

Microgeometry capture and RGB albedo estimation by photometric stereo without demosaicing

Yvain Quéau¹, Matthieu Pizenberg², Jean-Denis Durou² and Daniel Cremers¹

¹ Technical University Munich, Garching, Germany

² Université de Toulouse, IRIT, UMR CNRS 5505, Toulouse, France

ABSTRACT

We present a photometric stereo-based system for retrieving the RGB albedo and the fine-scale details of an opaque surface. In order to limit specularities, the system uses a controllable diffuse illumination, which is calibrated using a dedicated procedure. In addition, we rather handle RAW, non-demosaiced RGB images, which both avoids uncontrolled operations on the sensor data and simplifies the estimation of the albedo in each color channel and of the normals. We finally show on real-world examples the potential of photometric stereo for the 3D-reconstruction of very thin structures from a wide variety of surfaces.

Keywords: 3D-reconstruction, photometric stereo, non-uniform illumination, RGB images, Bayer filter.

1. INTRODUCTION

Among the numerous computer vision techniques for achieving 3D-reconstruction from digital cameras, photometric techniques such as shape-from-shading¹ and photometric stereo² are often considered as first choices when it comes to the recovery of thin structures. Indeed, they are able to estimate the surface normals in each pixel (in the literature, “pixel” usually refers to a cell in the red, green and blue channels interpolated from the Bayer matrix: we rather consider a pixel as a cell in the non-demosaiced RAW image). In this work, we focus on surfaces pictured by a device consisting of a digital camera and several LEDs arranged as described in Figure 1.

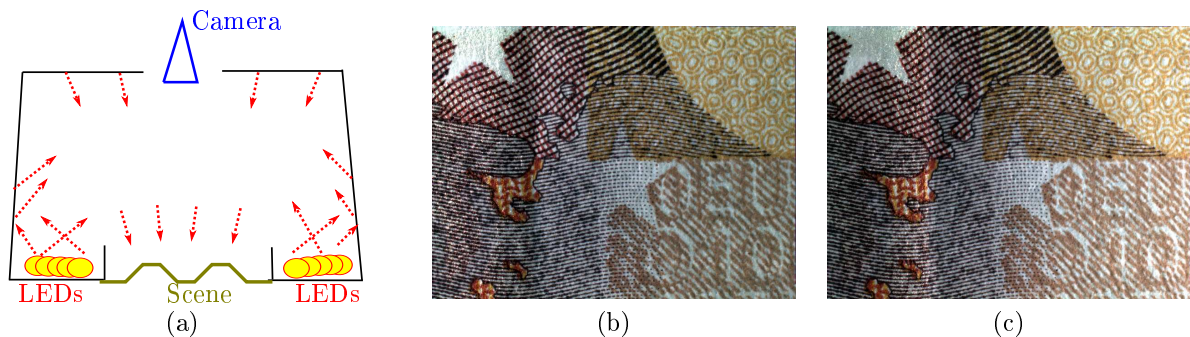


Figure 1. (a) Schematic representation of the device used for 3D-reconstruction by photometric stereo. Controllable LEDs are oriented towards the walls of the device, in order to illuminate the scene in a diffuse manner.³ A digital camera is used to capture $m = 15$ 12-bits RAW images of the scene under varying illumination obtained by successively turning on the different LEDs. (b-c) Two RGB images of a folded 10 euros banknote obtained with this device, which is intended for small-scale 3D-reconstruction: in full resolution (3664×2748 pixels), the imaged area is of size $1.6 \text{ cm} \times 1.2 \text{ cm}$, thus the surface area corresponding to a pixel is around $5 \mu\text{m} \times 5 \mu\text{m}$.

