https://doi.org/10.2352/issn.2168-3204.2023.20.1.3

Can Surface Topography give us Best Light Positions for Reflectance Transformation Imaging?

Muhammad Arsalan KHAWAJA [1&2], Sony GEORGE [2], Franck MARZANI [1], Jon Yngve HARDEBERG [2], Alamin MANSOURI [1]

[1] *ImViA Laboratory, Université de Bourgogne, Dijon, France.*
[2]Department of Computer Science, Norwegian University of Science and Technology (NTNU), Gjøvik, Norway.

Abstract

Reflectance Transformation Imaging (RTI) is a technique that provides an enhanced visualization experience. The current acquisition methods for Reflectance Transformation Imaging (RTI) are time-consuming and computationally expensive. This work investigates the idea of getting the best light positions for RTI acquisition using surface topography. We propose automating the RTI acquisition by estimating the surface topography using a deep learning method followed by estimating light positions using unsupervised clustering method. This is a one-shot method which only needs one image. We also created RTI Synthetic dataset in order to carry out experiments. We found that surface topography alone is not sufficient to estimate best light positions for RTI without putting constraints.

Introduction

Reflectance Transformation Imaging (RTI) is an imaging technique used to capture and visualize surfaces in varying light conditions. RTI has three major components. A camera, an object, and a light source. The setup for the RTI as shown in Figure 1 has the following features:

- The camera is fixed at the top of the surface.
- Each image is captured with a different light direction. The light can be anywhere in the imaginary hemisphere/ dome above the object.
- During the acquisition of the images, the camera and object remain fixed and only the light source is moving from one direction to another.
- The distance between the light source and the object remains constant throughout the acquisitions.

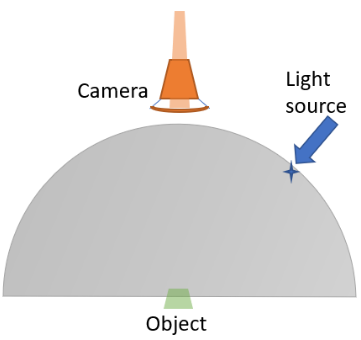


Figure 1. *RTI Setup*

The data acquired after the acquisition is called RTI data but it is also known by other names like Multi-Light Image Collection (MLIC), Single Camera Multi-Light (SCML), and Multi-Light Reflectance (MLR) [1]. The next step in the RTI pipeline is to process the data with relightening algorithms. The final step is to extract the features and use these feature maps to investigate and study surfaces. The pipeline for RTI is shown in Figure 2.

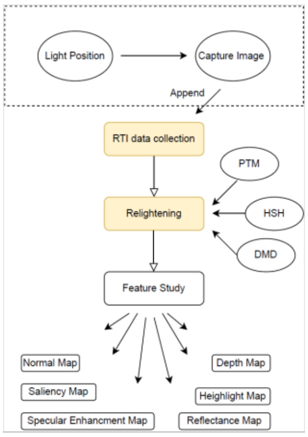


Figure 2. *RTI Pipeline*

The first practical implementation of Reflectance Transformation Imaging was carried out by Bell Labs [2]. After collecting the multi-light image collection (MLIC) dataset, Malzbender developed a re-lightening algorithm called Polynomial Texture Mapping (PTM). It interactively displays photorealistic renderings created for light positions that were not captured in the real dataset. The other methods that were developed later for re-lightening are Hemispherical Harmonics(HSH) [3] and Discrete Modal Decomposition (DMD) [4]. DMD is a superior algorithm relative to PTM and HSH [5].

RTI has wide applications ranging from archaeological investigation of ancient objects to quality inspection in Industry. It is a powerful technique that has got attention from industry and academia.

The documentation and digital preservation of cultural heritage objects have profound importance for historians as well as governments and international organizations. It is a great source of tourism as well as a tool for generations to connect and appreciate their culture and heritage. RTI can achieve visual recording of the artifacts which makes it a paramount technique in archeology[6]. RTI is also used to study the degradation in historical paintings and cultural heritage over a period of time [7]–[9]. It also provides important information to researchers studying the pigments/material used and helps in developing techniques used to preserve cultural heritage objects.

Recently, RTI has found a lot of applications in the industry. The prime use of RTI is in quality control [10]. RTI helps quality control inspectors to find defects that otherwise are challenging to find [11].

