Optics EXPRESS

Image processing pipeline for segmentation and material classification based on multispectral high dynamic range polarimetric images

MIGUEL ÁNGEL MARTÍNEZ-DOMINGO,[1,*] EVA M. VALERO,[1] JAVIER HERNÁNDEZ-ANDRÉS,[1] SHOJI TOMINAGA,[2] TAKAHIKO HORIUCHI,[2] AND KEITA HIRAI[2]

[1] *Color Imaging Laboratory, Department of Optics, University of Granada, Spain*
[2] *Color Image Engineering, Graduate School of Advanced Integration Science, Chiba University, Japan*
[*] *martinezm@ugr.es*

Abstract: We propose a method for the capture of high dynamic range (HDR), multispectral (MS), polarimetric (Pol) images of indoor scenes using a liquid crystal tunable filter (LCTF). We have included the adaptive exposure estimation (AEE) method to fully automatize the capturing process. We also propose a pre-processing method which can be applied for the registration of HDR images after they are already built as the result of combining different low dynamic range (LDR) images. This method is applied to ensure a correct alignment of the different polarization HDR images for each spectral band. We have focused our efforts in two main applications: object segmentation and classification into metal and dielectric classes. We have simplified the segmentation using mean shift combined with cluster averaging and region merging techniques. We compare the performance of our segmentation with that of Ncut and Watershed methods. For the classification task, we propose to use information not only in the highlight regions but also in their surrounding area, extracted from the degree of linear polarization (DoLP) maps. We present experimental results which proof that the proposed image processing pipeline outperforms previous techniques developed specifically for MSHDRPol image cubes.

OCIS codes: (100.2000) Digital image processing; (100.2960) Image analysis; (110.4234) Multispectral and hyperspectral imaging; (110.5405) Polarimetric imaging.

References and links

1. M. Martínez, E. Valero, J. Hernández-Andrés, J. Romero, and G. Langfelder, "Combining Transverse Field Detectors and Color Filter Arrays to improve multispectral imaging systems," Appl. Opt. **53**, C14–C24 (2014).
2. E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lightning* (Morgan Kaufmann, 2010).
3. J. J. McCann, and A. Rizzi, *The Art and Science of HDR Imaging* (John Wiley & Sons, 2011).
4. E. Hayman, B. Kaputo, M. Fritz, and J.O. Eklung, "On the significance of real-world conditions for material classification," in Proceedings of European conference on Computer Vision, (2004) pp. 253–266
5. O. Wang, P. Gunawardane, S. Scher, and J. Davis, "Material classification using BRDF slices," in Proc. CVPR IEEE (IEEE, 2009) pp. 2805–2811.
6. S. Tominaga, "Dichromatic reflection models for a variety of materials," Color Res. Appl. **19**, 277–285 (1994).
7. G. Healey,"Using color for geometry-insensitive segmentation," J. Opt. Soc. Am. A **6**, 920–937 (1989).
8. M. Varma, and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," Proceedings of European Conference on Computer Vision (2002) pp. 255–271.
9. H. Chen, and L. B. Wolff, "Polarization phase-based method for material classification in computer vision," Int. J. Comput. Vision **28**, 73–83 (1998).
10. L. B. Wolff, "Polarization-based material classification from specular reflection," IEEE T. Pattern Anal. **12**, 1059–1071 (1990).
11. G. Horvátz, R. Hegedüs, A. Barta, A. Farkas, and S. Âkesson, "Imaging polarimetry of the fogbow: polarization characteristics of white rainbows measured in the high Artic," Appl. Opt. **50**, F64–F71 (2011).
12. G. P. Können, "Polarization and visibility of higher-order rainbows," Appl. Opt. **54**, B35–B40 (2015).
13. S. Tominaga, and A. Kimachi, "Polarization imaging for material classification," Opt. Eng. **47**(12), 123201 (2008).
14. S. Tominaga, H. Kadoi, K. Hirai, and T. Horiuchi, "Metal-dielectric object classification by combining polarization property and surface spectral reflectance," Proceedings of SPIE/IS&T Electronic Imaging, 86520E (2013).

#297585 https://doi.org/10.1364/OE.25.030073
Journal © 2017 Received 8 Jun 2017; revised 23 Aug 2017; accepted 26 Aug 2017; published 16 Nov 2017

15. M. Martínez, E. Valero, and J. Hernández-Andrés, "Adaptive exposure estimation for high dynamic range imaging applied to natural scenes and daylight skies," Appl. Opt. **54**(4), B241–B250 (2015).
16. M. Martínez, E. Valero, J. Hernández-Andrés, and J. Romero, "HDR imaging - Automatic Exposure Time Estimation: A novel approach," Proceedings AIC conference in Tokyo **54**(4), pp. 603–608 (2015).
17. J. M. Medina, J. A. Díaz, and C. Vignolo, "Fractal dimension of sparkles in automotive metallic coatings by multispectral imaging measurements," ACS Appl. Matter. Inter. **6**, 11439–11447 (2014).
18. G. Ward, "Fast, robust image registration for composing high dynamic range photographs from hand-held exposures," Journal of Graphic Tools **8**, 17–30 (2003).
19. A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," Proceedings of International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG), pp. 49–56, (2007).
20. J. Im, S. Lee, and J. Paik, "Improved elastic registration for removing ghost artifacts in high dynamic imaging," IEEE T. Consum. Electr. **57**, 932–935 (2011).
21. R. Horstmeyer, *Multispectral Image Segmentation* (MIT Media Lab, 2010).
22. J. Theiler and G. Gisler, "Contiguity-enhanced k-means clustering algorithms for unsupervised multispectral image segmentation," Proc. SPIE Optical Science, Engineering and Instrumentation, 108–118 (1997).
23. S. Pal and P. Mitra, "Multispectral image segmentation using the rough-set-initialized EM algorithm," IEEE T. Geosci. Remote **40**, 2495–2501 (2002).
24. J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE T.Pattern Anal. **22**, 888–905 (2000).
25. S. Beucher and F. Meyer, "The morphological approach to segmentation: the Watershed transformation," Opt. Eng. **34**, 433–481 (1992).
26. H. Haneishi, S. Miyahara, and A. Yoshida, "Image acquisition technique for high dynamic range scenes using a multiband camera," Color Res. Appl. **31**, 294–302 (2006).
27. A. Kreuter, M. Zangerl, M. Schwarzmann, and M. Blumthaler, "All-sky imaging: a simple, versatile system for atmospheric research," Appl. Opt. **48**, 1091–1097 (2009).
28. W. Hubbard, G. Bishop, T. Gowen, D. Hayter, and G. Innes, "Multispectral-polarimetric sensing for detection of difficult targets," Proc. SPIE Europe Security and Defence, 71130L (2008).
29. Y. Schechner and S. K. Nayar, "Polarization mosaicing: high dynamic range and polarization imaging in a wide field of view," Proc. SPIE 48th Annual Meeting of Optical Science and Technology, 93–102 (2003).
30. P. Porral, P. Callet, P. Fuchs, T. Muller, and E. Sandré-Chardonnal, "High Dynamic, Spectral, and Polarized Natural Light Environment Acquisition," Proceedings of SPIE/IS&T Electronic Imaging, 94030B (2015).
31. : mirar pÃągina s S. Mann and R. Picard, "Being undigital with digital cameras," MIT Media Lab Perceptual (1994).
32. z J. McCann and A. Rizzi, "Veiling glare: the dynamic range limit of HDR images," Electr. Img. **6492**, 64913–64922 (2007).
33. J. McCann and A. Rizzi, "Camera and visual veiling glare in HDR images," J. Soc. Inf. Display **15**, 721–730 (2012).
34. A. Ferrero, J. Campos, and A. Pons, "Apparent violation of the radiant exposure reciprocity law in interline CCDs," Appl. Opt. **45**, 3991–3997 (2006).
35. P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in Proceedings of ACM SIGGRAPH pp. 31–40, (2008).
36. E. Talvala, A. Adams, M. Horowitz, and M. Levoy, "Veiling glare in high dynamic range imaging," ACM T. Graphic **37**, 26–37 (2007).
37. S. Wu, "Design of a liquid crystal based tunable electrooptic filter," Appl. Opt. **28**, 48–52 (1989).
38. Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of CCD imaging process," in Proceedings of Eighth International Conference on Computer Vision (ICCV) **1**, pp. 480–487 (2001).
39. D. Arnaud, *High Dynamic Range imaging: Sensors and Architectures* (SPIE, 2012).
40. M. Robertson, A. Borman, and R. L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," J. Electr. Img. **12**, 219–228 (2003).
41. T. Mitsunaga and S. K. Nayar, "Radiometric self-calibration," Proc. CVPR IEEE **1**, (IEEE, 1999) pp. 374–380.
42. M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. Seidel, and H. Lensch, "Optimal HDR reconstruction with linear digital cameras," Proc. CVPR IEEE (IEEE, 2010) pp. 215–222.
43. A. O. Akyüz and E. Reinhard, "Noise reduction in high dynamic range imaging," J. Vis. Commun. Image R. **18**(5), 366–376 (2007).
44. A. A. Goshtasby, *Image Registration: Principles, Tools and Methods* (Springer Science & Bussiness Media, 2012).
45. B. Zitova and J. Flusser, "Image registration methods: a survey," Image Vision Comput. **21**, 977–1000 (2003).
46. Y. Cheng, "Mean shift, mode seeking, and clustering," IEEE T. Pattern Anal. **17**, 790–799 (1995).
47. D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE T. Pattern Anal. **24**, 603–619 (2002).
48. D. L. MacAdam, "Color matching functions," in *Color Measurement* (Springer, 1981), pp. 178–199.
49. P. Soille, *Morphological Image Analysis: Principles and Applications* (Springer Science & Business Media, 2013).
50. J. Song, E. Valero, and J. L. Nieves, "Segmentation of natural scenes: Clustering in colour space vs. spectral estimation and clustering of spectral data," Proceedings of AIC Congress **12**, 2–8 (2014).
51. L. Hamers, Y. Hemeryck, G. Herweyers, M. Janssen, H. Keters, R. Rousseau, and A. Vanhoutte, "Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula," Comm. Com. Inf. Sc. **25**,