Correspondence:

Y. Quéau: yvain.queau@tum.de

M. Pizenberg: matthieu.pizenberg@enseeiht.fr

J.-D. Durou: durou@irit.fr

D. Cremers: cremers@tum.de

Photometric techniques invert a photometric model describing the interactions between the illumination and the surface. The usual assumptions of photometric stereo are that the data consist in $m \geq 3$ gray level images I^i , $i \in [1, m]$, obtained from a still camera, but under varying directional illumination (cf. Figure 1), that the surface is opaque and that its reflectance is Lambertian (perfectly diffusive). Under these assumptions, and neglecting shadowing effects, the gray level at pixel (u, v) in the i -th image is modeled as:

$$I_{u,v}^i = \rho_{u,v} \mathbf{n}_{u,v} \cdot \mathbf{s}^i, \quad i \in [1, m] \quad (1)$$

where $\rho_{u,v} > 0$ is the albedo, $\mathbf{n}_{u,v}$ is the unit outward normal to the surface, vector \mathbf{s}^i points towards the light source, its intensity is proportional to the luminous flux density, and \cdot is the Euclidean scalar product.

The albedo $\rho_{u,v}$ and the normal $\mathbf{n}_{u,v}$ can be estimated in each pixel (u, v) by solving system (1) in the least-squares sense in terms of the vector $\mathbf{m}_{u,v} = \rho_{u,v} \mathbf{n}_{u,v}$. Then, the albedo is obtained by $\rho_{u,v} = \|\mathbf{m}_{u,v}\|$ and the normal by $\mathbf{n}_{u,v} = \mathbf{m}_{u,v} / \|\mathbf{m}_{u,v}\|$. Eventually, the depth map is obtained by integration of the estimated normals.⁴ Figure 2 shows the albedo, normals and depth estimated from $m = 15$ images such as those in Figure 1.

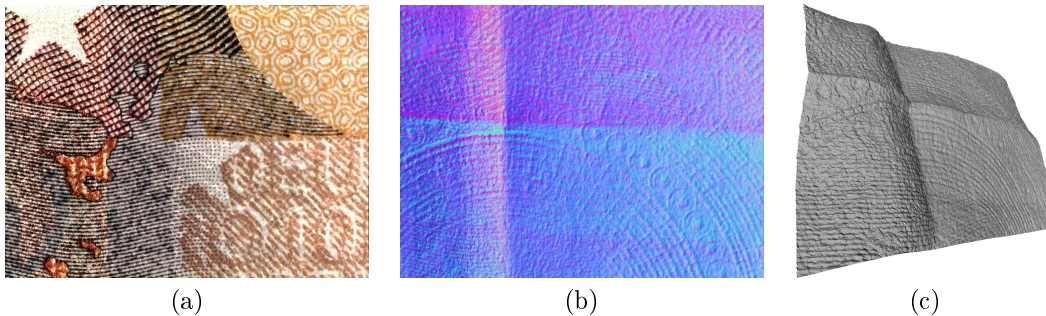


Figure 2. From a set of $m \geq 3$ RGB images such as those shown in Figure 1, obtained under varying illumination while keeping the camera still, our photometric stereo approach simultaneously recovers (a) an RGB albedo, (b) a normal map showing microgeometric structures which are hardly visible in the images, and (c) a 3D-reconstruction of the surface.

This procedure, which is the most considered in the literature on photometric stereo, relies on two restrictive assumptions. First, it is assumed that the data consist in gray level images. Yet, most modern digital cameras provide RGB images, which need to be converted to gray levels for the need of photometric stereo, inducing a loss of information. In addition, each illumination vector \mathbf{s}^i is supposed to be the same in all pixels. Since in our case, the effective incident illumination is actually the result of multiple reflections with the surrounding environment (cf. Figure 1-a), this is hardly justified. The main purpose of our contribution is to show that the gray level assumption can be avoided very simply by considering as inputs the RAW images from the sensor, without demosaicing (cf. section 3). This simplifies the estimation of the (color) albedo and of the normals, in comparison with methods dealing with interpolated RGB images (cf. section 2). Yet, our approach requires that each illumination is calibrated with respect to each pixel (cf. section 4).

2. RELATED WORK

Microgeometry recovery by photometric stereo has already been achieved by applying a chemical gel between the surface and the camera, in order to “lambertianize” the surfaces.⁵ In contrast, we use a non-intrusive method relying only on standard equipment such as LEDs, yet arranged in such a way that they create diffuse illumination, thus limiting specularities. In addition, when the appearance of a surface is modified by a chemical process, its original albedo cannot be estimated anymore, while our device is able to estimate it.

