https://doi.org/10.2352/ISSN.2470-1173.2019.6.MAAP-483


# Hair Tone Estimation at Roots via Imaging Device with Embedded Deep Learning

***Panagiotis-Alexandros Bokaris*, Emmanuel Malherbe*, Thierry Wasserman, Michaël Haddad, Matthieu Perrot; L'Oréal Research & Innovation; Clichy, 92110, France***
****Equal contribution***

## Abstract

*Establishing accurate hair diagnosis at roots is a significant challenge with strong impact on hair coloration, beauty personalization and clinical evaluation. The roots of hair - viz. the first centimeter away from the scalp - represent clean hair fibers that have not been subjected to color change due to hair dyeing or environmental conditions. Therefore, they are a measure of a person's baseline hair characteristics, including natural hair tone. A device that acquires high resolution macro images of hair roots under a well-defined illumination geometry has been designed in order to assess natural hair tones. Image analysis in this scenario is not a trivial task since the acquired images present an overlap of scalp and hair, with other possible artifacts due to dandruff and hair transparency. In this paper, we propose to train a Convolutional Neural Network (CNN) on a data-set composed of images from subjects who had their hair tone evaluated by trained color experts. Our method is compared with other popular CNNs as well as conventional color image processing approaches developed for this task. We found that the proposed model not only offers higher precision but also provides faster computation times, due to its lighter architecture in contrast to popular CNNs. Thus, we achieve sufficiently accurate results in real time on the low-end chip embedded in our device.*

## Introduction

Cosmetics, one of the largest and constantly growing industrial sectors, is going through a remarkable digital transformation driving product innovation [1]. A highly diverse and globally connected consumer market results in a constant demand for new products and novel experiences. Personalized solutions are key components in satisfying people who change rapidly their wishes and needs.

A cosmetics sector that faces many challenges in this digital era is hair care and hair coloration. Due to the complex nature of hair and the process of hair dyeing, accurate hair diagnosis is crucial in order to provide the clients with personalized hair care and coloration products [2]. Currently at hair salons, the first necessary step before applying any hair coloring product or recommending a hair treatment is hair diagnosis. Among the important features that the hairdresser needs to estimate are the hair tone, white hair percentage, hair diameter and density. Hair tone is the visual attribute of the hair lightness/darkness and is related to the hair melanin concentration. It is traditionally measured using a logarithmic scale starting at 1, for very dark hair, to 10 for very light blond hair [3]. The root of the hair (e.g. the very first part of hair that emerges from the scalp) is commonly the only region where we have access to hair that has not been altered by external factors, such as natural/artificial colorants. It is a region where we can measure natural hair color and white hair percentage as well as hair density. However, images in the roots region show not only hair but also the scalp that can vary highly in color, oiliness, and dandruff content. In addition, the hair fibers themselves are translucent, resulting in color dependence on the scalp background, and have a certain natural variability both in color and thickness. This diagnosis is currently made manually by hairdressers, and despite their expertise and training they are not always capable of making accurate estimations of all of these features, especially under the variable lighting conditions of hair salons. The high variability between different hairdressers regarding hair tone estimation in real life conditions was studied and is presented in Section Data Acquisition.

In order to help the hairdresser and improve client satisfaction, we present here an automated hair diagnosis device that can serve to accurately estimate hair and scalp attributes. The automation of the process is of paramount importance to ease the life of hairdressers and equip them with new tools that guarantee precision, robustness and efficiency. Moreover, such devices, as "neutral observers", go beyond the human visual perception and offer an objective notation that does not vary depending on the hairdresser and the conditions - in other terms, a standardized notation.

In this paper, we propose an automatic image analysis method we developed for this device in order to estimate the natural hair tone at roots. We focus on hair tone estimation since it is one of the crucial attributes that need to be standardized. Processing the hair-scalp images faces various computer vision challenges, as they present regions of scalp and hair pixels with various undesired effects, such as specularities, shadows, hair transparency, dandruff and skin irritations/redness. In order to tackle this challenge, we followed an approach that takes advantage of deep learning in computer vision. It is evaluated against conventional methods that are based on colorimetric features of hair and color image processing, as well as other popular CNN methods.

The main contribution of this work is addressing the challenging and complex problem of hair tone estimation in the context of an industrial application. Moreover, the proposed solution achieves the most accurate estimation of the compared approaches while guaranteeing low computation times, which makes the real-time estimation on the device feasible.

## Related Work

To the best of our knowledge, hair tone estimation has not been covered earlier in the literature. However, related work that this research builds upon can be separated in three categories: 1)

imaging devices for hair diagnostics, 2) similar applications in the fields of computer vision, recently benefiting from the advances in deep learning.

### Hair Diagnosis

In the last few decades, new technologies and digital processing have been used to address an increasing demand for personalized diagnosis and treatment, automatize and stabilize processes and allow as well the measurement of properties that are either significantly hard or impossible to assess by the naked eye. The concept of creating a device that is used for hair analysis has been mentioned several times in the literature. MacFarlane et al. [4] introduced the idea of a method and a device for acquiring the Hunter $L$, $a$, $b$ values [5] of hair and use these to propose an appropriate coloring agent that can achieve the desired hair tone from a predefined database. Leprince [6] presented the concept of an apparatus that gets as inputs the percentage of senescent hair, the natural hair tone, and a set of hair coloring products, in order to estimate the feasible results and and suggest the appropriate hair coloring solution. Ladjevardi [7] described a cosmetic color analysis system that analyzes the color of a three-dimensional object, such as hair, through acquisition(s) by a camera sensor and produces a cosmetic color determination by relative weightings of multiple cosmetic colors. In another approach, Sato [8] proposed a method and an apparatus for displaying images of hair regions in order to evaluate the hair styling using image processing.

However, despite the various patents and experimental trials, a functional device that succeeds in proposing a reliable hair diagnosis remains a challenging topic. The device presented in this work provides hair-scalp macro images of unique quality and resolution that allows the estimation of hair tone at roots. Hair-scalp images have been used mainly in videodermoscopy [9] that mostly focuses on skin analysis for identifying scalp or hair disorders. To the best of our knowledge, it is the first time such an estimation is proposed, which pushes the boundaries of hair treatment/coloration further.

### Similar Applications in Machine Vision

Making prediction from acquisitions of an imaging device, such as the one proposed in this paper, shares the same principles with other industrial applications in machine vision. The interested reader might find more information on such systems in the survey of Malamas et al. [10]. Extracting features and classifying attributes from color images in machine vision applications usually relies on image processing paired with machine learning.

