, and $\bar{x}, \bar{y}, \bar{z}$ the CIE 1931 2° color matching functions. We then use the *CIEXYZ* to *sRGB* transform specified in [1]. For each image, the target patches are extracted and the color transform $\underline{A}$ (1) is computed using the Moore-Penrose pseudo-inverse (2). This transform is then applied to the entire image prior to face pixel extraction and color classification.

Experiment

53 people holding a copy of the calibration target were imaged with four different RGB cameras. The cameras were an HP850 (3.9 Mpixels), a Nikon D1 (2.7 Mpixels), a Canon S400 (4.0 Mpixels), and a Nokia 6820 cell phone camera (0.1 Mpixels). The reflectance of each skin was also measured using a portable Microflash spectrometer with a 0°/45° geometry. The illumination conditions vary across subjects, but are constant in all four images of one subject imaged with the different devices. Figure 1

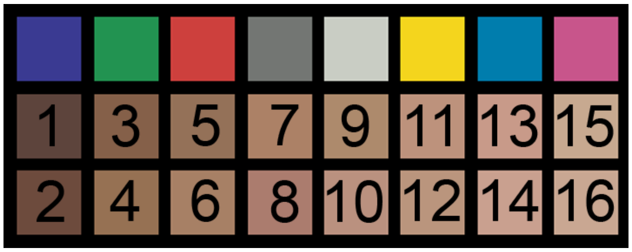

Figure 3. *The reference target contains three rows of eight patches. The last two rows are representative of human skin tones.*

shows an example of before and after color correction computed using the 16 skin tones patches. The first row shows the uncorrected images having resolutions from 0.1 to 4.0 Mpixels and different color qualities, and the second row the images corrected using the method described in this paper.

Optimal choice of target reference values

We investigated the colorimetry of the target color patches to determine which reference values should be used in the computation of the color correction matrix $\underline{A}$. Our target primarily consists of color patches representative of human skin tones. For an accurate color correction of skin tones, it is important to have a dense sampling in this region of the color space. We thus use skin tones for the computation of the color transform $\underline{A}$. The two shades of grey can be used for an additional white balancing of the images and may influence the color correction of skin color. The black and white background can also be used in the color correction computation, but those colors are more likely to be clipped in the images and were thus not considered. We want to determine which selection of patches should be used in the color transform and if the middle greys should be taken into account in order to obtain an accurate color correction of skin tones.

As the target is the only element in the image whose color properties are known, it is used to estimate the performance of the color transform. Several subsets of 8 skin patches, with and without the grey patches, are used to compute $\underline{A}$ using eq. (2). Each test transform is computed and applied to the 53 images taken with the HP 850 camera. The target values are re-extracted and used to compute the error on the 8 remaining skin colored patches not used in the transform calculation. The transform using all 16 skin patches was also computed for comparison.

The five tested transforms were computed using:

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

These five transforms were tested by estimating the error in computing ΔE^*_{ab} color differences in *CIELAB*, i.e. the euclidian distances between the values extracted from images and the target reference values.

Table 1 shows the values of ΔE^*_{ab} for the five transforms, computed on the 8 skin patches not used in the color transform and on all 16 patches. Comparing the results for the transforms using the grey patches (transforms 1 and 3) with the transforms using skin patches only (transforms 2, 4, and 5), we see that leaving out the grey patches gives a better color correction of skin tones, with a color prediction of $E^*_{ab} \approx 0.8$ as opposed to $\Delta E^*_{ab} \approx 1$ with the neutral tones. The difference among the ΔE^*_{ab} values obtained for the three transforms using solely skin patches is not large enough to be significant.