315–318 (1989).
52. R. Real, and J. M. Vargas, "The probabilistic basis of Jaccard's index of similarity," Syst. Biol. **45**, 380–385 (1996).
53. R. Real, "Tables of significant values of Jaccard's index of similarity," Misceltext period centered lania Zoologica **22**, 29–40 (1999).
54. S. Chandrasekhar, *Radiative Transfer* (Dover Publications, 1960).
55. A. Mehnert, and P. Jackway, "An improved seeded region growing algorithm," Pattern Recogn. Lett. **18**(10), 1065–1071 (1997).
56. J. Cao, P. Wang, Y. Dong, and Q. Xu, "A multi-scale texture segmentation method," Proceedings of 10th IEEE International Conference on Natural Computation (IEEE, 2014), pp. 873–877.
57. R. S. Medeiros, J. Scharcanski, and A. Wong, "Natural scene segmentation based on a stochastic texture region merging approach," Int. Conf. Acoust. Spee. (IEEE, 2013) pp. 1464–1467.

1. Introduction

Multispectral imaging as well as HDR imaging and polarimetric imaging are techniques developed to overcome the limitations of common scientific imaging systems. Multispectral imaging allows us to retrieve a larger amount of information compared with monochrome or RGB color images. It can help recovering spectral radiance or reflectance information pixel-wise, using some spectral estimation algorithms [1].

On the other hand, HDR imaging is used to capture useful image data from regions in the scene that present higher dynamic range than it is possible to capture with a single shot for non-HDR imaging systems [2, 3].

This work focuses on two main applications: image segmentation and objects' material classification. These two tasks have been addressed by many authors before.

Regarding material classification [4, 5], initially some authors studied it by using the dichromatic reflection model [6]. Using a spectroradiometer, they studied the spectral color signals coming from different objects and also its chromaticities. In general RGB or monochrome images were used to perform material classification [7]. Limitations of those methods are that the color reflected by objects is strongly dependent on the observation geometry. In [8], they based the classification in the texture of the objects. Assuming that objects with similar textures are made by similar materials.

Later on some authors approached the material classification problem via polarization-based imaging. In [9,10], they already tried to classify objects in two main classes: metals and dielectrics. They based their method in a thresholding of the Degree of Polarization (DOP). They used for the first time polarization features, which do not depend on object color and illumination. However their method was strongly dependent on the observation geometry, and only available for flat object surfaces. Besides, using polarimetric images also helps getting extra information about the content of the scene that is not visible with common imaging systems [11, 12].

In [13], they found that inspecting the curvature of DOP map around specular highlights, they could classify the objects into metal or dielectric, disregarding the observation geometry. This allowed them to perform classification including curved object surfaces as well [14], but sometimes the DOP calculation around the highlights is unstable. Specially if the image captured is somewhat noisy or the highlight region is rather large.

Based on this work, we aimed for overcoming these limitations, and as a first contribution, we proposed a new method which is also based on the study of the spatial distribution of the DOP map around the highlights. Moreover, in previous works [15, 16], we proposed an algorithm to automatically estimate the exposure times needed for the HDR capture. We applied it for monochrome as well as RGB cameras. Now, as second contribution, we have also adapted it for multispectral imaging channel-wise, making the capturing of HDR multispectral polarimetric images more automatic. Furthermore we also found and showed that when we rotate the LCTF for changing the polarization angle, we introduce some slight unwanted arbitrary translation, which was unaccounted for before. This translation is specially problematic if we want to get high resolution images and retrieve pixel-wise HDR multispectral polarimetric data (for instance for studying the metallic sparkling elements in special effects coating car paints [17]).

As explained in section 4, methods for registering the differently exposed LDR images captured to compose a HDR image [18–20] do not yield satisfactory results with already built HDR images. We therefore proposed, as third contribution, a pre-processing step before the registration, based in the compression of the dynamic range of the HDR scenes. This way, we also save computation time, since for each spectral band we only need to register one HDR image, and not many LDR images.

Regarding segmentation of multispectral images, other authors already tried conventional segmentation methods like k-means [21, 22], or more complex statistical methods like the rough-set-initialized EM algorithm [23]. As fourth contribution, we have proposed a novel segmentation full work-flow based on mean shift, skipping the computationally demanding Ncut method [24] used in [14]. We compared our results with those of Ncut algorithm as well as Watershed algorithm [25]. This last popular method is vastly used and directly implemented in most image processing programming libraries.

Few authors combined some of these techniques (multispectral, HDR and polarization imaging) before [26–29]. Even some authors tried combining all three of them together in a system able to capture all-sky HDR multispectral and polarimetric image data in less than two minutes, using two filter wheels and a fish-eye lens [30].

The remainder of the paper is structured as follows: in section 2, we explain the absolute radiometric calibration performed to the camera, as well as the spectral calibration of the system. In section 3 we briefly explain the capturing system and also the capture work-flow. In section 4, we explain all the image processing applied in order to get correctly registered HDR multispectral polarimetric images. In sections 5 and 6, we explain our proposed work-flows for the segmentation and classification applications respectively, showing the results obtained and comparing them with those of the methods proposed in [14]. Finally, in section 7, we draw the final conclusions of this work.