The gray level assumption has been relaxed in several ways in the literature on photometric stereo. One famous example is the real-time 3D-reconstruction of deformable white surfaces observed under different colored light sources.⁶ The standard photometric stereo procedure is used to estimate the normals (the estimated albedo is not relevant, since the surface is supposed to be uniformly white), considering the red, green and blue channels as three gray level images. The main advantage of this approach is that it can be applied from a single RGB image, an idea which was already suggested in the early work by Woodham.²

Another important work is that by Barsky and Petrou,⁷ who used four RGB images to estimate the three albedo values related to each channel and the normals. To this purpose, the photometric model (1) is written w.r.t. each color channel, yet considering that the illumination is “white”:

$$I_{u,v}^{i,\star} = \rho_{u,v}^{\star} \mathbf{n}_{u,v} \cdot \mathbf{s}^i, \quad i \in [1, m], \star \in \{R, G, B\} \quad (2)$$

where the albedo $\rho_{u,v}^{\star}$ is now relative to the color channel $\star \in \{R, G, B\}$. The standard photometric stereo procedure could be applied independently in each color channel, which would provide the desired values of the albedo, as well as three estimates for the normal. Unfortunately, this would lead to incompatible estimates of the normal: Barsky and Petrou proposed a principal component analysis-based procedure to overcome this drawback by simultaneously estimating the three values of the albedo and the normal vector. Ikeda suggested an alternative procedure consisting in estimating the color albedo first, and then the shape by resorting to a nonlinear PDE framework.⁸ It was also recently shown that considering ratios between pairs of color levels in the same channel yields a system of linear PDEs in the depth, which can be solved independently from the albedo.⁹

Nevertheless, all these works assume that the triplet of RGB values correspond to the same surface point. Yet, this is not a valid assumption with standard RGB cameras. Indeed, the color filters are usually arranged according to a Bayer pattern: one cell of the sensor can only receive information in one specific color channel. To obtain a triplet of RGB values, interpolation is required. This induces a bias, as each normal is estimated from color levels registered in neighboring pixels rather than only in the current one.

As for the directional illumination assumption, it has frequently been questioned in the context of photometric stereo. A lot of luminous sources, including pointwise and extended ones, can be approximated by a parametric model, see for instance¹⁰ for some discussion. The pointwise source model is often encountered in real-world applications, as LEDs represent cheap light sources which fit rather well this model. Calibrating the parameters of such sources (e.g., position, orientation and intensity) is a relatively well-known problem,¹¹ as well as the numerical resolution of photometric stereo under such an illumination model.⁹

Unfortunately, parametric models are not adapted to our use case illustrated in Figure 1: although the light reflections inside the device could be modeled using rendering techniques, inverting the rendering equation would be impossible in practice. Instead, we prefer to resort to a simple approximate model describing the resulting luminous flux reaching the surface. A first possibility consists in sampling the intensity of each illumination in a plane located in the area of interest, and then dividing each new image by these intensity maps.¹² Yet, this technique can only compensate for non-uniform illumination intensities, but not for non-uniform illumination directions. Another possibility would be to decompose the illumination direction on the spherical harmonics basis,¹³ and to calibrate the first coefficients of this decomposition. On the other hand, it is difficult to predict the number of harmonics required to ensure that the model is reasonably accurate. We will show in section 4 how to sample both the illumination intensities and directions, in order to obtain an accurate representation of the luminous flux reaching the surface.

3. RGB PHOTOMETRIC STEREO WITHOUT DEMOSAICING

Let us now introduce our RGB photometric stereo model, which relies on RAW inputs without demosaicing. As we shall see, avoiding demosaicing yields a much simpler approach to estimate the RGB albedo and the normals, in comparison with existing works discussed above.

