

ORIGINAL ARTICLE

Tracking skin colour and wrinkle changes during cosmetic product trials using smartphone images

Alan F. Smeaton¹  | Swathikiran Srungavarapu¹ | Cyril Messaraa²  | Claire Tansey²

¹Insight Centre for Data Analytics, Dublin City University, Glasnevin, Dublin 9, Ireland

²Oriderm R&D, Oriflame, Bray, Ireland

Correspondence

Alan F. Smeaton, Insight Centre for Data Analytics, Dublin City University, Glasnevin, Dublin 9, Ireland.
Email: alan.smeaton@dcu.ie

Funding information

Science Foundation Ireland, Grant/Award Number: SFI/12/RC/2289 (Insight Research Centre)

Abstract

Background: To explore how the efficacy of product trials for skin cosmetics can be improved through the use of consumer-level images taken by volunteers using a conventional smartphone.

Materials and Methods: 12 women aged 30-60 years participated in a product trial and had close-up images of the cheek and temple regions of their faces taken with a high-resolution Antera 3D CS camera at the start and end of a 4-week period. Additionally, they each had “selfies” of the same regions of their faces taken regularly throughout the trial period. Automatic image analysis to identify changes in skin colour used three kinds of colour normalisation and analysis for wrinkle composition identified edges and calculated their magnitude.

Results: Images taken at the start and end of the trial acted as baseline ground truth for normalisation of smartphone images and showed large changes in both colour and wrinkle magnitude during the trial for many volunteers.

Conclusions: Results demonstrate that regular use of selfie smartphone images within trial periods can add value to interpretation of the efficacy of the trial.

KEYWORDS

colour normalisation, computer vision, edge detection, skin physiology/structure

1 | INTRODUCTION

Cosmetic companies invest huge resources each year on researching new products that are specific to consumer requirements. The claimed efficacy of these products needs to be substantiated and documented in the Product Information File, as specified by Article 11 of Regulation (EC) No 1223/2009. Clinical trials on volunteers are one of the possible approaches to generate relevant data with regard to the efficacy of cosmetic products. Conventionally, such product trials begin by recruiting volunteers and measuring some baseline characteristics of their appearance in a laboratory setting before having them use the new product, or a placebo, in their everyday lives for a period of time, that can range from minutes (eg make-up, instant effect) to a number of weeks. At the end of the trial

period, the volunteers return to the laboratory for a re-test of their appearance, allowing for a before-and-after comparison, with and without the product being evaluated.

A major problem with this approach is that testers do not know what their subjects have been doing during the trial period because there are no possibilities to monitor their compliance to the protocol. With skincare as the leading category among cosmetic products, if subjects have been sunbathing, or spending time in cold and harsh outdoor conditions which present challenging conditions for skin products then not having control over, or even knowing, this variable reduces the usefulness of such product trials.

This paper explores the capabilities of a consumer smartphone (Huawei P20 PRO) in capturing information on facial characteristics during cosmetic product trials. For this study, volunteers took part in initial and end-of-trial assessments of their skin characteristics

and received skincare products to use during a 4-week trial period. Volunteers' faces were imaged regularly during the trial using the smartphone. Images were analysed by normalising and extracting facial features thus tracking changes in skin characteristics *during* the trial period. The work addresses if these intermediate images add value to before-and-after laboratory assessments as part of product trials.

2 | MATERIALS AND METHODS

2.1 | Subjects and equipment

We recruited 12 female volunteers with ages ranging from 30 to 60 years for this study. Volunteers attended an initial and an end-of-trial facial skin assessment in a laboratory environment and as many of the intermediate smartphone image capture sessions as they could. Each volunteer was given a randomly generated id for anonymity purposes. Ethical approval for this project was granted by the University's School of Computing Research Ethics Committee.

The Huawei P20 PRO smartphone¹ is co-engineered with Leica and operates with Artificial Intelligence (AI) to guarantee the highest outcomes in photography are always achieved. The phone comes with a dual-lens camera, one at 12 MP (RGB, f/1.8 aperture) and the other at 20 MP (Monochrome, f/1.6 aperture) which support contrast focus, phase focus, laser focus and deep focus. Its front camera is a single-lens camera of 24MP with an f/2.0 aperture and supports a fixed focal length. 2D Photographs were taken of our subjects with the camera at a resolution of 5120 × 3840 pixels with f/1.8 aperture.