2. Radiometric and spectral calibration

When capturing HDR images via combination of multiple LDR exposures, the resulting image data is in a different scale from the original LDR images [31]. The former scale is exposure-time-independent, and the latter is exposure-time-dependent. Usually for the HDR images, we aim to get pixel values in the final image that are proportional to the amount of light coming from each region of the scene. For this purpose we perform a radiometric calibration of the camera.

2.1. Radiometric calibration

Before starting with the capture of the scenes, we performed an absolute radiometric calibration of the imaging system [32, 33] by measuring the Camera Response Function (CRF) of our camera (i.e. the way our camera responds to light). We first detached the LCTF from the camera, because the CRF is unaffected by the LCTF, and detaching it allows us to work with lower exposure times, while the signal level is still high enough to get useful image data. We prepared a scene containing ten homogeneous gray patches of different gray levels (see Fig. 1 left).

The patches were non-homogeneously illuminated with a Xenon Lamp (different levels of irradiance from direct light to shadow). This way we generated radiance signals covering a wide range of values in the scene (up to 4 orders of magnitude). We captured several images of this static scene using different exposure times and aperture values. After that, we substituted the camera by a spectroradiometer model *Photo Research PR-645*, and measured the radiance coming from the ten patches in the scene. Since the scene was static, the radiance outgoing from each patch was constant in time. Therefore we did ten radiance measurements for each patch and averaged them in order to reduce noise. The spectral radiances measured for the ten patches can be seen in Fig. 1 right.

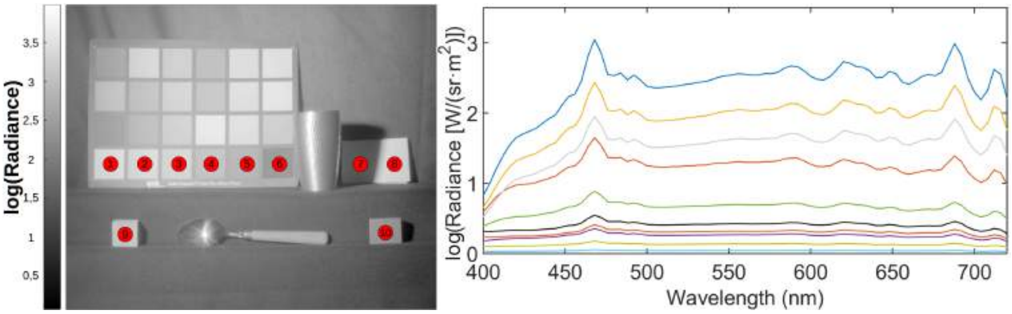


Fig. 1. Left: Tone-mapped version of monochrome HDR image of the scene used for radiometric calibration. The scale shows scene radiance (W/m^2sr) in a logarithmic scale. Right: spectral radiances of highlighted areas. The vertical axis is in logarithmic scale.

We integrated the measured spectral radiance values in the spectral range from 400 nm to 700 nm. Multiplying all these values by all exposure times used to capture the images for each aperture, we got a matrix of pseudo-exposure values. We call it pseudo-exposure because the exposure is the product of exposure time times irradiance ($T_{exp} \cdot E_e$), but our spectroradiometer measured radiance instead (L_e) of irradiance (E_e). Anyway this magnitude gives us information about the amount of light per unit area impinging on the sensor.

According to reciprocity law [34], halving the exposure time and doubling the radiance/irradiance, should result in the same sensor response. Since we can also average a small area in each patch to retrieve sensor responses for each exposure time and aperture value, then we could estimate the CRF for each aperture setting. The estimation of the CRF function was done by least squares fitting method. This function relates exposure/pseudo-exposure with sensor responses, and afterwards it is used to estimate a radiance map of the scene captured via multiple exposures technique [35].

Several factors influence the shape of the CRF. As *Ferrero et al.* pointed out in [34], reciprocity law does not hold when we study the response of the whole camera to light (rather than just the response of the sensor). This happens even working in raw mode. This is due to the fact that other phenomena also determine the way the camera responds to light. The strongest impact is that of optical veiling glare. Camera optics are not perfect, and as light goes through them, unwanted light from some regions of the scene, ends up impinging in pixels from other regions. This effect is specially severe when there are very brilliant regions present in the scene together with very dark areas. *McCann and Rizzi* have thoroughly characterized this phenomenon [3, 32, 33].

In this work, we have not considered correcting the effects of veiling glare (like other authors propose using some complex and not really practical system set ups [36]). We assumed that the imaging conditions are not so extreme so that the CRF curves measured for the different apertures hold for all the images captured. The CRF curves fitted from the measured data for the different aperture settings are shown in Fig. 2.

We can appreciate a clear linear behavior in the CRF curves. Depending on the aperture used during the capturing process, we used its corresponding CRF to estimate the HDR radiance map. We checked the accuracy of this absolute radiometric calibration including some gray patches in different scenes and measuring its integrated radiance with a spectroradiometer. The results are shown in Fig. 2. Since the purpose of this research was not to use the camera as an accurate imaging radiometer, we considered that the estimated and measured integrated radiances were similar enough for the purpose of our study. However if our aim had been to make a very accurate radiometric measurement using the camera, a more careful study of the CRF functions in different veiling glare conditions should be necessary.

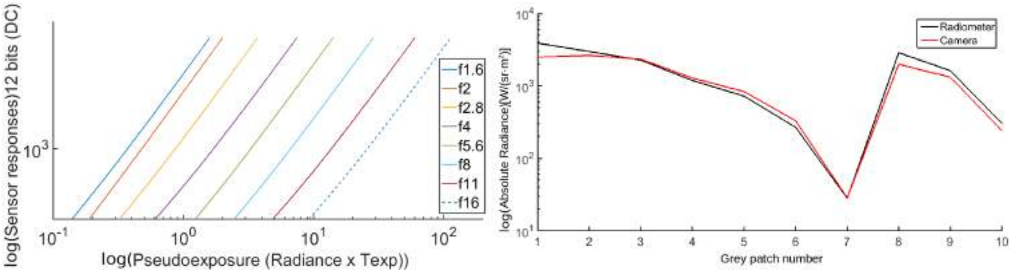


Fig. 2. Left: 12 bits Camera Response Functions (CRF) for different aperture settings. Both axes are displayed in logarithmic scale. Only the sensor response range above noise floor and below saturation is shown. X axis in $(W \cdot s/sr \cdot m^2)$. Right: comparison between radiances measured with the spectroradiometer (black) and estimated from the HDR radiance map (red). The vertical axis is in logarithmic scale.

The last part of the radiometric calibration consisted in identifying which pixels of the sensor were considered as hot pixels. These pixels yield a very high response even in darkness condition. This high response depends on the exposure time and temperature among other factors. We considered hot pixels those which average sensor response exceeded 5% of the dynamic range of our camera, or 205 digital counts, when we averaged 10 images with the aperture closed and 50 ms or higher exposure time. We stored the pixel coordinates of such pixels in order to make their impact negligible in the AEE method [15] during the capture.

Since the response of these pixels is not relevant, we located them in every captured image and substituted their pixel values by the mean value of their closest 8-neighbors. We found 35 hot pixels in total (less than 0.003% of the pixels contained in a captured image). However, since the AEE algorithm [15] keeps on looking for new exposure times until every pixel is correctly exposed in at least one shot, it is important to account for them in order to mitigate the impact of these pixels and achieve convergence.

2.2. Spectral calibration

The camera spectral sensitivity $S(\lambda)$, as well as the spectral transmittances of the different tune modes of the LCTF (weighted by the IR cut-off filter spectral transmittance) $T_i(\lambda)$, and the Spectral Power Distribution (SPD) of the lamp $L(\lambda)$ were measured using the spectroradiometer previously mentioned, and a monochromator. All these spectral measurements are used during the capturing process explained in section 3.