There are many acquisition setups used for RTI. The oldest one is Highlight RTI [12]. Human hands are used to hold and move the light source to acquire images. Since manual acquisition is a very tedious and painstaking task. Some domes have been developed to automate the acquisition process of RTI [12]. Recently, robotics arms have also been used to make RTI acquisition more efficient and fast in order to get better results [13].

The choice of light positions in RTI acquisitions can be classified in two categories. The first one is where the number and positions of the light source are pre-defined (generally equally spaced) The second one, generally automated, adapts the number and distribution of light position to the surface characteristics. So far, the field of automated acquisition has not been fully explored. We aim to investigate the problem of automated acquisition using surface topography information.

One of the major challenges in RTI is the processing of an enormous amount of data. The computation power is limited which serves as a bottleneck in RTI pipeline. The pipeline for RTI is shown in Figure 2. It requires extensive resources (memory, CPU, data storage), which limits the utilization of such a technique. This limitation serves as a bottleneck in RTI pipeline. Each dataset for RTI (Multi-Light Image Collection) can contain several images (ranging from 50 to 1000) and cameras mostly used for RTI tend to be highly defined producing high-resolution, large-size images. It is also observed that all images (each image corresponding to a unique light position) in the RTI dataset do not contribute equally to the re-lightening algorithm. Some images are more important than others. It is also observed that some regions in the hemisphere need more dense acquisition than others because reflectance is changing drastically in these regions [14].

The appearance of an acquired image depends on two factors. The first is the reflectance property of the object's material. Each material has unique reflectance properties like translucency, gloss, color etc. The second one is the surface topography of the object. The orientation of the surface, the roughness, bumps, and valleys highly influence the reflectance [15]. The function estimating the reflection of a light from surface is known as Bi-directional Reflectance Distribution (BRDF) function [16].

We are investigating an automated method to acquire an RTI dataset that can be more efficient. This area is not fully explored and the research question we ponder on is *"Is there a better way to acquire an RTI dataset with relatively less computational cost, better results, and smaller in size?"*. We have developed a method which takes one image as an input and estimates the light positions for the MLIC acquisition. The method is based on Surface topography. We have performed experimentation to test our method.

The nearest work related to our work is Next Best Light Position (NBLP) [14]. In this work, the author first acquires a small sparse dataset and then investigates how the reflectance changes with the light positions. It is possible to identify where there are abrupt reflectance changes, and he refers to those light positions as critical light positions. In the end, his algorithm identifies the interpolated light position using gradient descent and gives Next Best Light Position (NBLP) to acquire the image. This is an iterative method and adaptive method however it has the following disadvantages:

- It needs multiple images to initiate the algorithm.
- The algorithm is time taking since every time it has to calculate the differences in images and as the amount of images increases iteratively the complexity and computations are required to increase significantly.
- The algorithm is not conscious of the surface topography and only considers the reflectance properties.

In comparison, our algorithm only needs one image to initiate and estimates the light directions to acquire the RTI dataset.

In the next section, we explain our methodology, followed by dataset and experimentation. Finally, we present our results and discussion and conclude the research.

Methodology

We created an algorithm that gives us light positions for the acquisition of the dataset using surface topography information. The methodology of our algorithm is demonstrated in Figure 3.

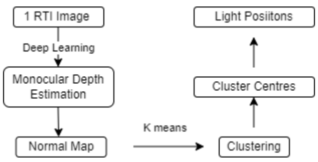


Figure 3. *The flowchart of our algorithm to estimate light positions.*

The first step is to acquire a single image of the object. In order to understand and estimate the topography of the surface we choose a recent deep-learning method [17]–[19]. This deep learning method is called Leres. The LeRes deep learning model is a convolutional neural network (CNN) architecture that is designed for estimating monocular depth from a single RGB image. Monocular depth estimation is a task of predicting depth of each pixel(image) from the camera. The LeRes deep learning network uses an encoder decoder with skip connections to preserve spatial information. The encoder is responsible for extracting features of the image and the decoder is responsible for estimating the depth map from the extracted features. It is state of the art monocular depth estimation method and that is why chose it. We feed the image to a deep learning network to estimate the depth map of the object. The depth map gives us information about surface topography.