The Antera 3D camera^{2,3} offers a powerful and easy-to-use technology for aesthetic medicine. It captures close-up images of the skin using an enclosed light hood which means it operates under controlled lighting. It can be used to measure skin wrinkles, texture, scars, the colour of the skin, redness and pigmentation.³ It offers a complete overview of the skin, and was used in the laboratory setting at the start and at the end of the trial period in the work reported here.

2.2 | Protocol

We captured facial image data during a cosmetic product trial using the Antera 3D camera on the first and last days of a 4-week trial period. As we are interested in facial skin changes during the trial, we also captured facial images using the smartphone during the trial period including the first and last days.

On the first day of the trial, volunteers were relaxed in a controlled environment before the left temple and cheek areas of their faces were imaged using the Antera 3D and also the smartphone camera. Following that, each volunteer received an anti-ageing product kit and usage instructions. The product kit comprised a cleanser, toner, eye cream, serum, and day and night creams.

During the trial period, cheek and temple images were recorded for each volunteer up to 3 times per week over a 4-week period, using the Huawei P20 PRO camera with the highest resolution. The

colour values in an image are dependent on the varying lighting conditions, location, environment and other factors. Because the smartphone images were taken under different lighting conditions during the trial period, to normalise colour values in the images and determine the true colour of the skin, we used colour cards⁴ composed of standard colours when taking the images. Volunteers held the colour card against their face, below the nose level for the image of their temple and across the eye for the image of their cheek.

During the trial period, volunteers were imaged at an indoor location on the University campus. The cheek and temple areas for two volunteers, respectively, are shown in Figure 1.

On the last day of the trial period, a similar procedure of image capture to the initial day was repeated using the Antera 3D and the smartphone camera. All 12 volunteers attended both the initial and the final imaging sessions, and volunteers also attended between 5 and all of the 10 scheduled within-trial imaging sessions (mean = 8.5, SD = 1.68). The full attendance record for volunteers is shown in Figure 2. This completed our data collection.

2.3 | Image processing for skin colour

The distribution of colour values in an image depends on the illumination, which varies widely under real-world conditions.⁵ We explored two known colour adjustment and transformation algorithms along with an approach based on normalisation using colour cards which can be applied within an image.

2.3.1 | Histogram equalisation

Histogram equalisation distributes the initial illumination intensity values in an image uniformly, nearly equal to the image's histogram.⁶ The algorithm follows a series of monotonic, non-linear mappings to re-assign the intensity values of pixels in the input image such that the output image contains a flat histogram of intensities.⁷ We used the `equalizeHist()` function implemented in the OpenCV software library.⁸ However, this function cannot be applied directly to images in the RGB or LAB colour space; therefore, we converted the images into YUV colour space, where Y is the luminance component and UV are the colour components. We then equalised the Y component alone. Later, the equalised images were converted back into RGB colour space. Figure 4 shows a sample image and its histogram equalised image.

2.3.2 | CLAHE (Contrast Limited Adaptive Histogram Equalisation)

Histogram equalisation considers global image contrast during the image transformation process. CLAHE is an adaptive method that computes several histograms corresponding to distinct sections in an image and then transforms these sections for redistribution

FIGURE 1 Cheek and temple areas captured using Huawei P20 PRO smartphone with colour card shown