A well-studied problem that is relatively close to the one we address here is the analysis of retinal images. This is due to the nature of the features to be classified and the similar topology between blood vessels in retina and hair on scalp. Goldbaum et al. [11] presented a system of automated diagnosis of retina images. It consists in segmenting and extracting objects of interest and classifying these objects using Bayesian networks that were trained on example images of each disease. Staal et al. [12] introduced a method for segmenting the vessels in color retina images. Their approach is based on the extraction of image ridges and the classification of pixel feature vectors using a kNN-classifier and sequential forward feature selection. The authors in [13] applied a cellular neural network in retina images for detecting various symptoms of diabetic retinopathy. In another approach, Ravishankar et al. [14] performed diabetic retinopathy detection using image processing techniques based on morphological operations and color properties. They performed their classification using an extracted color model and the geographical relationships between different features.

In other medical applications, Hance et al. [15] performed unsupervised color image segmentation applied to skin tumor borders. Lezoray et al. [16] presented a method for color object segmentation, feature extraction and classification for cellular sorting. Using color watershed and choosing the most appropriate color space, they performed the segmentation while the classification was carried out by a binary neural network. In [17], a shallow neural network is used to classify lung cancer in images after segmentation based on morphological operations.

An analogous concept to hair tone estimation is automatic grading of beef marbling, where a beef meat image is being classified in respect to a discrete scale of a perceptual attribute. Under this scope, Yoshikama et al. [18] proposed to segment the image in lean and fat regions by histogram thresholding and to classify the extracted features using regression on run-length histograms. Pang et al. [19] presented another approach that uses morphological operations and adaptive thresholding for fat segmentation and a support vector machine (SVM) classifier for grading the images.

More recently, great advances have been made in the field of computer vision by using deep convolutional neural networks (CNNs) [20, 21, 22]. Due to these accuracy and performance improvements of CNNs, a wider range of applications benefited from deep learning the past few years. In particular, U-Net [23] proposed a supervised neural network that segments cells on microscopic images, and achieved state-of-the-art results. A similar convolutional neural network has been applied for segmenting vessels on retina images [24] and obtained promising results. In skin applications, [25] recently outperformed skin cancer detection by using a convolutional neural network trained on melanoma pictures, acquired by a medical imaging device. Even though the previous architectures were proposed for classification and segmentation, deep learning methods gave interesting results as well in color-related problems. In the field of color constancy, [26] outperformed previous approaches for estimating the illuminant of a picture in-the-wild. However, the aspect of our work is to estimate attributes of the object's color and not the light source.

Despite the wide use of similar techniques in other computer vision and machine learning applications in the literature, a system for hair tone prediction remains an challenging issue that has not been addressed sufficiently. The complex nature of hair requires an appropriate method that is well optimized for this task. In this paper we propose a method that is a based on a CNN and provides sufficiently accurate results and low computation times.

## Problem

The objective of this research is to automatically estimate the natural hair tone of a person. This feature is expressed as a value from 1 to 10, ranging from black hair to the lightest blond (see Figure 1), as described in [27]. This perceptually linear notation applies well to most natural hair, since their color is directly related to the two melanin pigments that are present in human hair and is bound by the color space of hair colors existing in nature [3]. In practice, the hair tone varies from the roots to the

tips, mainly due to exposure to environmental conditions and the possible application of hair coloration products. In this paper, a special focus was made on hair tones at roots, since it is the area where the natural hair color can be observed avoiding, for example, color fading caused by the sun.

Figure 1: Hair tone color chart. For each hair tone from 1 to 10, a natural hair swatch provides a standard example.

The common approach in hair salons is to visually assess the hair tone. However, such assessment is difficult, since the human visual system depends on the illumination environment and may be perturbed by the background (e.g. the underlying scalp color). To demonstrate this limitation, a study was conducted with 6 hairdressers assessing the hair tones of 35 volunteers, under various light conditions. To compare with our reference notation, the hairdressers did the assessment twice, with and without the hair color chart (Figure 1). This chart consists of 10 natural hair swatches as described in [27]. We compare their average estimation with and without using the standard hair swatches. Figure 2 shows their predictions in the two cases; even for the same set of hairdressers on the same volunteers, we see a low concordance between the two protocols, with differences between two averages often close to 1 hair tone. We expect that the estimation using the color chart is more aligned to the standard hair tone notation [27], but we cannot be certain after this study since no ground truth was included. We observe that assessment by eye in salons is not standardized, and that using a color chart is not sufficient. More conclusions about the protocols based on this preliminary study will be discussed in the next section.

To solve this need for a standardized and accurate measurement, a dedicated device was designed to better equip the hairdressers. Standard spectrophotometer devices are capable of measuring the color with high accuracy but they are averaging its value over the complete area under measurement, usually on the order of 1 $cm^2$. Such devices cannot be used near the scalp for root measurements, as they would average the whole area, mixing hair and scalp values and being highly dependent on the exact area being measured. For this reason, we designed a non-invasive device that takes images on the subject's head. The device uses a telecentric lens design allowing the detection of hair from a single viewing angle over to the whole field of view. This allows the uniformity in lighting while minimizing distortion effects over the whole image. The illumination is done using High-CRI white LED lights at an azimuthal incidence angle of 45 deg and placed

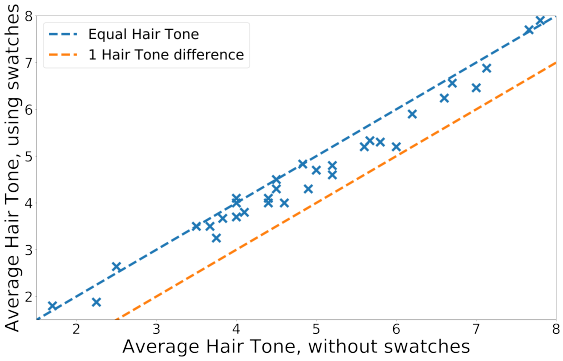



Figure 2: Hair dressers' discordance while evaluating hair tone with and without the standard hair color chart.