This depth map is converted to a normal map by taking the gradient of the depth map with respect to the neighboring pixels. Once, we have the normal map of an image. Our goal is to find the critical light position for each sub-image. For that matter, the question is what is a critical light position? We define the critical light position as a position that reveals the most possible information about the surface. If it's a surface or a painting, the critical light position should be able to reveal the minor textures and features of the image. If it's Reflectance Transformation Imaging for quality inspection, the critical light position should be able to reveal minor surface imperfections. This definition of critical light position agrees with the prime purpose of Reflectance Transformation Imaging i.e. "To enhance the visualization experience". The critical light positions can be found by clustering the normal map image.

We choose k-means method for clustering. K-means is one of the most popular algorithms used by the machine learning community. It categorizes in unsupervised learning [20]. It partitions the data into clusters based on their similarities. The 'K' in K-means refers to the number of clusters. K-means algorithm first selects the k initial cluster centroids randomly. It then assigns each data point to the nearest centroid iteratively. It continues to update the centroid of each cluster in every iteration by taking the mean of the data points in that cluster. The algorithm finally converges where no clusters further change and the centroid stops moving.

After the clustering is done, the cluster centers are retrieved, and they correspond to where most of the pixels in the normal map image are pointing towards. This vector intersects with the imaginary hemisphere (RTI dome) around the image. That intersection point is recommended light position to acquire the next RTI image from.

In order to test the algorithm, we used the synthetic dataset. The software used for the experimentation part are Blender, MSA Tool and Anaconda.

Result and Discussion

This section explains the experimentation and discussion of the results of our methodology. In order to test our methodology, we created two synthetic datasets. The following subsection explains the dataset and experimentation.

Dataset and Experimentation

We created a synthetic dataset to run the experiments. The 3D models were obtained from Sketchfab (a popular platform for 3D content sharing) under Creative Common Licence [21]. The 3D model was then used in blender software to create a ground truth RTI dataset. The toolbox used in the blender is SFF-RTI.

For each object, these acquisitions were made to create a ground truth Multi Light Image Collection (MLIC) dataset. Figure 4 shows some of the images of these ground truth dataset.

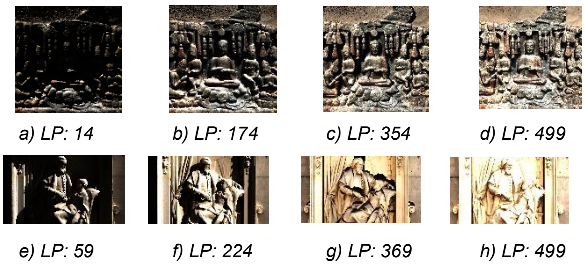


Figure 4. *Four random images from ground truth MLIC. The upper one is the Sculpture dataset and the lower one is the Buddha dataset. LP stands for Light Position.*

The ground truth contains 500 homogeneously equally spaced acquisitions. Figure 6f shows the light positions in space that were used to acquire the images. We chose 500 no of images because the reconstruction error in the normal map reduces significantly as the number of images (MLIC) increases. Figure 5 shows the structural similarity index of a normal map constructed with the ground truth normal map. We see that the normal map error reduces significantly after certain no of images. The normal map reconstruction converges and adding a greater number of images to MLIC does not improve the reconstruction. However, to be safer, and work with different datasets, we chose 500 image-dense acquisitions as a ground truth.

Figure 5. *The normal map reconstruction converges after certain no of images in MLIC. This is performed on sculpture dataset.*

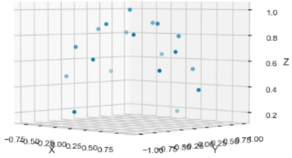
In order to test our algorithm, we designed experiments. The experimental protocol is explained as follows:

1. Take an acquisition from 45 to 60 degrees of elevation.
2. Estimate the light positions from a single image as described in the methodology.
3. Acquire the images from those estimated light positions.
4. Acquire the same no of images with homogeneous equally spaced light positions.
5. Estimate the normal map of both datasets (homogeneous equally spaced dataset and dataset created by our algorithm's recommended light directions)
6. Calculate PSNR and SSIM of the reconstructed normal maps with ground truth normal map (created with Relightening algorithm used for dense 500 acquisitions)
7. Compare our method with homogeneous equally spaced acquisition method.