In order for photometric model (1) to be satisfied when using real-world images, the camera response should be as linear as possible. In this view, all uncontrolled operations on the images should be avoided. This means that hardware automatic corrections such as exposure, white balance and gamma corrections should be disabled. Conversion from RAW data to JPEG images should also be avoided, since RAW images usually have a better depth (our RAW images are coded on 12 bits). In addition, we argue that demosaicing the RAW data should not be achieved, since this results in hallucinating missing data by interpolating the actual measurements registered by the sensor. Although elaborate demosaicing methods do exist, we believe that it is more justified to consider the values from the sensor as they are, without any kind of modification which might break the reliability of the response.

Hence, instead of considering that in each pixel (u, v) , a triplet of RGB values is available, we rather consider that a single measurement is available, yet this measurement should be understood as relative to one specific color channel, depending on the arrangement of the Bayer matrix, see Figure 3. We also add a space-dependency to the illumination vectors, since the intensity of each illumination vector \mathbf{s}^i is relative to the color channel, and since its direction varies because the illumination is diffuse (cf. Figure 1). This eventually leads to the following photometric model:

$$I_{u,v}^i = \rho_{u,v} \mathbf{n}_{u,v} \cdot \mathbf{s}_{u,v}^i, \quad i \in [1, m] \quad (3)$$

Note that in this new color PS model, there is no incompatibility in the estimated normals: if the illumination vector fields $\mathbf{s}_{u,v}^i$ are known (see section 4), we can apply the standard photometric stereo procedure to recover the albedo and the normal in each pixel. This yields a much simpler procedure, as compared with existing algorithms which require unmixing color and shape⁷ or resorting to image ratios.⁹

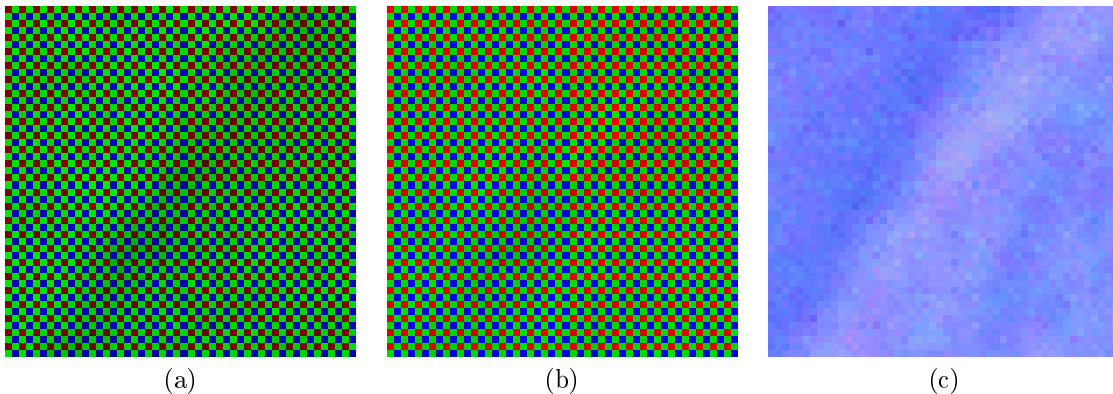


Figure 3. Photometric stereo without demosaicing. (a) Close-up on a 50×50 part of one of our input RAW images (represented in RGB for better visualisation). We suggest to use such non-demosaiced inputs as data for estimating (b) a Bayer-like estimate of the albedo (which can be further interpolated to obtain an RGB albedo) and (c) an estimate of the normal in each pixel. By avoiding demosaicing, we avoid any uncontrolled transformation of the data from the sensor, allowing the recovery of microgeometry structures, see section 5.

Eventually, if an RGB representation of the albedo is required, one should apply a demosaicing algorithm to the Bayer-like estimate of ρ . Since we are mostly interested in this work in recovering the surface shape, we apply a simple linear interpolation for this purpose. We emphasize that no demosaicing of the normal map is required: its direct estimation from the Bayer matrix already provides a dense map without missing data.

As a final stage in our pipeline, perspective integration of the normals into a depth map is performed. To this purpose, we apply the DCT method of Simchony et al.¹⁴ to the perspective gradients estimated from the normal field as described in.⁴ This yields an up-to-scale depth map, and we eventually deduce the scale from the mean camera-to-surface distance, which is estimated while geometrically calibrating the camera.