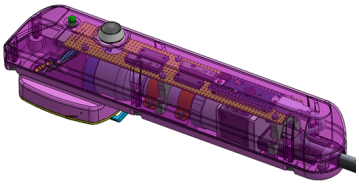


Figure 3: The proposed imaging device for hair-scalp images.

in two 90 deg quarter circles facing each other. The picture of an object area $12 \times 17mm$ is acquired by a $2000 \times 1200$ RGB CMOS sensor. One typical resulting picture is shown in Figure 4. Before acquisition, a line is parted on the subject's head in order to have an unobstructed view on the hair roots. As a consequence, pictures have an axial symmetry around their middle, and are thus generally oriented, with scalp in the middle and more hair in top and bottom.

The device has been designed to estimate an unbiased hair tone value. However, this hair tone is not directly readable from the acquired images since they present a mix between hair and scalp. Moreover, because of hair transparency the scalp color can influence the visible hair pixels. Hence, a method is required for analyzing images with the specific objective of obtaining accurate hair tones. For designing and evaluating such a model, we acquired an extensive amount of data, as described in the next section.

## Data Acquisition

Before acquiring images at roots paired with hair tone values assessed by color experts, we conducted an initial experiment to design the protocol. After concluding that hairdresser visual assessment highly varies with and without using a standard color chart (see previous section), we wanted to find the best conditions for visual human estimation of hair tone at roots. To do so, we

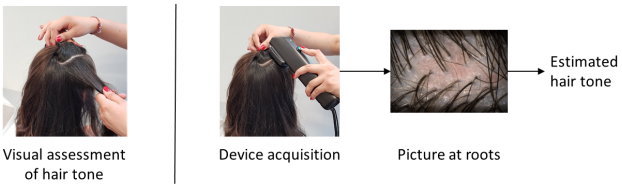



Figure 4: Left side: Visual assessment of hair tone (low precision, not calibrated). Right side: Estimate hair tone using our imaging device.

asked hairdressers to appraise the hair tone at roots on 35 volunteers. These hairdressers - hair colorists with years of experience in their field - are not color experts, so their assessment is representative of real-world salons. We compared 3 protocols:

- Protocol A : Hair salon conditions without hair tone chart
- Protocol B : Hair salon conditions using hair tone chart
- Protocol C : Controlled light conditions using hair tone chart

The uncontrolled light protocols were held inside two rooms with non-standardized light, being cold for a first half (fluorescent lights) and warm for the second half (incandescent lights), to reproduce the varying light conditions in salons. The last protocol was conducted in a gray room with standardized D65 illuminants, which are simulating the sun's mid-day spectrum and are used as a standard in colorimetric measurements. For each scenario, Table 1 shows the standard deviation among hairdressers. We observe that the most reproducible conditions are the ones of the last protocol stressing the importance of measurements under controlled lighting conditions. Using the hair color chart increases slightly the reproducibility as well, but mainly ensures that we are more aligned with our reference notation.

Table 1: Comparisons of protocols with 6 hairdressers assessing hair tone at roots of 35 volunteers. Standard deviation among hairdressers is computed for each volunteer, then averaged across all volunteers.

| Protocol | A | B | C |
|---|---|---|---|
| Average Standard Deviation | 0.65 | 0.61 | 0.35 |

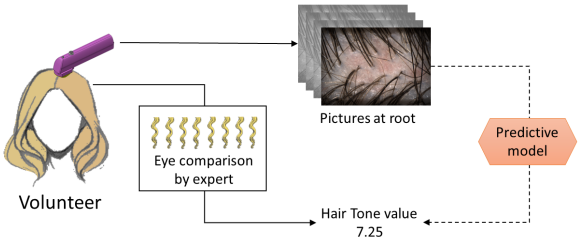



Figure 5: The acquisition protocol. Approximately 50 images are taken at roots by the device, while a color expert assesses the hair tone at roots using a hair color chart. Solid lines point to the input and output of our model, while the dashed line represent the relation that our model estimates.

During our data acquisition sessions, color experts assessed the hair tone of volunteers in a controlled environment under D65 illuminant and using the standard color chart (Protocol C). The first session was performed in our laboratory facility (St. Ouen, France) with 4 color experts, and a standard deviation of 0.12 was obtained between experts notations. This evoked that trained color experts were dramatically more reproducible and consistent than hairdressers. Therefore, the following sessions were conducted with a single color expert notation. Besides this hair tone assessment, approximately 50 images of hair near the scalp were taken on a parting line for each volunteer, as shown in Figure 5. Five sessions were organized in various places around the world in order to cover a wide range of hair tones in our data. Darker hair tones where measured in St. Ouen, France and Shanghai, China; and lighter hair in Warsaw, Poland; and, Karlsruhe and Munich, Germany. The locations were chosen following world hair color studies [28]. The diversity of the hair tones in our data set as well as the number of participants can be seen in the histogram of Figure 6. The total data set is composed of $11,175$ pictures from 407 different volunteers.

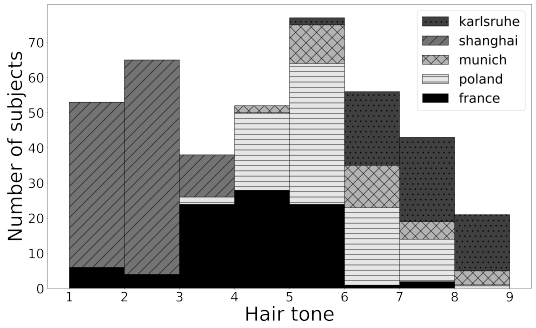



Figure 6: Volunteers histogram grouped by the 5 regions of our data set. On x-axis is the hair tone, which shows the complementarity of each region.

The obtained data are essential in order to address the hair tone estimation using statistical approaches. Our approach is described in detail in the next section.

## Model

In this section we present our method for hair tone estimation. Since the hair tone is separated in 10 categories in our hair color chart, one could address this problem as a classification problem. However, color experts assessed tones with a precision of $1/4$ hair tone, so considering 1 to 10 classes would reduce this accuracy in the labels $h$. Hence, we address the problem as a regression, which assumes an order between the values of $h$. Consequently, our models estimate a real-valued hair tone:

$$\hat{h}(I) \in [1, 10],$$

where $I$ is an input image acquired by our device.

We propose in the following a deep learning approach for hair tone estimation $\hat{h}(I)$. It is based on a Convolutional Neural Network (CNN) which treats the image $I$ in a holistic manner in order to statistically find the relevant patterns that are related to the hair tone $h$. It has no prior knowledge of the problem, but has a total flexibility in the patterns it finds.