Result and Discussion

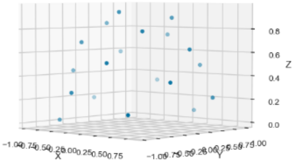
After conducting the experiments as explained in the previous section, on the Buddha and sculpture datasets. The estimated light positions for acquiring the images are obtained. The images are acquired using these estimated light positions for creating Multi Light Image Collection (MLIC) for RTI. The light positions obtained by our algorithm are compared with the uniformly spaced light positions by observing the normal maps reconstructed using respective MLIC's. One such experiment is demonstrated in Figure 6. The normal maps were reconstructed using DMD algorithm and data used for them was the respective MLIC's.

a) Normal Map



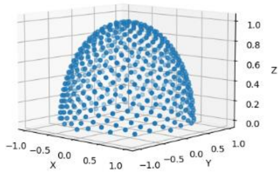
b) 19 Light positions estimated by our algorithm.

c) Normal Map



d) 19 Equally spaced Light positions

e) Ground Truth Normal map



f) 500 equally spaced light positions

Figure 6. *The right side sub-figures are light positions for the acquisition of MLIC. The MLIC data is acquired using these light positions. The left side sub-figures are corresponding reconstructed normal maps using respective MLIC's. The DMD algorithm is used for Normal map reconstruction.*

We did several experiments asking our algorithm to estimate light positions. The algorithm was tested for light positions ranging from 3 to 100. We found that for lower no of MLIC, our algorithm performed better than the homogenous equally spaced acquisition MLIC. The Figure 7 demonstrates this result for low no light positions.

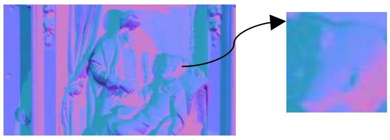
We believe that our algorithm performed better than homogenous equally spaced acquisition algorithm for low number of light positions (MLIC) because our algorithm was estimating light positions on the upper half of the dome mostly with elevation between 30 to 90 degrees since most of the surface normal point towards that direction. However, Homogenous equally spaced algorithm is not surface adaptive, and it estimates light directions with maximum possible distance between any two light positions. This makes homogenous equally spaced algorithm likely to choose light positions with very less elevation (near 0 degree). The images acquired corresponding to these low elevated light positions tend to be mostly dark thus contributing very little to Relightening algorithms.

a) Normal map reconstructed by 3 light positions estimated by our algorithm.

b) Normal map reconstructed by 3 light positions estimated by our algorithm.

c) Region of Interest

d) Normal map reconstructed by 3 Equally spaced light positions on Buddha Dataset



e) Normal map reconstructed by 3 light Equally spaced light positions on Sculpture Dataset.

f) Region of interest

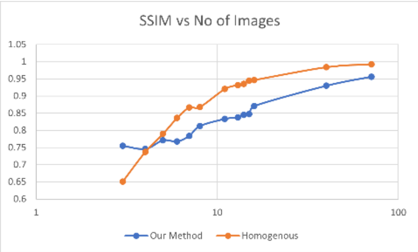
g) Ground truth normal map for Buddha

h) Ground truth normal map for Scupture.

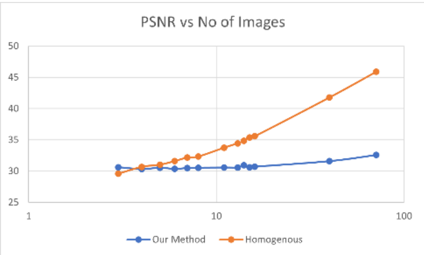
i) Ground truth Region of interest

Figure 7. *This figure compares the results of our algorithm with homogenous equal spaced algorithm at a lower no estimated light positions. The normal maps are reconstructed using DMD algorithm. The images for MLIC are acquired using our estimated light positions and equally spaced light positions. Our algorithm predicts better light positions for normal map reconstruction and is demonstrated further with region of Interests subfiguress.*

The Figure 8 and Figure 9 demonstrates the results for all the experiments we conducted on Buddha and Sculpture data respectively from light positions ranging from 3 to 100 with intervals. It can be seen in these results that homogenous equally spaced light positions are able to collect more information for more number of images in MLIC. We learn that after certain no of light positions it is difficult to optimize the light positions since homogenous equally spaced algorithm is optimal for capturing the most information unless the surface of object is very specular or has very specific reflectance profile. According to our results, we learn that it is only low number of light positions (MLICs) where light positions can be optimized to capture the most information within few images.