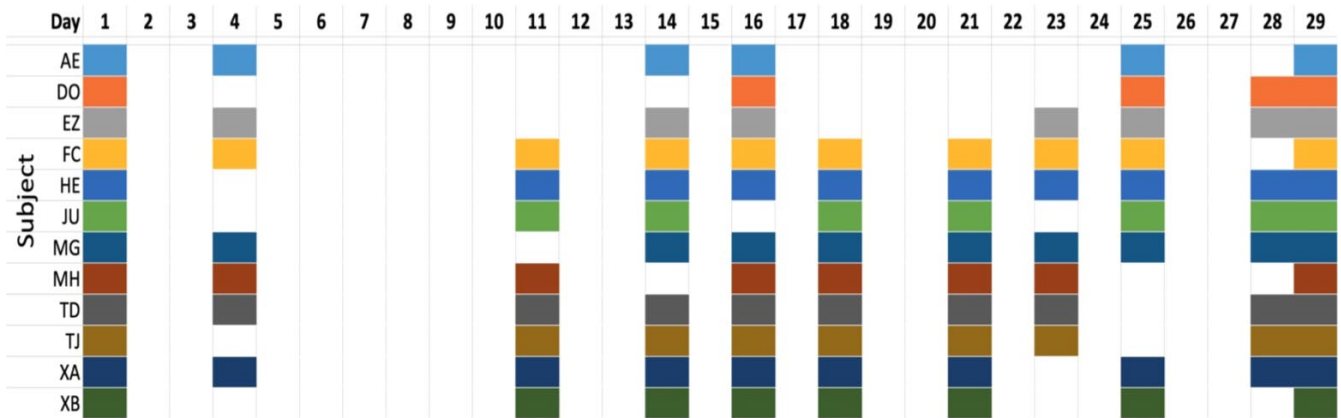


FIGURE 2 Subject participation for within-trial imaging

of the lightness values of the image.⁹ It is therefore suitable for improving local contrast and enhancing the definition of edges in regions of an image where the size of the regions is user-defined. In this work, we used regions of 4×4 pixels during the CLAHE transformations. Figure 4 shows a transformation result using CLAHE.

2.3.3 | Image normalisation using colour cards

Printed colour cards such as the one shown in Figure 3 were used during smartphone image capture. The aim is to measure the differences between the known true colour values from the cards and the values observed in the captured images and to adjust the images accordingly. Ideally, we try to adjust luminance and colours in the

image to account for the real-world conditions in which the images are taken.

If L , A and B are the true values of Black/White, Blue/Red and Yellow/Green, respectively, for the colour purple at row 2 and column 4 of the colour card in Figure 3 and, let P be the same area with the colour purple in a captured image I then.

$$\delta_L = L - L_1$$

$$\delta_A = A - A_1$$

$$\delta_B = B - B_1$$

where L_1 is the calculated mean value of Black/White for area P , A_1 is the calculated mean value of Blue/Red in area P , and B_1 is the

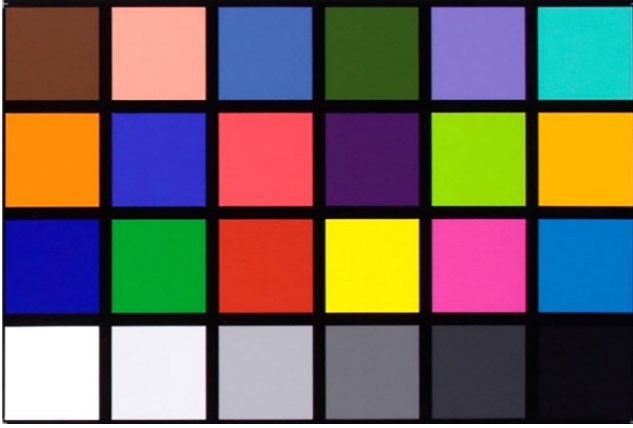


FIGURE 3 Colour card used by volunteers

calculated mean value of Yellow/Green in area P. We let δ be the set consisting of the difference values which are applied to an image captured with a smartphone.

$$\delta \subset \{\delta_L, \delta_A, \delta_R\}$$

Each image of a volunteer captured using the smartphone with the colour card included in the image was transformed using all three normalisation techniques. Figure 4 illustrates that there can be clear differences among the outputs of different techniques.

2.3.4 | Estimating skin colour

Once images are normalised for colour using each of the above methods, the next step is to calculate skin colour values using pixels in the cheek area. Normalised images are maintained in the LAB colour space rather than RGB because the LAB colour space is used by the Antera 3D camera. Every image pixel in the LAB colour space is described by the amount of black/white, green/red and yellow/blue, respectively. The cheek area is used for calculating the average L, A and B values that represent the volunteer's skin colour from each image. These values define the amount of brightness, redness and yellowness of the skin.