3. Image capturing

Our imaging device includes a 12-bits CCD monochrome camera (*QImaging Retiga* 1300) with hot mirror attached, and a LCTF model *CRI VIS* $10-20$ (see Fig. 3).

We illuminated the scenes using a 500 watts incandescent lamp. Thanks to the polarizing properties of the LCTF technology [37], we can capture images at different linear polarization angles just by rotating the LCTF along its optical axis. If an imaging device is to be used other than an LCTF, and it does not polarize the incoming light, then an extra rotating polarizing filter should be added in order to capture the polarization information. For a more detailed explanation about the capturing system please see [14]. A work-flow diagram for the capture process is shown in Fig. 3.

As explained before, we selected the exposure times needed to capture each spectral band using the Adaptive Exposure Estimation method (AEE) proposed in [15]. This method was

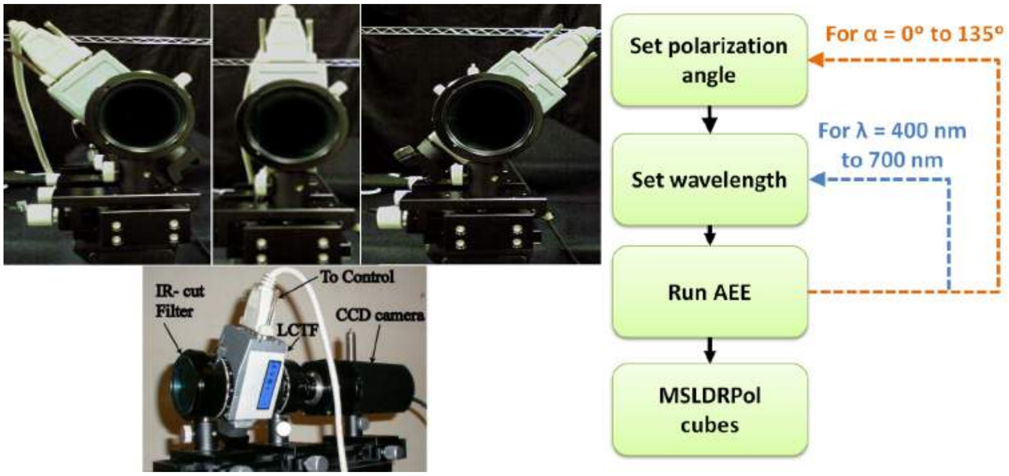


Fig. 3. Left: Top: front view of imaging system with different LCTF angles. Bottom: overall imaging system view. Right: workflow diagram of the capture process.

proposed and tested for a color RGB camera (*Canon EOS 7D*) and also for a monochrome camera. Both cameras were capturing all their spectral channels (3 or 1) in a single shot. Thus, we needed to adapt the method for the first time to a multispectral imaging system, which takes one shot to capture each spectral channel. We therefore run the AEE method for each individual spectral band.

As explained in [15, 16], the AEE method only needs as input one initial exposure time value. This value only has to accomplish one condition: at least some pixels of the resulting image must be properly exposed (*i.e.* neither below noise level, nor above saturation level). After this first shot is captured, the method computes the cumulative histogram, checking if still any pixels in our region of interest (ROI, which could be the entire image or just a portion of it), are either saturated or underexposed. If so, the algorithm computes a new shorter or longer exposure time (based on the camera's CRF properties). this new exposure time is computed to make the regions of the image close to saturation to be close to noise floor in the shorter exposure time, and the other way around in the longer exposure times. With the new exposure times, the algorithm continues capturing differently exposed shots. The iterative process goes on until it captures correctly the whole dynamic range of the scene, or it reaches a stopping condition (based on the percentage of correctly exposed total pixel population reached. This target percentage can be selected by the user, making the algorithm more flexible for different scene contents).

For the first shot of each band, due to the wide differences for different spectral bands in the spectral sensitivity of the sensor, the spectral transmittances of the LCTF and the SPD of the light, the exposure time value that accomplishes the mentioned condition for a spectral band (e.g. 550 nm), does not accomplish it for a different spectral band (e.g. 400 nm or 700 nm). This fact makes it necessary to adapt the initial exposure time value to run the AEE algorithm in each spectral band.

Our idea was to select a valid initial exposure time in the spectral band we usually use to point and focus the imaging system (550 nm since is the central band in the range from 400 nm to 700 nm). Once we chose this exposure time value, the AEE could run for this band. Afterwards, for each of the remaining spectral bands, we calculated a new initial exposure time value by weighting the chosen one for 550 nm, by a weighting function. This function was derived from the product of the relative weights of spectral sensitivity, SPD of light and the integral of the spectral transmittance of each mode of the LCTF multiplied by the IR cut-off filter used. The resulting relative weighting function is plotted in Fig. 5 left.

In order to record polarimetric information, we rotated the LCTF in front of the camera. As we assumed unknown the directions of maximum and minimum transmission through the polarizer, we make the capture in four relative rotation angles of the LCTF: $0°$, $45°$, $90°$ and $135°$. This yields the 4 relative polarization images $I(\lambda,\theta)$, $I(\lambda,\theta+45°)$, $I(\lambda,\theta+90°)$ and $I(\lambda,\theta+135°)$ respectively. Each of these images is a HDR image built from several LDR images captured with different exposure times. In section 6 we explain how to calculate the DoLP map out of these 4 polarization images.

4. Image pre-processing

There are three pre-processing steps in our proposed pipeline, which are applied to the MSLDR-Pol cubes resulting from the capture process: dark image subtraction, HDR image building and polarimetric registration (see Fig. 4).

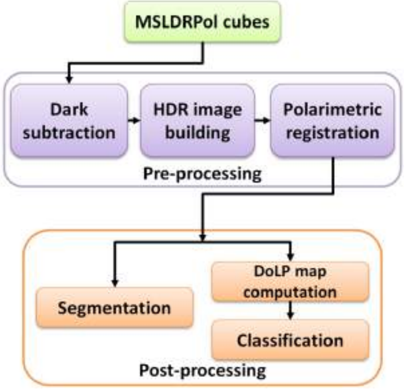


Fig. 4. Work-flow diagram of the whole image processing pipeline from capture to the final classification step.

Once the capture is finished, for each spectral band and polarization angle, we get a set of images captured with different exposure times. The number of images for each spectral band and polarization angle would depend on the dynamic range of the scene we are capturing, as well as the parameters we set for the AEE method, and the capturing device responsivity. Since our AEE parameterization was configured to get minimum bracketing sets (i.e. the set of different exposure times needed to capture correctly every point of the scene with the lowest number of shots), we obtained from two to three images per band and polarization angle.

4.1. Dark subtraction

As it usually happens in most digital imaging systems, there are some sources of noise which make the sensor yield a non-zero response even in the absence of light signal [38,39]. The impact of this dark noise component can be reduced by subtracting black images from the different images captured from the scene. These black images are captured using the same exposure times as the scene images, but the aperture of the camera is completely closed. For each exposure time value used during the capture of all LDR images, ten black images were captured and averaged (in order to reduce noise). Afterwards, and before the next step of building the HDR images, these averaged black images were subtracted from every LDR image captured.