### *Convolutional Neural Network*

Following recent advances of deep learning in computer vision (see Related Work), we present the architecture we designed for this problem. The intuition behind such an approach is to learn which patterns in the image are related to the hair tone, with no prior assumption on the patterns to extract. In particular, the model does not have prior knowledge of the problem, the optimization process is purely statistical and learns the patterns in an image $I$ autonomously in order to estimate correct hair tone $\hat{h}(I)$. The basic idea of such a CNN is to apply successive linear

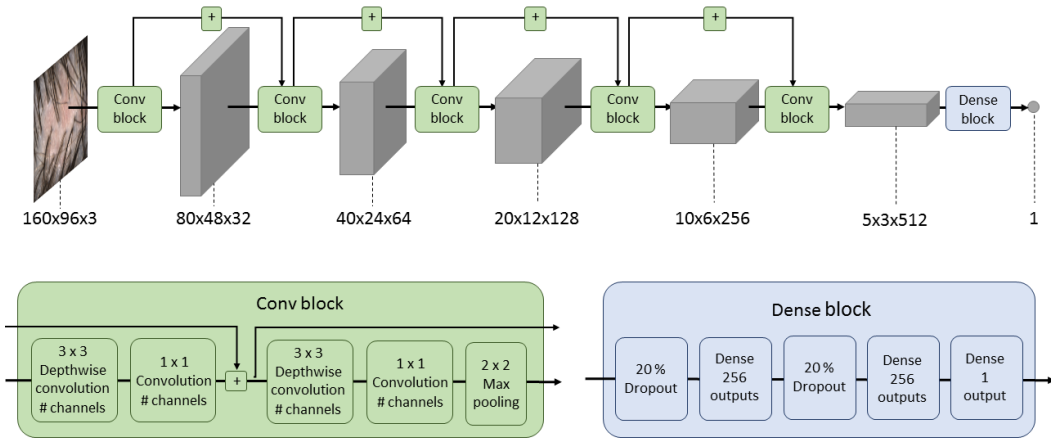



Figure 7: Neural network architecture proposed for hair tone estimation.

operations, such as $3 \times 3$ convolutions, with non-linear functions in-between them as well as operations reducing the spatial dimension of the image representation. Such operations produce feature maps of the image, which is a representation into visual patterns. Such representation is then taken as input of some linear regressions, whose real-valued output is statistically mapped to labels $h_i$ during the training process.

All the operations of the neural network we designed are represented as successive layers in Figure 7. It retakes the basic principles of VGG16 architecture, enriched with residual connections [21]. Indeed, they ease the optimization - direct links in back propagation - and extend model representation with simpler patterns. We opted for simple additions of some layers outputs to a latter layer, since it is a light operation requiring no additional weights.

Compared to state-of-the-art CNN models (see Related Work), we opted for a lighter architecture in two aspects. First, it has less convolutional layers; this is motivated by the fact that we are looking for simpler patterns such as edges and thin fibers, requiring fewer successive convolutions to be represented by the network. The final dense layers are also reduced, since we have an one-dimensional output $\hat{h}(I)$. Second, for speed issues, we use separable 2D convolutions, which has been proposed in [29] as a lighter alternative to regular 2D convolutions. They have less parameters to learn and require less computation. The idea is to first apply a $3 \times 3$ convolution channel by channel (also called depthwise), thus lighter than a full convolution, and then combine the resulting channels by a $1 \times 1$ full convolution. This variant of convolution will speed up the training process as well as the embedded prediction on device.

We now briefly explain the optimization process of such a model, in order to provide some intuition of how the neural network learns the relevant patterns for our problem. If we take a step back from the successive operations, we see that our estimated hair tone can be written as

$$\hat{h}(I) = g(I, \Theta), \tag{1}$$

where $g$ represents all successive operations of our neural network and $\Theta$ is a vector that regroups all the variable parameters, *viz.* convolutions and dense layers weights. In order to find appropriate values for $\Theta$, we define a loss function $\mathscr{L}(\hat{h}(I), h)$ that penalizes predictions $\hat{h}(I)$ far from ground truth $h$. In our case, we minimize the mean square error $\mathscr{L}(\hat{h}(I), h) = (\hat{h}(I) - h)^2$ over all images $I_i$ and hair tones $h_i$ of our data-set:

$$\min_{\Theta} \sum_{i=1}^{M} \mathscr{L}\left(\hat{h}\left(I_i\right), h_i\right) = \min_{\Theta} \sum_{i=1}^{M} \left(g\left(I_i, \Theta\right) - h_i\right)^2. \tag{2}$$

Our strategy to solve this minimization is to update $\Theta$ using an iterative stochastic gradient descent:

$$\Theta' = \Theta + \varepsilon \widehat{\nabla_{\Theta} \mathscr{L}}, \tag{3}$$

where $\varepsilon$ is the learning rate and $\widehat{\nabla_{\Theta} \mathscr{L}}$ is an approximation of the gradient on the full data set. Indeed, this complete gradient $= \sum_i \nabla_{\Theta} \mathscr{L}\left(\hat{h}\left(I_i\right), h_i\right)$ would be too long to compute for each update. Using back-propagation [30], the gradient can be computed efficiently for a mini-batch of size $m$ images. $\Theta$ is thus updated at each mini-batch by considering a moving average of the gradient

$$\widehat{\nabla_{\Theta} \mathscr{L}}' = \lambda \widehat{\nabla_{\Theta} \mathscr{L}} + (1 - \lambda) \frac{1}{m} \sum_{i \in \mathscr{B}} \nabla_{\Theta} \mathscr{L}\left(\hat{h}\left(I_i\right), h_i\right), \tag{4}$$

where $\lambda$ is the momentum and $\mathscr{B}$ is the set of indices $i$ of the mini-batch - $\mathscr{B} \subset \{1..M\}$ and $|\mathscr{B}| = m$. Each mini-batch update of the weights is thus based on a history of the gradient and not only on the mini-batch gradient $\sum_{i \in \mathscr{B}} \nabla_{\Theta} \mathscr{L}\left(\hat{h}\left(I_i\right), h_i\right)$, which presents high variations through mini-batches.