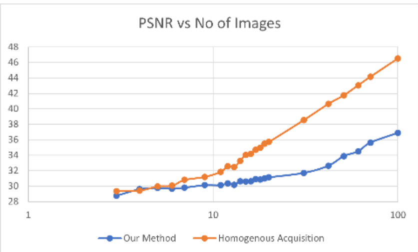


a)

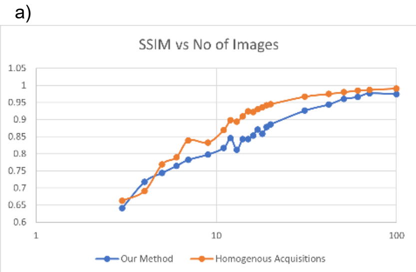


b)

Figure 8. *The a) SSIM and b) PSNR values are obtained by comparing the normal map reconstructed using our algorithm (orange) and homogenous equally spaced algorithm (blue) with the ground truth normal map. Our algorithm estimates better light positions for low no of images in MLIC but underperforms for high no of light positions. These results are for Buddha dataset.*



a)



b)

Figure 9. *The a) PSNR and b) SSIM values are obtained by comparing the normal map reconstructed by our algorithm and homogenous acquisition with the ground truth normal map. These results are for Sculpture dataset.*

Conclusion

The appearance of an object is a complex phenomenon that depends on multiple factors such as surface topography, transparency, gloss, color and etc. In this study, we attempted to use surface topography to estimate light positions. Our results indicate that our algorithm works better with only a low number of Multi Light Image Collections (MLICs) which might not be suitable for practical applications. As the number of images for MLICs increases, it becomes evident that only equally spaced light positions can gather more information unless there is a highly specular region. Furthermore, we observed that our algorithm tends to converge to equally spaced light positions when estimating higher numbers of light positions. It can be reflected in the SSIM graph.

We challenged ourselves to find the best light positions by using only one image. One can get only limited information about the object with one image. The estimation of translucency, gloss, and other appearance phenomena are complicated to understand with only one image except for color. With modern deep-learning methods, monocular depth estimation can provide an estimation of surface topography, but it is still not robust and accurate.

Although surface topography is important for material appearance, it is not enough to determine the optimal light positions for Reflectance Transformation Imaging. Based on our findings, we acknowledge that our initial hypothesis was only partially correct, and without putting in place specific restrictions, it is difficult to obtain the best light positions.

References

[1] R. J. Woodham, “Photometric method for determining surface orientation from multiple images,” *Optical engineering*, vol. 19, no. 1, pp. 139–144, 1980.

[2] T. Malzbender, D. Gelb, and H. Wolters, “Polynomial texture maps,” in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 519–528.

[3] P. Gautron, J. Křivánek, S. N. Pattanaik, and K. Bouatouch, “A Novel Hemispherical Basis for Accurate and Efficient Rendering.,” *Rendering Techniques*, vol. 2004, pp. 321–330, 2004.

[4] G. Pitard *et al.*, “Discrete modal decomposition: a new approach for the reflectance modeling and rendering of real surfaces,” *Machine Vision and Applications*, vol. 28, pp. 607–621, 2017.

[5] A. Zendagui *et al.*, “Quality assessment of dynamic virtual relighting from rti data: application to the inspection of engineering surfaces,” in *Fifteenth International Conference on Quality Control by Artificial Vision*, SPIE, 2021, pp. 94–102.

[6] H. Mytum and J. Peterson, “The application of reflectance transformation imaging (RTI) in historical archaeology,” *Historical Archaeology*, vol. 52, pp. 489–503, 2018.

[7] S. Saha, A. Siatou, A. Mansouri, and R. Sitnik, “Supervised segmentation of RTI appearance attributes for change detection on cultural heritage surfaces,” *Heritage Science*, vol. 10, no. 1, pp. 1–15, 2022.

[8] A. Siatou *et al.*, “A Methodological Approach for Multi-Temporal Tracking of Silver Tarnishing,” in *Proceedings of the 4th ACM International workshop on Structuring and Understanding of Multimedia heritAge Contents*, 2022, pp. 5–13.