4.2. HDR image building

Once the values of the hot pixels are corrected (as explained in section 2.1) and the dark noise is subtracted from every LDR image, we are ready to build the HDR image of each spectral band and polarization angle. We did so using Eq. (1), previously used in [2, 35].

$$E'_{\lambda\theta}(x,y) = \frac{\sum_{n=1}^{N} \omega(\rho_{n\lambda\theta}(x,y)) \cdot \frac{CRF^{-1}(\rho_{n\lambda\theta}(x,y))}{\Delta t_n}}{\sum_{n=1}^{N} \omega(\rho_{n\lambda\theta}(x,y))} \tag{1}$$

where E' is the value of HDR radiance map generated, λ is the spectral band index, θ is the polarization angle index, (x, y) are pixel coordinates, N is the number of shots captured for each spectral band, ρ is the sensor response in each LDR image captured, CRF^{-1} the inverse camera response function estimated as explained in subsection 2.1, Δt_n the exposure time used for shot number n, and ω the weighting function used to construct the HDR image (see Fig. 5). In the literature we find some other shapes for this function like the triangular or hat shape [35], derivatives of the CRF [31, 40, 41], or more complex shaped functions [42]. We found the smooth (also called broad hat function) function proposed in [2, 43] and shown in Fig. 5 right, was less prone to introduce artifacts in the final HDR image than the other alternatives.

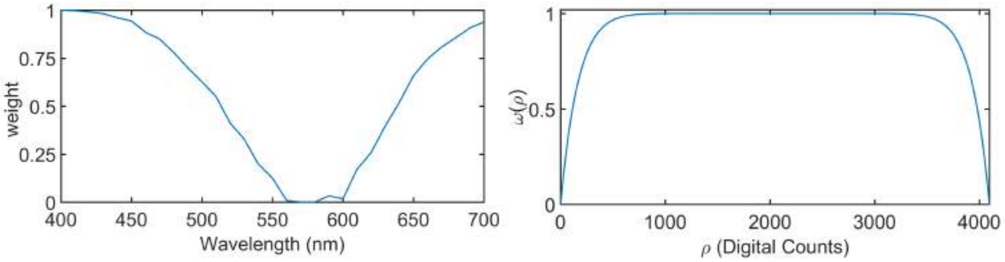


Fig. 5. Left: Weighting function used to scale the initial exposure time values to run the AEE algorithm in each spectral band. Values around [560 nm, 600 nm] are non-zero values. Right: Weighting function used to build HDR images from multiple 12-bits LDR images.

4.3. Polarimetric registration

Once the HDR images were built for each spectral band, and each polarization angle, we checked for misalignments between different bands and polarization angles within each band. We found no noticeable misalignment between spectral bands. The depth of field was larger than the longitudinal chromatic aberration in our capturing conditions. Thus, there was no need for registration among different bands captured with the same polarization angle.

On the other hand, we did find misalignment between the images corresponding to different polarization angles for a given spectral band. Apparently, the rotation of the LCTF introduced small translations of the images. Thus, we had to perform image registration. Since all spectral bands were correctly aligned for each polarization angle individually, as long as we could find the transformation needed to align the four polarization images for one spectral band, we could apply the same transformation to every spectral band. These misalignments could potentially worsen the metal-dielectric classification accuracy in the post-processing steps of the proposed work-flow. The difficulty in this step lies in the fact that the images we are going to align are already HDR images.

Several algorithms have been proposed for the registration of the differently exposed LDR images before building the final HDR image. Some of them are based in percentile threshold bitmaps like [18]. Some others are based on SIFT key points extraction like [19], or novel proposed methods such as optimum target frame estimation [20]. However all these methods are not applicable for the registration of different HDR images already built. This is due to the fact that when the HDR images are already built, their pixel values are too different from the darkest regions to the brightest ones. Key features such as edges or corners are based in the high contrast between different objects. However, the magnitude of this contrast is very different for HDR images depending whether we are in a very dark or a very bright region.

We tried standard feature-based registration methods for estimating the transformation needed in order to align the HDR images of the 4 different polarization angles. We considered translation, rotation, scale and shear transformations, as well as combinations of them, but none succeeded. Therefore we decided to modify the registration work-flow by including a three-stage pre-processing step before applying the standard key-point based registration procedure [44, 45].

The first stage is dynamic range logarithmic compression. We compress the dynamic range in logarithmic scale to reduce the large difference between very bright and very dark regions. The second stage is normalization. We scale the image values in the range [0, 1], by subtracting minimum value and dividing by maximum value. The third stage is contrast enhancement. We stretch the contrast of the image so that we allow 5% underexposed pixels and 5% saturated pixels. This way, even if up to 5% of pixel population is either very bright or very dark, we will still highlight the details in the middle exposure region. Figure 6 shows a portion of overlay between the polarization images corresponding to 0° and 135° for the spectral band of 550 nm, before and after pre-processing and registration are applied.

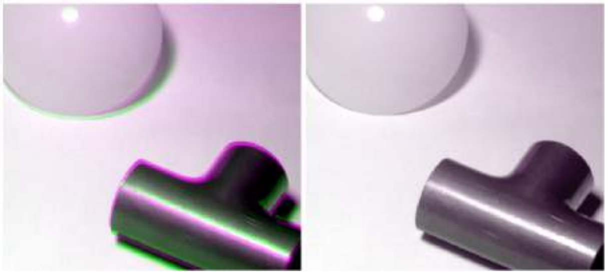


Fig. 6. Image overlay 0° over 135° for 550 nm. Left: before registration. Right: after registration.

We see that before registration, there are colored bands close to the edges of the objects and their highlights. This means there is misalignment between the two polarization images. A similar situation is found for all pairs of images corresponding to different polarization angles. After the registration procedure, we can see in Fig. 6 (right) how the two images are correctly aligned. So we can conclude that the registration was performed satisfactorily.

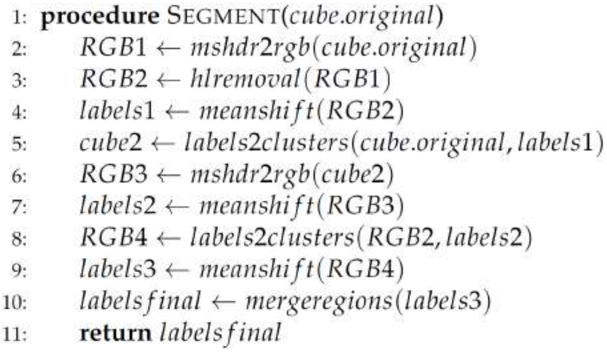
The final transformation found for all images was a translation. The magnitude of this translation depended on the scene, ranging from 5 to 15 pixels both in horizontal and vertical directions.

After all the described pre-processing steps are finished, our Multispectral High Dynamic Range polarimetric image cubes are ready to use. We focused our work in two main applications: segmentation and material classification. For each of those, we based our efforts in trying to overcome the limitations found in the methods proposed in [14], which is to our knowledge the only reference available so far which deals with MSHDRPol image cubes for segmentation and material classification. We explain both applications in the next sections. The first one is dedicated to the segmentation method description and validation experiments, and the last one to the classification method and its validation experiments.

5. Segmentation method and evaluation

5.1. Summary of the proposed method

The segmentation procedure we propose is composed by several steps. We used Matlab for processing the captured image data. In the pseudo-code of algorithm in Fig. 7 we describe the different steps of the segmentation. We explain each of them in the following subsections from 5.2 to 5.6. The input to the algorithm is a MSHDRPol cube obtained after applying the pre-processing method explained in section 4.



```
procedure SEGMENT($cube.original$)
    $RGB1 \leftarrow mshdr2rgb(cube.original)$
    $RGB2 \leftarrow hlremoval(RGB1)$
    $labels1 \leftarrow meanshift(RGB2)$
    $cube2 \leftarrow labels2clusters(cube.original, labels1)$
    $RGB3 \leftarrow mshdr2rgb(cube2)$
    $labels2 \leftarrow meanshift(RGB3)$
    $RGB4 \leftarrow labels2clusters(RGB2, labels2)$
    $labels3 \leftarrow meanshift(RGB4)$
    $labelsfinal \leftarrow mergeregions(labels3)$
    return $labelsfinal$
```

Fig. 7. Segmentation algorithm.