In practice, we passed 800 times on all samples of our data set, with a learning rate of $\varepsilon = 10^{-4}$ and using a momentum of $\lambda = 0.9$. At the end, in order to tune further the parameters $\Theta$, we passed 20 additional times on our data set with a learning rate divided by 10. During each pass, each image $I_i$ is randomly flipped horizontally and vertically in order to augment artificially our data set. For our successive layers and the optimization process we used the implementation of Keras [31] based on the Tensorflow back end [32]. We did not use pre-trained weights since our architecture does not match those of existing CNNs. Moreover, ImageNet pictures and classes that are used for pre-training are importantly different from our images. In particular, the visual patterns our model needs to focus on are simpler in terms of shape, but more related to colors compared to available data sets.

## Experiments

### *Conventional Approaches*

In order to better evaluate the proposed method various conventional approaches were considered. The idea behind these approaches is to transform the imaging device into a colorimetric device that is able to provide standard color values of hair and relate them with the perceptual attribute of hair tone. Two factors whose impact was particularly tested are the segmentation of hair pixels in the image and the color space under which the hair pixel values are examined. These two steps are described in the two following subsections.

#### *Hair segmentation*

The main objective of this step is to demonstrate the effect of considering only the hair pixels in the image. Various segmentation methods can be applied but in this section in order to avoid outliers we tune our method to instead of segmenting all hair pixels in the image to robustly segment a large number of hair pixels. This ensures that segmentation avoids specularities and other artifacts such as dandruff. Our segmentation method is based on adaptive thresholding [33] in order to vastly segment hair form scalp areas. Before thresholding, a Gaussian filter is applied to reduce the noise in the image. Finally, the resulting hair regions are ranked by size using connected components analysis in order to eliminate small detected areas that could be specularities or other artifacts falsely segmented as hair. The output of this method can be seen in Figure 8. Different segmentation methods were considered and we found that the accuracy of the segmentation is not crucial. This is due to the fact that taking into account the median hair value one can neglect possible outliers and the error in the final result is negligible. However, in order to isolate the segmentation impact, in our experiments we consider this method that eliminates outliers.

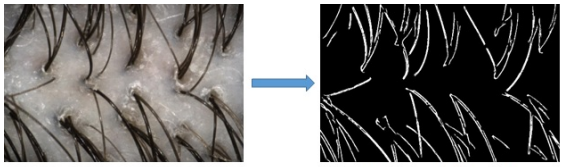



Figure 8: Our segmentation method for extracting a large number of hair pixels while avoiding outliers.

#### *RGB to CIELAB*

The device needs to be calibrated in order to transform the device-dependent *RGB* values to a device-independent color space. The CIELAB color space was selected for this transformation since it is relatively perceptually uniform and is widely used in the industry. There are various methods in the literature that provide $L^*a^*b^*$ measurements of hair, such as in [34] and [35]. For the calibration process, we estimate the $L^*a^*b^*$ value $\mathbf{r} \in \mathbb{R}^3$ of each *RGB* pixel expressed as $\mathbf{x} \in \mathbb{R}^3$, using the following model:

$$\hat{\mathbf{r}}(\mathbf{x}) = \mathbf{C_r}\boldsymbol{\phi}(\mathbf{x}), \tag{5}$$

where $\boldsymbol{\phi}(\mathbf{x}) \in \mathbb{R}^N$ is a third degree polynomial ($N = 20$) of the *RGB* pixel values $\mathbf{x} \in \mathbb{R}^3$ and $\mathbf{C_r} \in \mathbb{R}^{3\times N}$ are the corresponding coefficients, learned statistically on a set of $K$ training pairs. Those pairs were obtained by acquiring images of predefined color patches with measured $L^*a^*b^*$ values $\mathbf{r}_i$ ($i = 1,\ldots,K$). For each patch $\mathbf{x}_i$ acquired by the device we take the median *RGB* pixels in order to solve the regression problem:

$$\underset{\mathbf{C_r}\in\mathbb{R}^{3\times N}}{\arg\min} \sum_{i=1}^{K} \|\hat{\mathbf{r}}(\mathbf{x}_i) - \mathbf{r}_i\|^2, \tag{6}$$

which is the closed-form equation of $\mathbf{C_r}$ and $\|.\|^2$ denotes the $L^2$ norm. The patches used in our experiments were the 24 patches of the Macbeth Color Checker and the 110 pacthes of the PANTONE SkinTone$^{\text{TM}}$ Guide. The latter was used due to its similarity with the scalp and hair colors that are present in our images. The accuracy of this transformation was evaluated on a small test set of 20 hair colors where the average error was found to be $\Delta E_{ab} = 1.31$.

### *Compared Models*

In order to evaluate the performance of our CNN, we compared it with the conventional methods described above and with existing popular deep learning models. We thus considered the following models:

- *RGB* : As a baseline, we evaluated a regression on the mean red, mean green and mean blue values for a given input picture. We used a Ridge regression on a 3rd-degree polynomial of these mean *RGB* values. This method mixes scalp pixels with hair pixels.
- *RGB*-Seg : Regression of 3rd-degree polynomial based on mean *R*, *G* and *B* values of the segmented hair pixels, for a given input picture.
- $L^*a^*b^*$-Seg: Regression of 3rd-degree polynomial based on mean $L^*$, $a^*$ and $b^*$ values of the segmented hair pixels, for a given input picture.
- Deep-HT : Our proposed method that is based on a CNN.
- VGG 16 [20], Inception [22], MobileNets [29] : These baseline architectures were used for the first part of the network, the convolutional part, which learns a representation of the image as feature maps. The final dense block is kept as for Deep-HT, without the dropout regularization. Regularization is indeed present in the convolutional blocks of Inception and MobileNets, and absent from the original architecture of VGG. The training procedure is kept the same, as detailed in the description of our model. All CNN models are trained from scratch, with no pre-training, since the images acquired by our device are very different with ImageNet pictures that are used for pre-training.

### *Evaluation Strategy*

In order to evaluate and compare our models, we performed a 5-fold cross-validation [36] on the 11,175 images of our data set acquired following the protocol in Section Data Acquisition. To avoid bias of having different pictures of the same volunteer in both training and testing, images from each of the 407 volunteers were used only in the same fold. Moreover, we stratified the validation, by distributing equally the 10 classes of hair tone into each fold.

After the cross-validation, we computed three metrics on hair tones $h_i$ predicted on test folds: maximum error MAX $= \max_i |\hat{h}(I_i) - h_i|$ , mean absolute error MAE $= \sum_i |\hat{h}(I_i) - h_i|$ and root mean square error RMSE $= \sqrt{\sum_i (\hat{h}(I_i) - h_i)^2}$ against the ground truth. We computed these errors in the 2 different cases:

- /Img : Error comparing each image to its corresponding ground truth. When considering RMSE, this metric is the training loss of the deep learning models.
- /Vol : Error comparing the average hair tone over all pictures of the same volunteer to the ground truth. This case is optimistic in real-world usage since this would require approximately 50 pictures per subject to achieve similar errors.