2.4 | Image processing for Wrinkle measurement

Feature detection is the process of automatically computing an abstraction of an image and making a local decision at every point in the image to determine whether there is an image feature of a given type which exists at that point.¹⁰ We explored methods to automatically align images of the same volunteer taken during the trial period so we could exactly match the skin areas, and then extract a common area of the face for wrinkle comparison across the set of images thus tracking wrinkle changes across the trial period. In the end, we used a semi-automated technique based on SIFT matching.

Scale Invariant Feature Transform (SIFT)¹¹ is a popular local feature descriptor which can be used to match similar or near-identical images. The general approach of this algorithm is to extract many key points from images and then compute a local descriptor for each of these key-points. Image-to-image matching or scoring is then performed by matching each key-point descriptor in a query image against all descriptors extracted from a target image and summing the matches across point descriptors. Figure 5 shows image matching based on SIFT features.

Following feature detection, we aligned images of each volunteer and detected the common wrinkle area across images taken during the trial period. The wrinkle area in our study is the cropped portion of the temple image consisting only of wrinkles. Once this is completed, images are processed to measure wrinkle magnitude using edge detection, with the following steps.

Edge detection is less complex when working with grayscale images compared to other colour spaces; therefore, the gradient magnitude in this study is calculated using grayscale images.¹² Gradient magnitude is a quantification of the edges in an image where a gradient operator simply produces large amplitudes in the output image indicating edges in the input image. Therefore, the more edges in the image the greater the amplification and vice-versa. We apply the Laplacian method to compute the gradient magnitude.¹³

The Sobel edge detector is a classic computer vision operator which uses a derivative approximation to find the edges in an image. The Sobel operator in a 2-dimensional image works in both directions, that is Sobel-X for left-to-right or horizontal scanning and Sobel-Y for top-to-bottom or vertical scanning for edge detection. Edges detected in both directions at a given point in an image can be combined using the logical OR operator, and this is referred to as Sobel combined. Figure 6 illustrates the output of these operators on a sample image.

Using these attributes of an image, we can estimate a measure for a volunteer's skin wrinkles as follows:

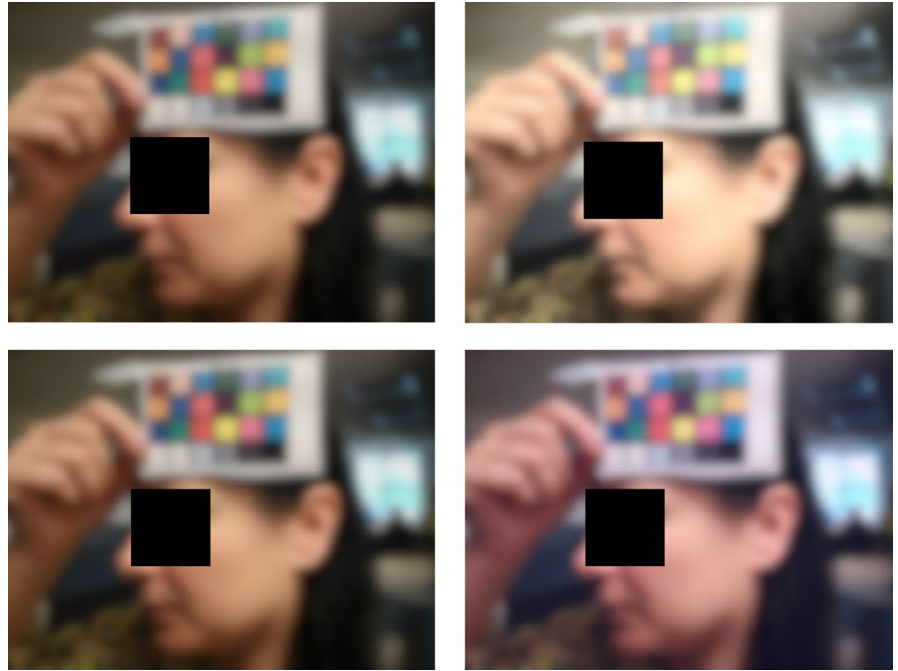
$$W_G = S_G / I_G \quad (3)$$

2.5 | Statistics

The following sets of statistics were generated:

- Antera 3D at baseline and after 4 weeks: Antera 3D parameters were analysed at baseline and following 4 weeks of treatment of the skincare treatment using paired t test, or Wilcoxon matched pairs test if normality of the data was not observed with Shapiro-Wilk. Level for significance was set at $\alpha = 0.05$.
- Discrepancies between Smartphone parameters and Antera 3D parameters: Differences between Antera 3D parameters and Smartphone parameters were evaluated by the mean of mean squared error (MSE).
- Diurnal variation of skin parameters with Smartphone parameters: Regular evaluation of the parameters from the smartphone (colour and wrinkles) was plotted for each volunteer in order to visualise their individual diurnal variation.

FIGURE 4 Normalising images showing the original, histogram equalised, CLAHE and colour card normalisation. Note that this and subsequent images are blurred for volunteer anonymity



3 | RESULTS

3.1 | Antera 3D at baseline and after 4 weeks

Before we examine the per-participant results, we will look at overall changes in colour and in wrinkles observed in results from the Antera 3D camera taken on the first and last days of the trial. Table 1 presents a summary. A significant reduction of wrinkle max depth and skin yellowness b^* was observed following 4 weeks of using a skincare routine (paired t test, $P < .05$). Other parameters were not found to change significantly.

LAB colour values for skin regions in an image are made up of an L value for describing the lighting, an A value for redness of the skin and a B value for describing the yellowness. We compared colour values from images taken with the controlled lighting Antera 3D camera on the first day of the trial with those taken on the last day on a per-volunteer basis. These show that between the beginning and the end of the 4-week trial period, from among all volunteers for the A and B colour values there is a slight reduction for almost all

volunteers except JU, AE and HE. For the L values, the overall trend is also slightly downward except for FC and to a lesser extent EZ, TD and JU who have greater declines, while volunteer HE has the most noticeable increase. Thus, there are lots of observed colour changes among volunteers but we do not know the nature of these changes during the trial.

3.2 | Discrepancies between smartphone parameters and Antera 3D parameters

We examined the colour values for images taken using the smartphone which had already been normalised using each of the three within-image normalisation methods described in section 2.3 and compared these against colour values from the Antera 3D camera. Those differences were measured using mean squared error (MSE), and Table 2 shows the MSE for each colour parameter and each colour normalisation technique with a range of differences for different normalisation techniques.

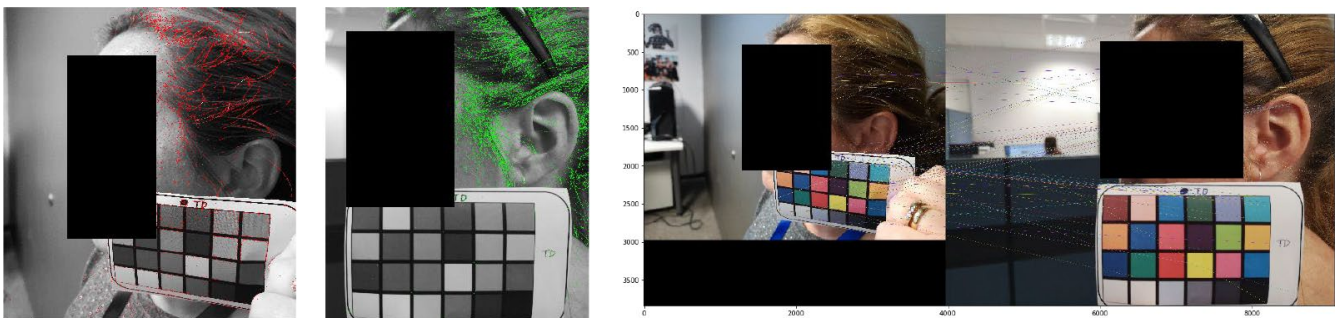


FIGURE 5 Key points and matching using SIFT showing training image, a query image and SIFT-kNN matching