Table 1: ΔE^*_{ab} estimated on 8 and 16 patches

ΔE^*_{ab} estimated on 8 patches		
Transform	mean ΔE^*_{ab}	var ΔE^*_{ab}
1. skins 1 + grey	1.19	0.12
2. skins 1	1.05	0.05
3. skins 2 + grey	1.45	0.05
4. skins 2	1.18	0.26
5. skins in Lab	0.81	0.03
ΔE^*_{ab} estimated on 16 patches		
Transform	mean ΔE^*_{ab}	var ΔE^*_{ab}
1. skins 1 + grey	1.08	0.11
2. skins 1	0.82	0.04
3. skins 2 + grey	1.17	0.11
4. skins 2	0.79	0.20
5. skins in Lab	0.82	0.04

Correlation between spectrally derived and image extracted color values

We extracted face pixels from the color corrected images and compared these values with spectrometric measurements of the subject's skin. Spectrometry allows objective measurement of color and has traditionally been used in many applications. However, spectrally derived skin colors may not always accurately represent the average color of a subject's face. The skin spectra measurements were thus systematically done on a uniform area of the face with minimal probe pressure.

The faces are extracted from the images using the Viola-Jones face detection [9] and only the face pixels having a lightness between 10% and 90% (computed as $Y = (R+G+B)/3$) are considered in order to remove outliers due to hair, eyebrows, eyes, lips, teeth, shadows, and specularities. The mean *sRGB* values of the remaining pixels are then used as the skin color estimate.

The skin reflectance spectra were used to compute *sRGB* values (3). These values were then compared with mean values of face pixels extracted from each of the 53 images taken with the HP850 camera and color corrected with the transform $\underline{A}$ computed using all skin colored patches. The *CIEXYZ* to *sRGB* transform is computed as indicated in [1]. *sRGB* values are converted into normalized color coordinates $Y = (R+G+B)/3$, $r = R/(R+G+B)$, and $g = G/(R+G+B)$. 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 Figure 4). However, spectrally derived values have systematically smaller r components and larger Y components than values extracted from the images. This difference can have several causes. The skin is not lambertian, flat, and untextured, and its reflectance measurements can be biased by the pressure of the probe. More-

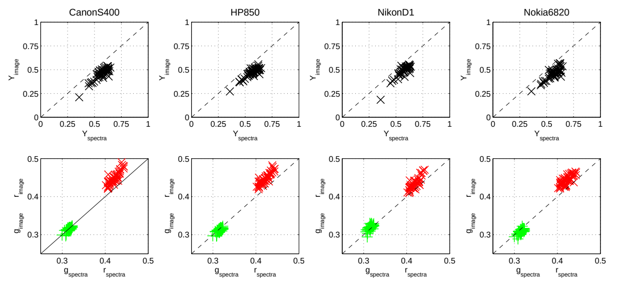

Figure 4. Spectrally derived values (x-axis) vs. image extracted values (y-axis) for Canon S400, HP850, Nikon D1, and Nokia 6820. The top row shows lightness values $Y = (R+G+B)/3$. The bottom row shows normalized color coordinates $r = R/(R+G+B)$ and $g = G/(R+G+B)$. The black dotted line indicates the linear relation.

over, the average of face pixel values is a simple estimate but may not optimally represent skin color, as it is generally uneven across a subject's face. There may still be a significant amount of shadows despite selecting skin pixels according to their lightness.

Consistency of extracted skin color values across a variety of cameras

Despite the discrepancy in lightness observed between spectrally derived and image face color values, the high consistency in color prediction across subjects indicates that the color correction is coherent. Most importantly, we want to make sure that the method is also consistent also across devices, such that with proper training, the system can assign skin tones for any new unknown RGB camera. The correlation is high across values derived

Table 2: Correlation coefficients across (Y,r,g) components for each pair of cameras and the corresponding $CIELAB$ ΔE^*_{ab} color difference averaged over all images