### *Comparative Results*

Table 2: Comparisons of the different models for hair tone prediction. In bold are the best results for each metric.

| Model | RMSE | | MAE | | Max | |
|---|---|---|---|---|---|---|
| | /Img | /Vol | /Img | /Vol | /Img | /Vol |
| RGB | 1.21 | 1.09 | 0.96 | 0.90 | 5.31 | 3.11 |
| RGB-Seg | 0.73 | 0.66 | 0.58 | 0.48 | 4.06 | 1.88 |
| L*a*b*-Seg | 0.74 | 0.67 | 0.58 | 0.48 | 4.42 | 1.92 |
| Deep-HT | **0.51** | **0.43** | **0.39** | **0.33** | 3.25 | 1.63 |
| VGG16 | 0.55 | **0.43** | 0.43 | **0.33** | **3.16** | **1.61** |
| Inception | 0.60 | 0.51 | 0.45 | 0.37 | 4.20 | 2.50 |
| MobileNets | 0.57 | 0.46 | 0.44 | 0.35 | 2.85 | 1.77 |

We see in Table 2 that the highest accuracy is obtained by our Deep-HT model. This appears to us as the shallowest CNN model, with fewer operations and parameters used, that is sufficient for our simple patterns in hair/scalp images, and is therefore easier to optimize. For similar reasons, the second best model is VGG, despite its lack of regularization. The conventional approaches provide worse results but of the same order, validating our simple hair image analysis models that require significantly fewer parameters. This confirms that mean values of hair pixels are strongly related to volunteer hair tone and validates the role of hair segmentation in our problem. However, the regression performed in the CIELAB color space does not increase the accuracy as it is further argued in the next section.

Specifically, we see that the mean absolute error on one picture is of 0.39 for Deep-HT; extremely close to the one obtained by hairdressers under standard light condition and using the hair color chart. This is very satisfactory as we highly improve on the accuracy of hairdressers in real-world conditions (Section Data). Moreover, our approach provides a standardized notation, since it has been trained on our color experts labels. Besides this accuracy comparison, we evaluate the performance in terms of computational speed in the next section.

### *Computational Costs*

In this section, we compare the performance of our models in three setups, on the training hardware and on the low-end mobile chip used in the device. We focus on the computational time of the model itself, and discard all necessary steps such as I/O. More precisely, we computed the following times:

- Train : the total training time on one fold of the evaluation. GPU Nvidia Tesla K80 was used for the CNN calculations. For the colorimetric approach, the hardware used for training was a 24 CPU cores Intel E5-2690v3.
- Test : The average prediction time per picture computed on the full data set. The same setup was used as above. This prediction time is optimistic since it was computed on high-end hardware as well as on mini-batches of 16 images instead of single ones.
- On-device: The calculation was performed on a mobile SoM to evaluate the feasibility of performing hair tone predictions on the device at a hair salon. For this, we measured average time required over 10 successive predictions using a Qualcomm Snapdragon 410 - a quad-core 1.2 GHz CPU with 1GB of RAM. Such a chip is already required in the device for managing the captors, screen, and lights. We implemented the neural network models using TensorFlow Mobile [32] for Android.

Table 3: Comparisons of computational costs for each model, on the hardware used for training and on low-end device chip. Only the CNN models were implemented on the chip.

| Model | #Parameters | Train | Test | On-device |
|---|---|---|---|---|
| RGB | 20 | 34s | 3ms | - |
| RGB-Seg | 20 | 36m | 239ms | - |
| L*a*b*-Seg | 20 | 42m | 277ms | - |
| Deep-HT | 2,571,004 | 15h48m | **1.9ms** | **177ms** |
| VGG16 | 16,747,073 | 13h24m | 7.1ms | 1141ms |
| Inception | 31,271,841 | 27h06m | 10ms | 1280ms |
| MobileNets | 7,205,441 | 7h17m | 3.1ms | 275ms |

The results are displayed in Table 3. The Deep-HT provides the fastest calculations, even when compared to MobileNets that includes more layers and parameters. When considering accuracy and performance, Deep-HT is thus the most suited to our problem. We note that computation time on device is satisfactory for real-world usage, since 170*ms* would be barely noticeable by the user.

## Discussion

Figure 10 represents the boxplot of average absolute errors by volunteer, for all predictions of Deep-HT on the test folds. We see that for most ranges we achieve approximately 0.25 error, apart for very high tone values that were less frequent in our data set (see Figure 6). This meets our industrial requirements as the hairdressers had a standard deviation of 0.65 in salon conditions. Moreover, our model error is computed against a reliable ground truth, given by color experts using hair color chart. This ensures our prediction to be a standard and objective estimation, with satisfactory accuracy.

To understand better the difference between CNN models and conventional approaches, we computed the saliency map [37] of the Deep-HT model on a given picture $I_i$, as

$$\frac{\partial \hat{h}}{\partial I}(I_i)$$

which has the same size as $I_i$. In order to display a 2D heat-map with one channel, the maximum value over the three input channels is considered. This derivative is computed with a single back

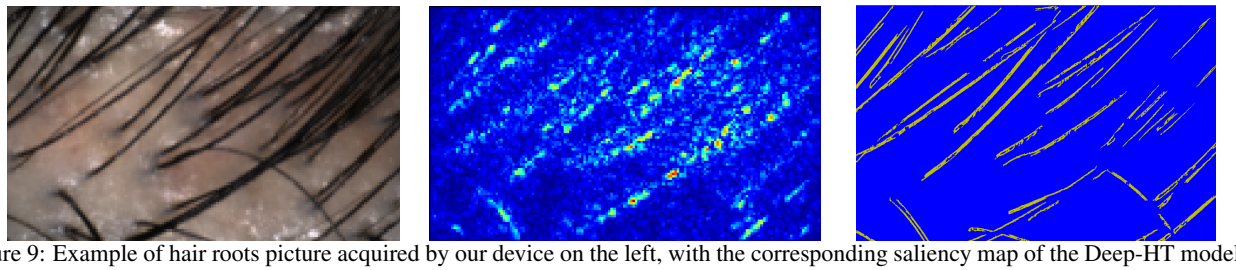


Figure 9: Example of hair roots picture acquired by our device on the left, with the corresponding saliency map of the Deep-HT model in the middle, and the hair pixels segmentation used in our conventional approaches on the right.