4. SAMPLING THE ILLUMINATION DIRECTIONS AND INTENSITIES

We now describe a practical way to sample the illumination directions and intensities i.e., the m vectors $\mathbf{s}_{u,v}^i$ appearing in (3), in each pixel (u, v) .

To this purpose, it would be necessary to invert the model (3) in terms of the $\mathbf{s}_{u,v}^i$ vector, in each pixel (u, v) and for each illumination i . This can be achieved independently for each illumination, by using a calibration object with known albedo $\rho_{u,v}$ and known normals $\mathbf{n}_{u,v}$. Unfortunately, since only one normal is available in each pixel, solving (3) is an under-constrained problem. Ensuring a correct estimation of each illumination vector $\mathbf{s}_{u,v}^i$ would require to use a series of Lambertian calibration objects whose shape and color are known, to picture each calibration object under each illumination, and eventually to invert the Lambertian model. Yet, this procedure would be very time-consuming.

Instead, we designed a simple calibration method which requires a single calibration object, consisting in an array of hexagonal structures machined in a diffuse white material. Without loss of generality, we assume that this “white” color has albedo equal to 1 w.r.t. all three color channels, that is to say $\rho_{u,v} \equiv \rho = 1$. This means that any color will then be estimated with this “white color” as reference.

Then, we divide the 2D grid into 30 rectangular parts Ω^j , $j \in [1, 30]$. In each of these rectangular parts, up to seven different normals (the fronto-parallel hexagonal part and six sloped faces), along with the corresponding color image values, are available. We assume that the rectangular parts are small enough so that the illumination can be locally considered as directional and uniform in each color channel. For each channel $\star \in \{R, G, B\}$, each illumination $i \in [1, m]$, and each rectangular part Ω^j , we approximate the illumination $\mathbf{s}_{u_0^j, v_0^j}^{i, \star}$ in the center pixel (u_0^j, v_0^j) of Ω^j by solving in the least-squares sense the following system of linear equations:

$$\mathbf{n}_{u,v} \cdot \mathbf{s}_{u_0^j, v_0^j}^{i, \star} = I_{u,v}^i, \quad (u, v) \in \Omega^{j, \star} \quad (4)$$

where $\Omega^{j, \star}$ is the set of pixels in Ω^j for which an information in channel \star is available.

This gives us a sparse estimation of each illumination in each color channel. Each of these scattered data is further interpolated and extrapolated to the whole grid by using a biharmonic spline model, resulting in three \mathcal{C}^2 -smooth vector fields per illumination. These three fields $\mathbf{s}_{u,v}^{i, \star}$, $(u, v) \in \Omega$, are eventually combined into a single one $\mathbf{s}_{u,v}^i$, by using the same Bayer arrangement as in the images. By repeating this procedure for each illumination, we obtain the vectors $\mathbf{s}_{u,v}^i$ which arise in model (3), and the 3D-reconstruction procedure described in section 3 can be applied.

Let us note for completeness that a similar idea was proposed by Johnson et al.,⁵ in order to calibrate a spherical harmonics model, and that our approach can be viewed as an extension of flatfielding techniques¹² aiming at sampling not only the illumination intensities, but also their directions.

5. EMPIRICAL EVALUATION

Let us now show on real-world examples that our device is capable of recovering very thin structures for a wide variety of surfaces.

We first show in Figure 2-c the 3D-reconstruction of the 10 euros banknote of Figure 1. This experiment shows us the potential of photometric stereo for surface inspection applications, as for instance verifying the presence of thin structures which should be present in real banknotes.

Then, we show in Figure 4 the estimated albedo maps and 3D-reconstructions of two metallic euro coins. Obviously, metallic surfaces do not hold the Lambertian assumption upon which our approach relies. Nevertheless, the estimated albedo maps remain satisfactory (although albedo is clearly over-estimated around sharp structures, because of strong inter-reflection effects), while the shaded depth maps nicely reveal small impacts on the surface of the coins which may be due to usury (e.g. between the “n” and the “t” of the Cervantes portrait) or may be part of the original engraving (e.g. on the left cheek of this character).