Y	HP	nokia	nikon	canon
HP	1			
nokia	0.93	1		
nikon	0.98	0.90	1	
canon	0.98	0.91	0.96	1
r	HP	nokia	nikon	canon
HP	1			
nokia	0.80	1		
nikon	0.89	0.57	1	
canon	0.96	0.71	0.94	1
g	HP	nokia	nikon	canon
HP	1			
nokia	0.80	1		
nikon	0.86	0.82	1	
canon	0.91	0.84	0.85	1
av. ΔE^*_{ab}	HP	nokia	nikon	canon
HP	0			
nokia	3.59	0		
nikon	3.28	5.32	0	
canon	1.90	4.00	3.40	0

from all four cameras and the relation is linear (see Figures 5 and 6). Table 2 reports the correlation coefficients across the (Y,r,g) values for each pair of cameras and the corresponding $CIELAB$ ΔE^*_{ab} color difference averaged over all images. Correlation coefficients range from 0.57 to 0.98 and most values are above 0.8. The lowest correlation is obtained with the Nokia cell phone camera, which was expected considering that it has the lowest quality. At this resolution, the size of one target patch is about 10×10 pixels (see Figure 7). JPEG artifacts then become important and may introduce an error in the estimation of the patch color values. The results are noisy but still give an estimate of skin tones with a $\Delta E^*_{ab} \simeq 5$. To compare this value against the range of all possible human skin tones, we compute $CIELAB$ values using the spectral measurements of the reference target for the lightest ($L^* = 71.6$, $a^* = 9.01$, $b^* = 14.8$) and darkest ($L^* = 31.17$, $a^* = 8.91$, $b^* = 9.47$) skin colored patches, giving a color difference $\Delta E^*_{ab} = 40.7$. Even though the error $\Delta E^*_{ab} \simeq 5$ is quite large, it still allows 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. The results using the Nikon D1 camera are also quite low, despite the much higher resolution, but this set of pictures has systematically important clipping in the three color channels.

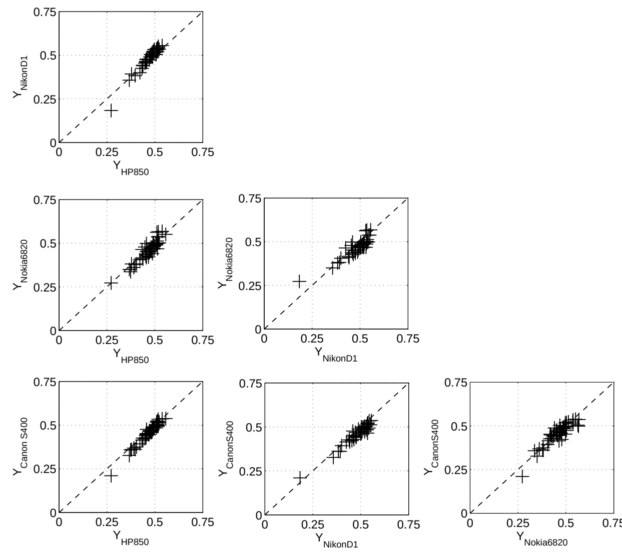

Figure 5. Lightness $Y = (R+G+B)/3$ values compared for each pair of cameras. The x-axis shows (from left to right) cameras HP850, Nikon D1, and Nokia 6820 and the y-axis shows (from top to bottom) cameras Nikon D1, Nokia 6820, and Canon S400. The black dotted line indicates the linear relation.s

The error in the skin color assessment has several sources of error adding up. There is an error in the estimation of the color correction transform inherent to the least mean square estimation. It is related to the quality of the imaging devices and can also be worsened by mixed illuminants, clipping in the images, or shadows projected on the target. There is also an error in the estimation of skin pixels. Shadows and specularities may not be completely eliminated by the lightness bounds and shift the mean face color.

Note though that these results were obtained using four cameras with very different characteristics and resolutions without bypassing any of the in-camera image processing.