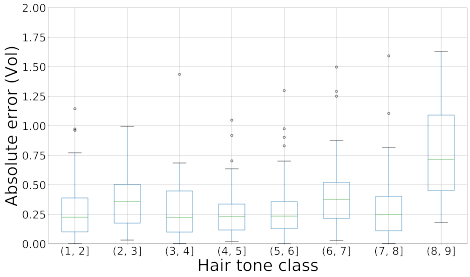



Figure 10: Boxplot on all data-set, deep-HT

propagation, and gives higher weights to pixels in $I_i$ that have higher influence on hair tone prediction $\hat{h}(I_i)$. This visualization is to be compared with the segmented hair used in our conventional methods as shown in the right of Figure 9.

One can observe that no scalp pixels are segmented as hair in our segmentation method since the objective was to focus on the impact of considering only the hair pixels. In comparison, deep learning seems to focus on hair pixels as well, but the selection is more variable, with a few broader areas of interest. To some extent, this confirms that hair contains the information as we assumed in the conventional approaches. We also see that the convolutional model manages to capture subtler details which might explain the differences in estimation accuracy.

The assumption that segmenting the hair has an important impact in the precision was validated by the results of our experiments, However, as can be seen in Table 2, the device-independent CIELAB color space does not perform better than the device-dependent *RGB* values. This is related to the findings of Lozano et al. [35] that despite the strong correlation between the values $L^*b^*$ and natural hair tone ($a^*$ plays a significant role only on red hair), the $L^*a^*b^*$ clusters between different hair tones are highly overlapping. Especially in the range of darker hair colors the distinction is not accurate by simply using these colorimetric measurements and other attributes need to be taken into account (diameter, shape and shine). These attributes can be better captured by the more holistic approach of a CNN. Moreover, it should be noted that the transformation from *RGB* to CIELAB space, even though it was specifically designed for this problem, is always subject to additional errors. Since CIELAB is a relatively perceptually uniform color space, differences in dark colors are suppressed. This might explain the similar performance between the *RGB* and $L^*a^*b^*$ values.

In terms of computational costs, we observe that the segmentation we used was the most expensive part of the conventional approaches. Despite having the benefit of being unsupervised, the segmentation step is significantly longer to compute than the CNN models. For this reason, at test time, the CNN models are faster. Furthermore, they benefit from the recent development of deep learning frameworks. We found that Deep-HT is the fastest model due to its lighter architecture that still captures enough information of the hair roots images to provide high accuracy.

## Conclusion

Despite the complex challenging nature of hair, we propose an accurate method based on deep learning for automatic hair tone estimation using hair-scalp images. Moreover, we evaluated our method in comparison with conventional color image processing methods and other popular deep learning techniques. The results show that our deep learning model is not only more robust and precise than our accuracy objective (the subjectiveness of the average hair dresser) but also outperforms the other tested approaches. Moreover, the light network architecture of our model achieves a low calculation time that allows the real time estimation on the acquisition device. The conventional approaches perform worse due to their simpler pattern extraction based on unsupervised hair/scalp segmentation. Indeed, focusing on the color of hair pixels achieves a decent estimation but fails to take into account more global effects such as transparency and geometry. The complex hair-light interaction cannot be adequately described by a single value in a color space. On the other hand, the deep learning method takes better advantage of our large annotated database to capture subtler patterns in hair pictures.

Such an imaging device with embedded computation has an immense potential in hair salons and breaks new ground in hair care and hair color personalization. Our future work will be to extend this deep learning approach on new attributes of the hair at their roots, including fiber diameter, density and white hair percentage.

## Acknowledgments

We would like to thank all the contributors, color experts and evaluation teams in France, Poland, Shanghai and Karlsruhe. Special thanks go to Tal Marcu who conducted the initial experiments with hairdressers and to Boris Krasny who started the algorithmic analysis of the data, participated in the first measurement sessions in Paris and helped building the first device. Our gratitude goes to Asiya Shakhmametova and Yohanna Pachas for accepting to be photographed for the needs of this paper.

## References

[1] D. Szutowski and J. Szułczyńska, "Product innovation in cosmetic industry – case study of major cosmetic companies," pp. 19–21, 01 2017.

[2] S. Harrison and R. Sinclair, "Hair colouring, permanent styling and

hair structure," *Journal of Cosmetic Dermatology*, vol. 2, no. 3-4, pp. 180–185, 2003.

[3] C. Bouillon and J. Wilkinson, *The Science of Hair Care*. Taylor & Francis, 2005.

[4] D. S. MacFarlane, D. K. MacFarlane, and F. W. Billmeyer Jr, "Method for correctly identifying hair color," May 23 2000, uS Patent 6,067,504.

[5] R. S. Hunter and R. W. Harold, *The Measurement of Appearance*, 2nd ed. New York: John Wiley & Sons, 1987.

[6] S. Leprince, "Apparatus and method for providing hair tinting information," Apr. 3 2003, uS Patent App. 10/098,574.

[7] M. Ladjevardi, "Advanced cosmetic color analysis system and methods therefor," Dec. 19 2006, uS Patent 7,151,851.

[8] H. Sato, "Hair image display method and display apparatus," Apr. 23 2013, uS Patent 8,428,382.

[9] E. K. Ross, C. Vincenzi, and A. Tosti, "Videodermoscopy in the evaluation of hair and scalp disorders," *Journal of the American Academy of Dermatology*, vol. 55, no. 5, pp. 799–806, 2006.

[10] E. N. Malamas, E. G. Petrakis, M. Zervakis, L. Petit, and J.-D. Legat, "A survey on industrial vision systems, applications and tools," *Image and Vision Computing*, vol. 21, no. 2, pp. 171–188, 2003.

[11] M. Goldbaum, S. Moezzi, A. Taylor, S. Chatterjee, J. Boyd, E. Hunter, and R. Jain, "Automated diagnosis and image understanding with object extraction, object classification, and inferencing in retinal images," in *Proceedings of 3rd IEEE International Conference on Image Processing*, vol. 3, 1996, pp. 695–698.

[12] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, April 2004.

[13] D. Matei and R. Matei, "Detection of diabetic symptoms in retina images using analog algorithms," *World Academy of Science, Engineering and Technology*, vol. 45, pp. 408–411, 2008.