Our proposed segmentation method involves sequential application of the mean-shift algorithm [46, 47], together with the clustering of all pixels belonging to the same superpixel after a mean shift iteration is applied. Each time we apply a new iteration of mean-shift, we reduce the number of superpixels in the pre-segmented image.

We first produce a usable RGB image from the original cube with highlights removed (see subsections 5.2 and 5.3). Afterwards we apply mean-shift to get an initial crude estimation of labels for the different regions (see subsection .5.4). Later on, we use the original spectral cube and these labels to perform an initial clustering of superpixels data. From here on, we use only the RGB image of this initial clustering to further apply mean-shift and clustering to obtain a second set of more refined labels (see subsections 5.4 and 5.5). These labels are finally segmented and merged (see subsection 5.6) to obtain the output of the segmentation algorithm.

5.2. Multispectral HDR to RGB (mshdr2rgb)

This function receives as input a HDR multispectral image cube of integrated radiances, and uses the CIE 1931 Color Matching Functions [48], to get a color image with HDR information. This is not a standard XYZ image. We did not include any standard illuminant in the calculation because we are dealing with integrated radiance signals channel-wise, which include illuminant information. This color image is afterwards transformed into RGB color space. Later on we perform a tone mapping to get finally a LDR RGB image. The tone mapping includes logarithmic compression, normalization, contrast enhancement, histogram equalization and saturation boost. Even though the final RGB image is not a standard RGB color space image, it is still valid for our segmentation purpose. However if we wanted to get a standard RGB color space image, then we should normalize the multispectral HDR image cubes dividing by the corresponding illuminant present in the scene first, and then use a standard CIE illuminant to compute the tristimulus values. In Fig. 8 left, we can see an instance of the *RGB1* image for one of the scenes.

5.3. Highlights removal (hlremoval)

This function receives an RGB image containing objects with highlights, and returns another RGB image where the highlights are reduced or eliminated. This is done by computing the negative of the image (for an $8 - bits$ image: $Image_{negative} = 255 - Image$). Then we perform a morphological operator called region filling [49], to the negative image, and finally calculate again the negative of the filled image. Figure 8 shows the *RGB1* image for one scene and the resulting *RGB2* image after removing highlights. It is important to note that we remove the highlights to help in the segmentation task. As we will explain in subsection 6, we will use the information in the highlight regions for this task, getting it again from the original pre-processed images (*cube.original*).

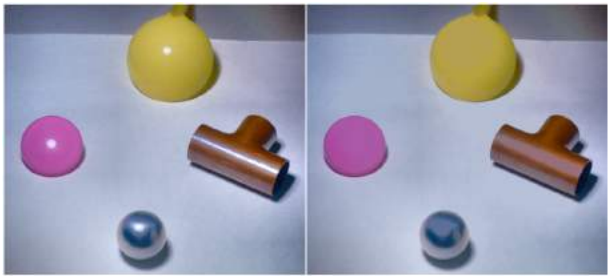


Fig. 8. Left: RGB1 image rendered. Right: RGB2 image after highlights removal.

5.4. Mean shift

This function receives as input and RGB image and returns a label image. A label image is an image where each pixel value corresponds to a label. Mean shift algorithm [46, 47] is a pre-segmentation performed to find areas of the image where the pixels have similar values, and group them together under the same label. This cluster is often called superpixel.

5.5. Labels to clusters (labels2clusters)

Once the labels images are generated, we can use a multispectral image cube or a RGB image to average the pixel values in each spectral or color channel in order to generate a new cube or image which is much more simple than the original one. This is what this function does. It receives a labels image and an original image (color or multispectral), and returns an image of the same kind of the original, where the pixel values for those pixels under the same label (belonging to the same superpixel) are the same. This value is the average of all pixels in the same superpixel.

This new image, can in turn be an input for a new mean shift iteration. In Fig. 9 left, we can see an RGB renderization of one of the highlights-removed images (*RGB2*), and the resulting RGB image (*RGB3*), after applying the first mean shift iteration.

The performance was found to be better if we apply *labels2clusters* method using the labels images and the spectral cubes in the first iteration, and then the labels images and the RGB highlight-removed rendered image for the second iteration. The reason why the first time we use the spectral cube (*cube.original*) and the second time we use the RGB image (*RGB2*) to perform the *labels2clusters* method, is that when the image is still rather complex, averaging spectral information poses a better input for the next step. However once the second mean-shift iteration has been performed, the RGB images are already simpler, yielding satisfactory results. Thus we do not need to use the spectral cubes anymore. This way we can save a step of *mshdr2rgb* method.

5.6. Region merging (mergeregions)

After mean shift is applied and we have used the labels image to cluster regions in our image, we aimed to reduce the number of clusters as much as possible, yet keeping the different objects apart both from each other and from the background. We then apply an iterative region merging process which consists in the following steps. First of all we find the background, assuming it is the region with the biggest pixel area. We know that this might not occur in some cases where there is a very large and homogeneous object in the scene. However for the general case this condition worked fine. Afterwards we compute the adjacency matrix of the labels image ignoring the background. This means that any region adjacent with the background would account as no adjacent, since we will never want to merge any object with the background. Once we have the adjacency matrix, we compute a similarity matrix. We treat the image as a graph, in which the weight ($\omega_{i,j}$) between nodes i and j is ∞ for non adjacent regions, and calculated as shown in Eq. (2) for adjacent regions.

$$\omega_{i,j} = \sqrt{\sum_{k=1}^{K} (\rho_{k,i} - \rho_{k,j})^2} \tag{2}$$

where K is the number of color or spectral channels, and $\rho_{k,a}$ is the average sensor response for image channel k and region a. The fourth step is thresholding the similarity matrix. Those regions which similarity lies below a threshold value found by trial an error to be convenient for all our scenes were merged. Merging two regions means computing the average sensor response values for each channel and giving the same label to both of them. Therefore after each merging, we recomputed the adjacency and similarity matrices again. This iterative process goes on until there are no adjacent regions for which similarity metric falls below the threshold. In Fig. 9, we can see a false color labels image before and after applying the region merging. We can see how the number of regions in the image was reduced from 36 to 8.

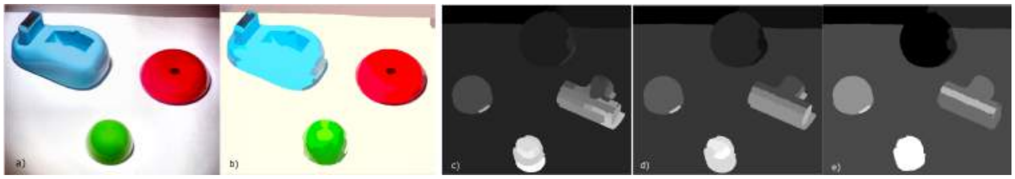


Fig. 9. Left: *RGB2* image after highlights removal (a) and *RGB3* image after the first iteration of mean-shift and *labels2clusters* processing (b). Right: Grayscale label images. Before region merging (c with 36 regions). After thresholding with $th = 60$ (d with 15 regions remaining), and after thresholding with $th = 120$ (e with 8 regions remaining).

In this scene, the optimal number of regions would be 6 or 7. They would correspond with the 4 objects, the ground, and the wall (divided in two sides by the top-most object). We are reasonably close to this number using a threshold value of 120. Increasing this value would make the walls merge with the top-most object. The threshold value used performs reasonably well even with the residual highlights or shadows remaining in the scene after the previous steps.

In general, a threshold value of 120 performed rather well for all scenes tested. However a fine adjustment of this value would increase the performance of the segmentation for certain scenes. The region merging is performed attending to the similarity between adjacent regions, therefore it is independent of the number of objects present in the scene. A limitation of this method is that two very similar adjacent objects would be merged as single object, and also an object with two very different regions would be segmented as two separate objects.