[9] A. Siatou *et al.*, “New methodological approaches in Reflectance Transformation Imaging applications for conservation documentation of cultural heritage metal

objects," *Journal of Cultural Heritage*, vol. 58, pp. 274–283, 2022.
[10] A. Zendagui *et al.*, "Reflectance Transformation Imaging as a Tool for Computer-Aided Visual Inspection," *Applied Sciences*, vol. 12, no. 13, p. 6610, 2022.
[11] G. Pitard *et al.*, "Robust anomaly detection using reflectance transformation imaging for surface quality inspection," in *Image Analysis: 20th Scandinavian Conference, SCIA 2017, Tromsø, Norway, June 12–14, 2017, Proceedings, Part I 20*, Springer, 2017, pp. 550–561.
[12] G. Earl *et al.*, "Reflectance transformation imaging systems for ancient documentary artefacts," *Electronic visualisation and the arts (EVA 2011)*, pp. 147–154, 2011.
[13] R. Luxman *et al.*, "LightBot: A Multi-Light Position Robotic Acquisition System for Adaptive Capturing of Cultural Heritage Surfaces," *Journal of Imaging*, vol. 8, no. 5, p. 134, 2022.
[14] R. Luxman, M. Nurit, G. L. Goïc, F. Marzani, and A. Mansouri, "Next Best Light Position: A self configuring approach for the Reflectance Transformation Imaging acquisition process," *Electronic Imaging*, vol. 2021, no. 5, pp. 132–1, 2021.
[15] F. E. Nicodemus, "Directional reflectance and emissivity of an opaque surface," *Applied optics*, vol. 4, no. 7, pp. 767–775, 1965.
[16] R. Montes and C. Ureña, "An overview of BRDF models," *University of Grenada, Technical Report LSI-2012-001*, 2012.
[17] W. Yin *et al.*, "Learning To Recover 3D Scene Shape From a Single Image," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 204–213. Accessed: May 14, 2023. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/html/Yin_Learning_To_Recover_3D_Scene_Shape_From_a_Single_Image_CVPR_2021_paper.html?ref=https://githubhelp.com
[18] S. M. H. Miangoleh, S. Dille, L. Mai, S. Paris, and Y. Aksoy, "Boosting Monocular Depth Estimation Models to High-Resolution via Content-Adaptive Multi-Resolution Merging," in *Proc. CVPR*, 2021.
[19] M. Welponer, E. K. Stathopoulou, F. Remondino, and this link will open in a new window Link to external site, "Monocular Depth Prediction in Photogrammetric Applications," in *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Gottingen, Germany: Copernicus GmbH, 2022, pp. 469–476. doi: 10.5194/isprs-archives-XLIII-B2-2022-469-2022.
[20] J. MacQueen, "Classification and analysis of multivariate observations," in *5th Berkeley Symp. Math. Statist. Probability*, University of California Los Angeles LA USA, 1967, pp. 281–297.
[21] L. Reid, S. McDougall, and C. Erolin, "Sketchfab: An educational asset for learning anatomy," *Journal of Anatomy*, vol. 236, no. S1, p. 267, 2020.

Author Biography

Muhammad Arsalan KHAWAJA is joint PhD candidate at Université de Bourgogne, France and Norwegian University of Science and Technology (NTNU), Norway. He has M.Sc. in Computer Vision from Université de Bourgogne and B.Sc. in Aerospace Engineering from Institute of Space Technology, Islamabad. He is interested in Artificial Intelligence, Cultural heritage and Poetry.

Sony GEORGE, PhD is an Associate Professor at the Colourlab, Norwegian University of Science and Technology (NTNU), Norway. His research interests include color imaging, multi/hyper spectral imaging, imaging applications in cultural heritage. He has been involved in several national and EU projects in multiple roles, including MSCA CHANGE-ITN and PERCEIVE.

Franck MARZANI obtained his Ph.D. in computer vision and image processing in 1998. He is full professor at the Université de Bourgogne, Dijon, France since 2009. He is currently the head of the ImViA research laboratory (Imaging & Computer Vision). His research interests include acquisition and analysis of images. He has been developing an activity on feature extraction from color and multispectral images for classification purposes. These methodologies have been proposed in the frame of different applications.

Jon Y. HARDEBERG received his PhD degree from Ecole Nationale Supérieure des Télécommunications in Paris, France in 1999. He is now Professor of Colour Imaging at the Colourlab at NTNU - Norwegian University of Science and Technology, Gjøvik, Norway. His current research interests include spectral imaging, image quality, colour management, material appearance, and cultural heritage imaging, and he has co-authored more than 300 publications within the field.

Alamin MANSOURI is a full professor at the university of Bourgogne since 2015 and a member of ImViA laboratory where he is co-leading the CORES team. His current research is focused on multimodal imaging for appearance capture and modelling with main applications in Cultural Heritage and Industry.