[14] S. Ravishankar, A. Jain, and A. Mittal, "Automated feature extraction for early detection of diabetic retinopathy in fundus images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 210–217.

[15] G. A. Hance, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Unsupervised color image segmentation: with application to skin tumor borders," *IEEE Engineering in Medicine and Biology Magazine*, vol. 15, no. 1, pp. 104–111, 1996.

[16] O. Lezoray, A. Elmoataz, and H. Cardot, "A color object recognition scheme: application to cellular sorting," *Machine Vision and Applications*, vol. 14, no. 3, pp. 166–171, 2003.

[17] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for ct images," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 202–209, 2014.

[18] F. Yoshikawa, K. Toraichi, K. Wada, N. Ostu, H. Nakai, M. Mitsumoto, and K. Katagishi, "On a grading system for beef marbling," *Pattern Recognition Letters*, vol. 21, no. 12, pp. 1037–1050, 2000.

[19] B. Pang, X. Sun, C.-W. Ye, and K. Chen, "Grading of beef marbling based on image processing and support vector machine," *Mathematical and Computer Modelling*, vol. 17, no. 3, pp. 87–92, 2013.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.

[24] M. Melinščak, P. Prentašić, and S. Lončarić, "Retinal vessel segmentation using deep neural networks," in *VISAPP 2015 (10th International Conference on Computer Vision Theory and Applications)*, 2015.

[25] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[26] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 81–89.

[27] S. Panhard, I. Lozano, and G. Loussouarn, "Greying of the human hair: a worldwide survey, revisiting the '50'rule of thumb," *British Journal of Dermatology*, vol. 167, no. 4, pp. 865–873, 2012.

[28] G. Loussouarn, I. Lozano, S. Panhard, C. Collaudin, C. El Rawadi, and G. Genain, "Diversity in human hair growth, diameter, colour and shape. an in vivo study on young adults from 24 different ethnic groups observed in the five continents," *European Journal of Dermatology*, vol. 26, no. 2, pp. 144–154, 2016.

[29] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[30] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, p. 533, 1986.

[31] F. Chollet *et al.*, "Keras," 2015.

[32] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning." in *OSDI*, vol. 16, 2016, pp. 265–283.

[33] D. Bradley and G. Roth, "Adaptive thresholding using the integral image," *Journal of Graphics Tools*, vol. 12, no. 2, pp. 13–21, 2007.

[34] H. L. Norton, M. Edwards, S. Krithika, M. Johnson, E. A. Werren, and E. J. Parra, "Quantitative assessment of skin, hair, and iris variation in a diverse sample of individuals and associated genetic variation," *American Journal of Physical Anthropology*, vol. 160, no. 4, pp. 570–581, 2016.

[35] I. Lozano, J. B. Saunier, S. Panhard, and G. Loussouarn, "The diversity of the human hair colour assessed by visual scales and instrumental measurements. a worldwide survey," *International Journal of Cosmetic Science*, vol. 39, no. 1, pp. 101–107, 2017.

[36] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[37] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

## Author Biography

*Panagiotis-Alexandros Bokaris holds a Diploma in Engineering from the University of Patras (2011) and the Erasmus Mundus Master Degree CIMET (2013). He obtained his PhD in computer science from the*

*University of Paris-Saclay (2016), during which he was a visiting PhD student at the University College London. Since then he joined L'Oréal Research and Innovation where he focuses on augmented reality, computer vision and graphics for beauty personalization.*

## Author Biography

*Emmanuel Malherbe received his engineering degree from Ecole Polytechnique (2012), his MSc from Imperial college (2013) and his PhD in machine learning from Centrale Supelec (2016). Since then he has worked in the Research and Innovation Division at L'Oréal in Saint-Ouen, France. His work has focused on the development of deep learning models for hair images.*

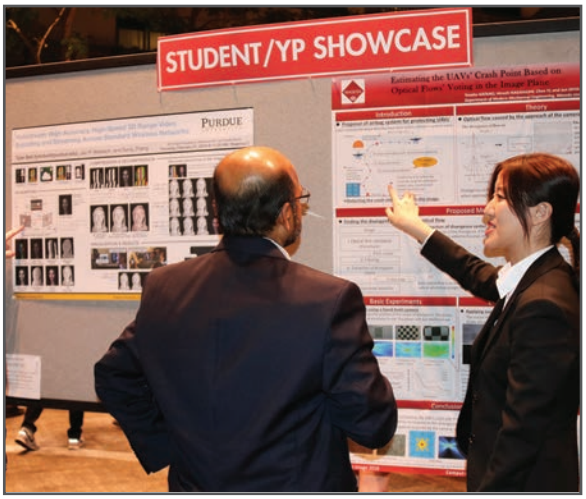
# JOIN US AT THE NEXT EI!

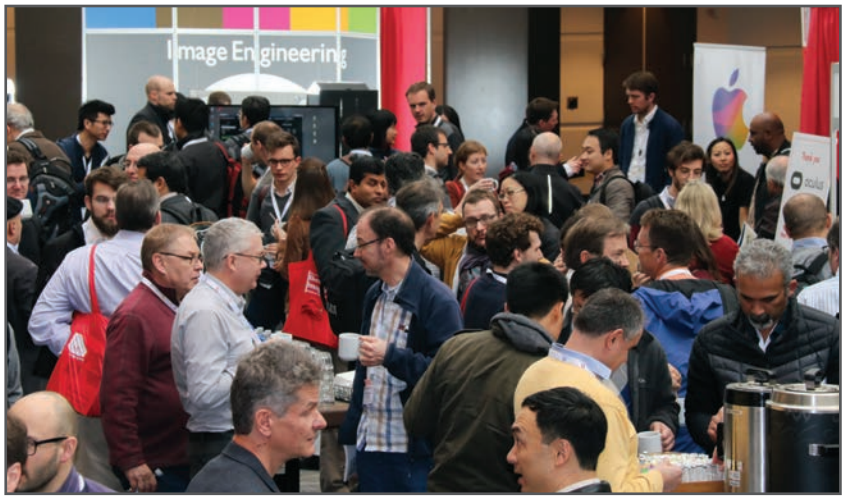
IS&T International Symposium on

# Electronic Imaging

SCIENCE AND TECHNOLOGY

*Imaging across applications . . . Where industry and academia meet!*





**• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
**• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

**www.electronicimaging.org**

IS&T
imaging.org