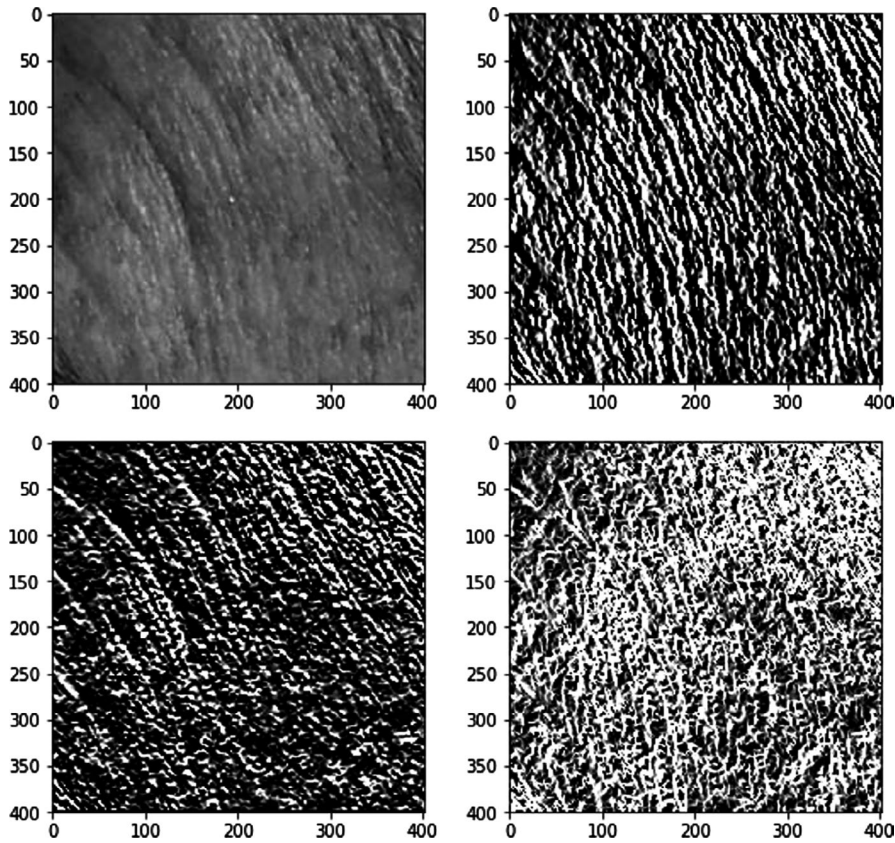


FIGURE 6 Gradient edge detection stages showing grayscale, Sobel X, Sobel Y and Sobel combined where, W_G is the Wrinkle ratio of the image G , S_G is the calculated mean of the Sobel Combined Image, and I_G is the calculated mean of the Original Image G

TABLE 1 Changes in colour and wrinkle assessment between start and end of trial using the Antera 3D camera

Parameters	Colour (L)	Colour (A)	Colour (B)	Wrinkle overall size	Wrinkle depth	Wrinkle max depth
% variation	0.8%	1.7%	4.0%	1.6%	2.0%	11.2%
Significant	No	No	YES	No	No	YES
P-value	0.067	0.303	0.010	0.664	0.669	0.048
Statistical test	paired t test	paired t test	paired t test	paired t test	paired t test	paired t test

We calculated correlation between the colour parameters observed by the Antera 3D camera at the start of the trial, and the colour parameters from the smartphone on day 1, normalised using three different approaches, and the results are shown in Table 3.

What this shows is that colour normalisation techniques have varying levels of effectiveness and that the B element, corresponding to skin yellowness, is particularly difficult to replicate with a smartphone. The A element corresponding to skin redness, which

TABLE 2 MSE of normalisation methods

Image type	MSE (L)	MSE (A)	MSE (B)
Original	19.25	42.05	22.51
CLAHE	61.75	45.29	21.74
CC normalised	130.89	48.13	54.92
Histogram	255.52	99.16	14.18