Eventually, to quantitatively assess the overall accuracy of the 3D-reconstruction, we show in Figure 5 the results obtained with a machined surface. The 3D-model which has been sent to the machine being known, we can match our 3D-reconstruction with it and evaluate the absolute cloud-to-mesh (C2M) distance between both surfaces. The results show that 99% of the estimated points have a 3D-reconstruction error below 0.1 mm, and that the median value of the error is around 20 μm . A closer look at the spatial distribution of the errors shows that the high errors are localized around sharp corners: this is probably due to inter-reflection effects and to the fact that we used least-squares integration of the normals,¹⁴ which tends to smooth the 3D-reconstruction,⁴ but on the other hand it may also be due to the inaccuracy of the machine itself. Hence, our overall error is probably even lower than that measured.



Figure 4. 3D-reconstructions of two metallic coins. From left to right: one of the $m = 15$ input images, estimated albedo and 3D-reconstruction. First row: Italian 1 euro coin. Second row: Spanish 50 cents coin.

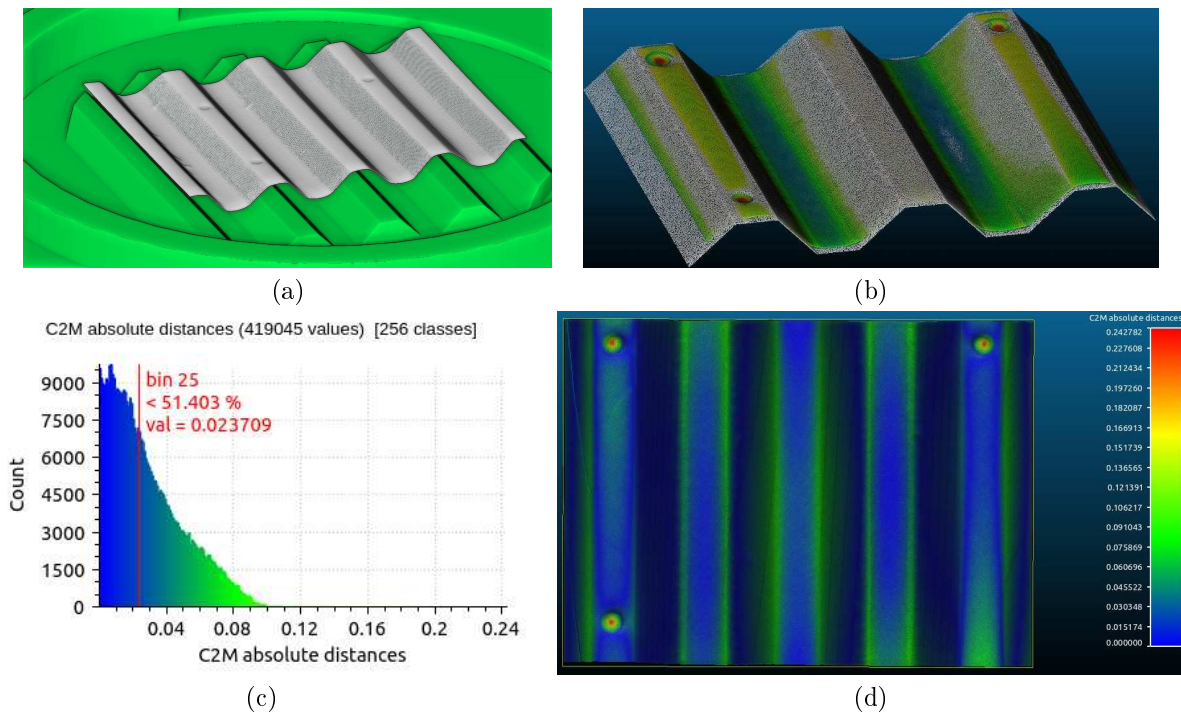


Figure 5. Quantitative evaluation of the 3D-reconstruction using a machined surface. (a) Our 3D-reconstruction is matched with the 3D-model. (b) Close-up on the matched shapes. (c) Histogram of the absolute C2M distance between both surfaces (in mm): the median value is around $20 \mu m$. (d) Spatial distribution of the 3D-reconstruction errors.