5.7. Evaluation of segmentation procedure

In order to evaluate our segmentation results, we need to compare them with a reference segmentation or ground truth. For this purpose we segmented manually the images captured (see Fig. 10), with the help of a graphical user interface [50]. We considered this segmentation as a perfect segmentation, and then compared how close our segmentation was to this ground truth (see Fig. 10). We considered to be background whatever was not an object in the scene.

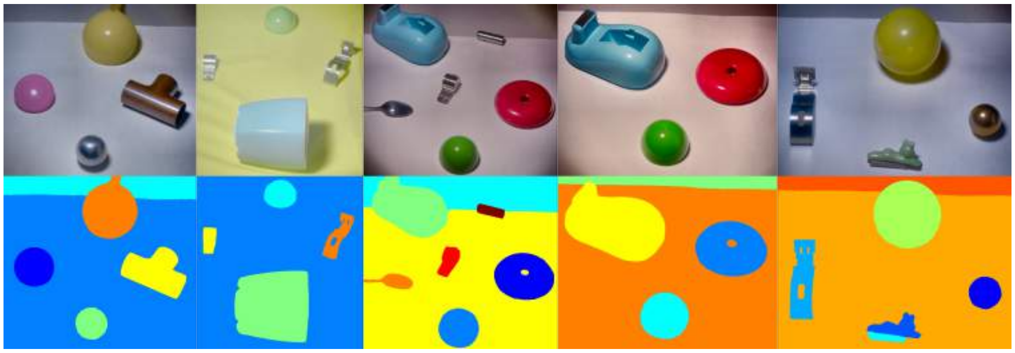


Fig. 10. Top row: RGB renderization of original spectral cubes. Bottom row: benchmark manually segmented. From left to right, scenes from 1 to 5.

The figure of merit we used for evaluation was the Jaccard's similarity index [51–53]. This metric measures the similarity between finite sample sets. It is defined as the size of the intersection between sets divided by the union of these sets. In our case, we use it to compare the benchmark labeling and the automatically segmented labeling for each scene. It is a measure of how much each labeled object overlaps with its ground truth labeling. A Jaccard's index value of 100% would mean a perfect match, while the lower this value is, the more dissimilar the two segmentation results would be.

In order to compare our segmentation with that proposed in [14], we also used Ncut algorithm with the RGB highlights.removed images (RGB2) of the scenes captured. Besides, we included a third widely used image segmentation method implemented in most image processing toolboxes like that of Matlab . This method is called the Watershed method, and it is based on morphological Watershed transforms, embodying edge detection, thresholding and region growing [25]. Table 1 shows the Jaccard index values for different scenes captured and segmented using the two methods.

Table 1. Jaccard index values for the three segmentation methods compared.

Scene	1	2	3	4	5
Our method	92.49%	87.62%	91.99%	95.23%	90.82%
Ncut	81.11%	84.72%	52.94%	90.01%	39.01%
Watershed	59.61%	71.85%	52.77%	54.41%	35.20%

We can see how, scene-wise, the smallest improvement of our proposed method is 2.9% better than Ncut, which we could consider almost negligible, and 15.77% better than Watershed. The largest improvement is 51.81% from Ncut and 55.62% from Watershed. If we average results, we find that our proposed method yields a mean segmentation accuracy of 91.63%, while Ncut method yields a mean accuracy of 69.56% and Watershed of 54.77%. That means that applying our method resulted in a relative 31.73% increase in segmentation accuracy from Ncut, and 36.86% from Watershed.

Regarding computation time, the slowest step is the meanshift which is performed 3 times. However there were no significant differences found in computation time for the different scenes and the different parameters used. In a 64 bits Intel i7 computer with 16GB of RAM, the process takes always less than 10 seconds (8.6 seconds the fastest and 9.8 seconds the slowest out of all conditions tested). Watershed method was also quite fast, in the order of 7 seconds minimum to 10 seconds maximum for all scenes, and Ncut was the slowest with a run time of more than 1 minute in all cases. No trend was found relating computation time to number of objects present in the scenes.

6. Classification method and evaluation

6.1. Summary of the proposed method

When dealing with the classification problem, we based our work in the method used in [13, 14], which is based in studying the DoLP map curvature around the highlights. The DoLP map is computed from the four polarization angles images. Using these four images, we can calculate the Stokes parameters images [54], as well as the DoLP map for wavelength λ using Eqs. (3 - 6).

$$S_0 = I(\lambda, \theta) + I(\lambda, \theta + 90^\circ) \tag{3}$$

$$S_1 = I(\lambda, \theta) - I(\lambda, \theta + 90^\circ) \tag{4}$$

$$S_2 = I(\lambda, \theta + 45^\circ) - I(\lambda, \theta + 135^\circ) \tag{5}$$

$$DoLP(\lambda) = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \tag{6}$$

In [13, 14], they classified the objects into metal and dielectrics by studying an area of 9 x 9 pixels centered in the brightest point of the highlights, detected by simple thresholding in the luminance image. Then the curvature along the direction of maximum variation of the DoLP surface was computed. According to the sign of this curvature (which indicates if the surface is concave or convex), the object can be classified into metal or dielectric material.

When we implemented this method, we found some difficulties in the calculation of this magnitude. Sometimes, the shapes of the highlights and the DoLP maps are not regular or rounded, but rather spiky and having elongated or non-rounded shapes. In these cases, calculating the curvature in both the direction of maximum variation of the DoLP map or even the direction of the highlight in the HDR image, was not yielding satisfactory classification results. We even tried studying the mean curvature along all directions around the brightest point, but results were still not satisfactory.

In Fig. 11(a), we can see two examples of DoLP map surfaces together with their corresponding HDR image areas (Fig. 11(b) and Fig. 11(c)) and highlight surfaces. Both these highlights belong to two different metal objects. However, using the curvature of the DoLP map resulted in wrongly classifying one of these two objects as a dielectric. With this kind of DoLP pattern (with the brightest point in the middle of a ramp-like surface), judging the curvature is somewhat fuzzy, and it varies quite a lot depending on the direction we choose. We found this pattern rather often, and not only in objects presenting elongated highlights.

We propose a new method of material classification based on the DoLP maps which consists in thresholding a ratio between the DoLP values within the highlight and in the surrounding area (see Fig. 11(d) and Fig. 11(e)). Eq. (7) shows how to calculate this DoLP ratio (R) value for each highlight. N is the total number of pixels in the highlight central area. M the total number of pixels in the diffuse surrounding area. δ_n^{hl} is the DoLP value in pixel n of the highlight area and so δ_m^{su} is the same for the surrounding area. As we can observe this is just a ratio of average DoLP values in the two areas.

$$R = \frac{\sum_{n=1}^{N} \delta_n^{hl}}{\sum_{m=1}^{M} \delta_m^{su}} \cdot \frac{M}{N} \tag{7}$$

Between these two areas we keep a safe zone to avoid including pixels of a highlight area in the computation of the surrounding area or the other way around. The central and surrounding areas are selected via a double region growing morphological image processing operator [55], around the brightest points of each detected highlight. We used twice a disk-shaped growing element of radius 5. First we grow to get the safe area out of the central area, and then we grow once more to get the surrounding area. Afterwards we just subtract both to end up with the ring-shaped surrounding area.

Once the ratios for all the objects included in our scene set are computed as shown in Eq. (7), we noticed that their values for metal objects were lower than for dielectric objects. Thus, our idea was to make a thresholding of this ratio value to classify each highlight. In order to find an optimal threshold value, we propose to use a set of highlights which labels are known (train set). For this set, the optimal value is found by brute force approach in a fast and easy way. Once we find the threshold value yielding the highest classification accuracy for the training set, we compute its percentile value for the whole distribution of ratio values in this set. Afterwards, when we get a set of testing highlights, we can compute the optimal threshold value using the same percentile with the new distribution of ratio values and check the classification accuracy for the validation of our proposed method.