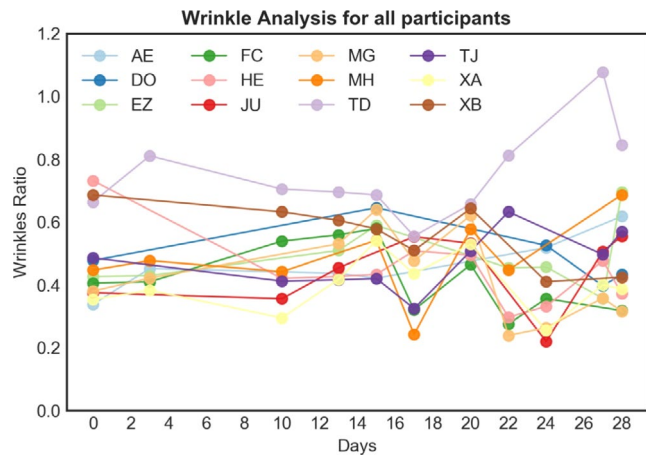
we are most interested in, correlates well with the Antera 3D when CLAHE normalisation is used with a correlation of 0.860.

3.3 | Diurnal variation of skin parameters with smartphone parameters

In the second of our experiments, we focused on wrinkle analysis. In Table 1, we saw there is an 11.2% variation in wrinkle maximum depth between start and end of the trial with other features like wrinkle size and depth not showing significant difference. The per-volunteer result shows that those volunteers with initially low values tended to stay low at the end of the trial (e.g. EZ, XA, HE and AE) while those with initially higher values tended to reduce wrinkles (e.g. FC, TJ and TD). This is in line with what would happen physiologically: people having a low severity of wrinkles would hardly see an improvement of their wrinkles, while people having deeper wrinkles

TABLE 3 Correlation between Antera 3D and Smartphone skin colours

Colour parameter	ColorNorm	Histequal	CLAHE	Original
L*	-0.159	0.344	0.628	0.719
A*	0.237	0.439	0.860	0.771
B*	-0.254	-0.423	-0.421	-0.404

**FIGURE 7** Wrinkle ratio results for all volunteers during the trial period

would be far more likely to see improvements, following the usage of cosmetic products.

Using the within-trial smartphone images, we identified the wrinkle area and computed its gradient magnitude. If the appearance of wrinkles in the image is greater, then more edges will have been detected by the gradient operator. In such a case, the value from the Sobel combined is large, resulting in a higher value for Wrinkle ratio. What this analysis shows is that almost all volunteers have some perturbations in their wrinkle analysis during the 4-week trial and there are no "flatlines" in the graph in Figure 7. Between days 18 and 22 (which was May 17th, a Friday) something happened to impact everybody's assessment. As with the analysis of skin colour, follow-up interviews with individual volunteers would be needed to explore the causes for these changes but once again we see that the analysis of smartphone images taken during the trial gives insights into within-trial changes of volunteers' skin.

4 | CONCLUSIONS

This research aimed to study the intermediate progress of cosmetic laboratory product trials using smartphone images so as to gather additional information during the trial period allowing better analysis. Working with a cohort of 12 female volunteers, we focused on skin colour which required colour normalisation on the images to adjust for varying daylight conditions, and automatic analysis of wrinkles based on measuring lines in the greyscale version of the images. We used images from an Antera 3D camera taken at start and end of

the trial and then images taken from a consumer-level smartphone taken throughout the trial period.

When we examine per-volunteer changes in both skin colour and wrinkle ratio we see that for many of the volunteers there are noticeable perturbations in both their skin colour and wrinkle ratio indicating non-linear transitions from the start to end points. This clearly highlights the contribution of other confounding variable that cannot be controlled/recorded accurately as part of regular clinical trials with fixed appointments: weather fluctuations, food intake, mood, sleep, exercise and so on. A face-to-face exit interview with each volunteers, or the volunteers keeping a log or a diary of events during the trial would allow those conducting the product trial to get a deeper understanding of skin changes not just at the two fixed timepoints of start and end, but throughout the trial.

The results obtained in this paper indicate there is potential for using within-trial facial image data gathered directly by participants using a consumer-level smartphone. While there are limits on the ability of such devices to capture detailed high-resolution, 3D facial surface textures such as described in,¹⁴ the correlation between normalised smartphone images and the Antera 3D camera for skin colour parameters are very encouraging.