6. CONCLUSION AND PERSPECTIVES

We have shown the potential of photometric stereo for microgeometry capture, while relying only on standard equipment such as a digital camera and LEDs. Unlike previous work, we do not need to resort to any chemical process in order to enforce the Lambertian behavior of the surface. This is made possible by a well-engineered device which illuminates the scene with controllable diffuse light, and by directly modeling the photometric stereo problem from the RAW, non-demosaiced images. This new model simplifies the estimation of the RGB albedo in comparison with other color photometric stereo models, provided that a dense estimation of the incident luminous flux is available. In this view, we also described a calibration procedure for sampling the illumination intensities and directions on the acquisition plane, while previous methods only sample the intensities.

Nevertheless, our model remains valid only for shapes with limited slopes. Indeed, with steepest surfaces, inter-reflections, shadows or penumbra will occur. As future work, we plan to improve the robustness of our method w.r.t. such effects. A first strategy would consist in improving our simple regression procedure based on least-squares by using more robust estimators. Another option would be to iteratively refine the illumination estimation, by alternating it with the estimation of the shape.

ACKNOWLEDGMENTS

This research was funded by the Toulouse Tech Transfer company (Toulouse, France).

REFERENCES

- [1] Horn, B. K. P., *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*, PhD Thesis, MIT, Cambridge, USA (1970).
- [2] Woodham, R. J., “Photometric method for determining surface orientation from multiple images,” *Optical Engineering* **19**(1), 134–144 (1980).
- [3] George, J. and Delalleau, A., “Visual observation device, especially for a dermatological application,” (2016). EP Patent App. EP20,140,800,095.
- [4] Durou, J.-D., Aujol, J.-F., and Courteille, F., “Integrating the Normal Field of a Surface in the Presence of Discontinuities,” in [*Energy Minimization Methods in Computer Vision and Pattern Recognition (EMM-CVPR)*], (2009).
- [5] Johnson, M. K., Cole, F., Raj, A., and Adelson, E. H., “Microgeometry capture using an elastomeric sensor,” *ACM Transactions on Graphics* **30**(4), 1 (2011).
- [6] Hernández, C., Vogiatzis, G., Brostow, G. J., Stenger, B., and Cipolla, R., “Non-rigid Photometric Stereo with Colored Lights,” in [*IEEE International Conference on Computer Vision (ICCV)*], (2007).
- [7] Barsky, S. and Petrou, M., “The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(10), 1239–1252 (2003).
- [8] Ikeda, O. and Duan, Y., “Color Photometric Stereo for Albedo and Shape Reconstruction,” in [*IEEE Workshop on Applications of Computer Vision (WACV)*], (2008).
- [9] Quéau, Y., Mecca, R., and Durou, J.-D., “Unbiased photometric stereo for colored surfaces: A variational approach,” in [*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], (2016).
- [10] Quéau, Y. and Durou, J.-D., “Some Illumination Models for Industrial Applications of Photometric Stereo,” in [*Quality Control by Artificial Vision (QCAV)*], (2015).
- [11] Xie, L., Song, Z., Jiao, G., Huang, X., and Jia, K., “A practical means for calibrating an LED-based photometric stereo system,” *Optics and Lasers in Engineering* **64**, 42–50 (2015).
- [12] Sun, J., Smith, M., Smith, L., and Farooq, A., “Sampling Light Field for Photometric Stereo,” *International Journal of Computer Theory and Engineering* **5**(1), 14–18 (2013).
- [13] Basri, R., Jacobs, D. W., and Kemelmacher, I., “Photometric Stereo with General, Unknown Lighting,” *International Journal of Computer Vision* **72**(3), 239–257 (2007).
- [14] Simchony, T., Chellappa, R., and Shao, M., “Direct analytical methods for solving Poisson equations in computer vision problems,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(5), 435–446 (1990).