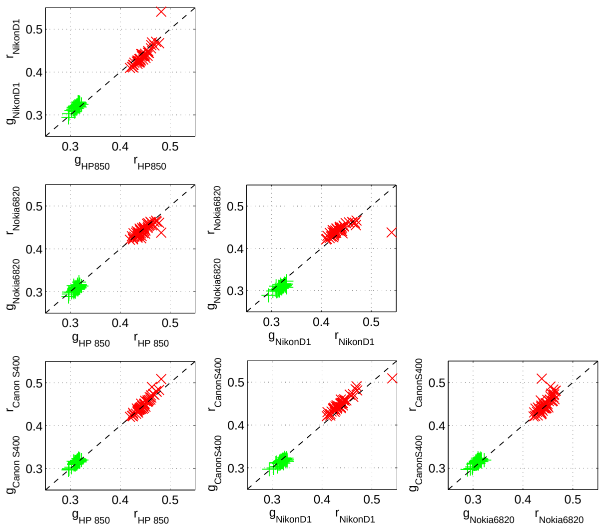

Figure 6. *Normalized color coordinates $r = R/(R+G+B)$ and $g = G/(R+G+B)$ compared for each pair of cameras. The x-axis shows (from left to right) cameras HP850, Nikon D1, and Nokia 6820 and the y-axis shows (from top to bottom) cameras Nikon D1, Nokia 6820, and Canon S400. The black dotted line indicates the linear relation.*

Figure 7. *When using low resolution cameras, JPEG artifacts become important and can introduce errors in the color transform estimation.*

Conclusion

We present a method that allows accurate assessment of skin color from uncalibrated images taken with uncalibrated consumer cameras under unknown illuminants. Images are color corrected using pre-computed reference values present in the image in the form of a calibration target consisting of patches representing a range of human skin tones. Skin color estimated with this technique correlates well with spectral data and across a variety of uncalibrated devices with resolutions as low as 0.1 Mpixels, without control of the illuminant.

The accuracy of color assessment is lowered by the unevenness of skin and the error in the extraction of face pixels. Thus our method allows consistent, though not perfect, skin color assessment without requiring any expensive calibrated imaging devices or control of the illuminant and can be performed with any consumer camera. This method can be applied to other limited color gamuts.

References

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

[2] 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., vol. 7, p. 24-31, 2001.

[3] L. Caisey, F. Grangeat, A. Lemasson, J. Talabot, and A. Voirin, Skin color and makeup strategies of women from different ethnic groups, Int. J. Cosmetic. Sci., vol. 28 (6), pp. 427, 2006.

[4] M. Harville, H. Baker, N. Bhatti, and S. Süsstrunk, Image-based measurement and classification of skin color, Proc. of IEEE Int. Conf. on Image Processing, vol. 2, p. 374-377, 2005.

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

[6] J. Marguier, N. Bhatti, H. Baker, M. Harville, and S. Süsstrunk, Assessing human skin color from uncalibrated images, to be published in Int. J. Imag. Syst. Tech., special issue on Appl. Color Image Processing, 2007.

[7] G.E. Piériard, EEMCO guidance for the assessment of skin colour, J. Eur. Acad. Dermatol., vol. 10 (1), p. 1-11., 1998.

[8] S. Taylor, W. Westerhof, S. Im, and J. Lim, Noninvasive techniques for the evaluation of skin color, J. Eur. Acad. Dermatol., vol. 54 (5), p. S282-S290, 2006.

[9] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, Conf. Comput. Vis. Pattern Recogn., vol. 1, pp. 511, 2001.

Author Biography

Joanna Marguier received her MS in physics from the Ecole Polytechnique Fédérale de Lausanne (EPFL) in 2003. She worked two years as process engineer in the Displays Department at Swatch Group's R&D Laboratories. Since 2005, she has been a Research Assistant in Color Imaging, pursuing a PhD degree in the Audiovisual Communications Laboratory, EPFL. In 2006, she spent three months as an intern at Hewlett-Packard Laboratories, Palo Alto.