Lifelogging represents a phenomenon whereby people can digitally record their own daily lives, ordinary lives, in varying amounts of detail, and for a variety of purposes.¹⁵ This can include using simple wearable devices like fitness trackers, step counters and sleep monitors, right up to front-facing cameras and wearables to record heart rate, respiration, stress, etc Including lifelog data into clinical trial analysis, even data gathered from simple wearable devices would allow real investigation of the diurnal patterns of the skin and link them to stress, sun exposure, air pollution, food consumption, cosmetic usage and many more features of daily life. This would help cosmetic research organisations to add an extra layer of information on compliance of volunteers as part of regular clinical trials Without the need of having volunteers coming on a daily basis to a clinic facility.

ACKNOWLEDGEMENTS

Funding for the work described in this paper was provided by Science Foundation Ireland under grant number SFI/12/RC/2289 (Insight Research Centre) and equipment, and laboratory expertise was provided by Oriflame R&D.

DATA AVAILABILITY

The data used in this research including images and analysis is archived in a public repository and is available at <https://doi.org/10.6084/m9.figshare.11881059.v1>

ORCID

Alan F. Smeaton  <https://orcid.org/0000-0003-1028-8389>

Cyril Messaraa  <https://orcid.org/0000-0002-4948-9949>

REFERENCES

1. Huawei. Huawei Camera Specifications – 2019. [Online]. <https://consumer.huawei.com/ie/phones/m/p20/camera>. Accessed: February 12, 2020.
2. Linming F, Wei H, Anqi L, et al. Comparison of two skin imaging analysis instruments: the VISIA® from canfield vs the ANTERA 3D® CS from Miravex. *Skin Res Technol*. 2017;24(1):3-8. <https://doi.org/10.1111/srt.12381>
3. Messaraa C, Metois A, Walsh M, et al. Wrinkle and roughness measurement by the Antera 3D and its application for evaluation of cosmetic products. *Skin Res Technol*. 2018;24:359-366.
4. Gardner R. Adobe Photoshop ACR Camera Calibration Basic Adjustments. [Online]. 2006. <http://www.rags-int-inc.com/PhotoTechStuff/AcrCalibration/>. Accessed: February 12, 2020.
5. Csink L, Paulus D, Ahlrichs U, Heigl B. Color normalization and object localization. *Mathematics*. 1998;49-55.
6. Finlayson G, Hordley S, Schaefer G, Tian GY. Illuminant and device invariant colour using histogram equalisation. *Pattern Recogn*. 2005;38(2):179-190.
7. Ali AR. Histogram Equalization in Python. [Online]. 2018. https://code.tutsplus.com/tutorials/histogram-equalization-in-python-cms-30202#post_comments. Accessed February 20, 2020.
8. Bradski G, Kaehler A. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.; 2008.
9. IITMCVG. Histogram Equalisation, Filters and Image Enhancement. [Online]. 2018. https://iitmcvg.github.io/summer_school/Session2/. Accessed February 16, 2020.
10. Karami E, Prasad S, Shehata MS. Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images. *CoRR*. 2017;abs/1710.02726.
11. Ji H, Wu YH, Sun HH, Wang Y. SIFT feature matching algorithm with global information. *Optics Precis Eng*. 2009;17(2):439-444.
12. Glover N. *Gradients and Edge Detection [Online]*, 2016. Available <https://handmap.github.io/gradients-and-edge-detection/>. Accessed February 12, 2020.
13. Mlsna PA, Rodriguez JJ. Gradient and Laplacian edge detection. In: *The Essential Guide to Image Processing*, Elsevier; 2009: 495-524.
14. Seck A, Hannah D, William S, Tiddeman B. 3D surface texture analysis of high-resolution normal fields for facial skin condition assessment. *Skin Res Technol*. 2020;26(2):169-186.
15. Gurrin C, Smeaton AF, Doherty AR. Lifelogging: personal big data. *Found Trends®. Inf Retrieval*. 2014;8(1):1-125.

How to cite this article: Smeaton AF, Srungavarapu S, Messaraa C, Tansey C. Tracking skin colour and wrinkle changes during cosmetic product trials using smartphone images. *Skin Res Technol*. 2020;00:1–8. <https://doi.org/10.1111/srt.12928>