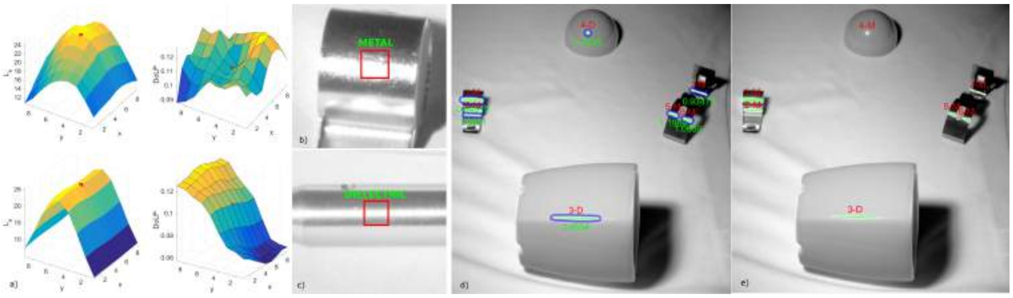


Fig. 11. a) Example of two HDR highlight surfaces (left) of a metal object, and DoLP surfaces (right). b) and c) 550 nm band images. The brightest point is highlighted with a red mark on the surfaces. The top example (b) is correctly classified as metal by the previously proposed method, but the bottom example (c) is not. d) Example of highlights detected (green) in scene 2 and their surrounding areas (blue). Each highlight is automatically classified as metal (M) or dielectric (D), according to its ratio value (green text). Proposed method using a threshold value of 1.2. e) Method in [14]

6.2. Evaluation of classification method

We have created a benchmark dataset of manually classified highlights detected in the scenes. After running the automatic classification using our proposed method (explained above), and the method proposed in [14], we compared their respective classification accuracies.

We randomly divided the full highlights data set (72 highlights from 10 scenes captured only in the spectral band of 550 nm) into two halves. One of them was the training set, and the other one the test set. In the training set, we looked for the optimal threshold value by a brute force approach. We tested all threshold values from 0 to 4 by steps of 0.1. In Fig. 12 left, we can see the classification accuracy in the training set for each threshold value.

As we can see, the optimal threshold value calculated was 1.5 with an accuracy of 88.57% for the training set. Since we divided the data set into train and test sets randomly, different instances of this process yielded different overall classification accuracies. After 20 trials, we found that the same threshold value was estimated in all of them. Applying this threshold value for the classification of the test sets, gave us accuracy values ranging between 86.11% and 91.89%. The mean accuracy for the 20 trials was 89.19%. As an overall accuracy for the whole data set, using a threshold value of 1.5, we got a 90.28% classification accuracy. In Fig. 12 right, we see all the highlights in a cloud of points relating DoLP ratio and radiance of brightest area in each highlight. The threshold value found as optimal for the full set is drawn in green color.

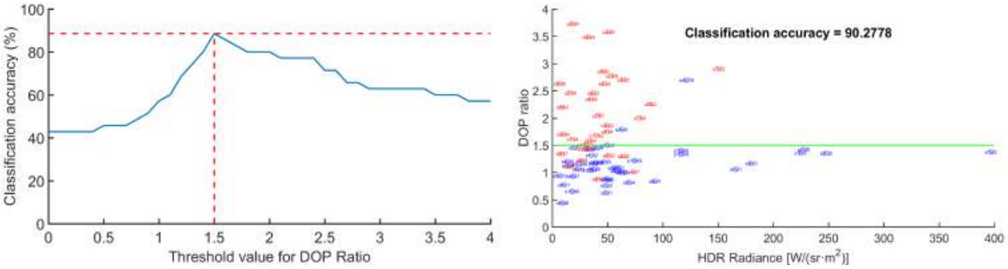


Fig. 12. Left: Classification accuracy vs threshold value for training set. Rigth: DoLP ratio vs HDR Highlight radiance for the whole data set of highlights. Red: dielectric objects. Blue: metal objects.

On the other hand, using the classification method in [14] based on curvature of DoLP map, we got a total accuracy of 66.67% for the full set. We found that this method was failing to classify most objects for which the brightest point in the highlight did not correspond to a local maximum or minimum in the DoLP map. On the other hand, a fair comparison should mention that this method does not require for a training set.

In general, using our method we also observe that the classification accuracy for the dielectric objects is lower (77.42%) than for the metal objects (95.12%). In other words, the number of false positives in the metal class (classifying a dielectric as a metal) is 28.52%, and the number of false positives in the dielectric class (classifying a metal as a dielectric) is only 4.88%. This could be due to the fact that dielectric objects are from very different base materials (plastics, ceramics, etc, which can also be coated). Thus they behave in a more heterogeneous way in terms of DoLP ratio than metals.

We have tried performing error analysis based on the material composition both of metals (copper, steel, aluminum) and dielectrics (plastic, ceramics), but we did not find any significant trend in the classification error according to material composition. Therefore we can not conclude that certain specific materials present a higher or lower classification accuracy in the proposed method.

Finally, if we only consider the results of the five scenes which were used as well in the segmentation procedure, we get a global classification accuracy of 90.91%. Therefore for the full framework, and in realistic conditions, we would get roughly more than 9 out of every 10 objects correctly classified.

Regarding computation time, the whole classification was done in less than a third of a second for all scenes (from 109 ms minimum to 312 ms maximum), with no significant differences or trends found between scenes or number of highlights.

7. Conclusions

We have introduced a complete image processing pipeline from image capture to object segmentation and classification, with a remarkable number of novel features involved.

Our capture device was successfully calibrated and the CRF determined. We used this function to build four HDR images at different polarization angles for each spectral band, which were afterwards correctly registered after range compression, normalization and contrast enhancement. Our method succeeded in registering the already built HDR images where common methods proposed for registering LDR images failed.

We also have introduced new segmentation and classification approaches using as building blocks existing techniques which are combined aiming to simplify the computational load of previous approaches to this problem for MSHDRPol image cubes. We got a relative improvement of 31.73% in segmentation accuracy and 35.41% in classification accuracy.

As future work, we will try to improve the segmentation procedure by assessing the clustering made after each mean-shift, so we can automatically perform a fine tune of the threshold for the final region merging step. How to tune this parameter automatically to find its ideal value is a matter of future study. We could use clustering quality metrics to make the full segmentation process completely automatic. Besides, including further information during this step, such as texture information as they do in [56, 57], could improve the limitations of the proposed method when two adjacent objects are very similar or there are more complex objects with two different colored regions.

Regarding the classification, we believe that including also infrared information by extending the spectral range we are capturing, could help us including more candidate classes. However the LCTF model we used in this study could not be used above 720 nm.

We would also like to test our processing pipeline in more challenging conditions, including more scenes like outdoor imaging. For this purpose, a new capturing architecture would probably be needed in order to accelerate the capturing process. This way we would reduce the impact of the uncontrolled changing conditions like illumination, moving elements, atmospheric phenomena, etc. On the other hand, uncontrolled conditions could lead us to the absence of highlights for some objects. This would make even more challenging the classification task.

Finally, studying how well the system recovers spectral radiance signals pixel-wise from calibrated HDR multispectral camera responses by means of regression methods such as kernels or neural networks, is a promising line of research for spectral estimation in natural conditions (including HDR scene information).

Funding

Spanish Secretary of State of Research, Development and Innovation (SEIDI), within the Ministry of Economy and Competitiveness (DPI2015-64571-R).

Acknowledgments

We would like to acknowledge Professor Abdelhameed F. A. Ibrahim from the Computer Engineering Deptartment, of College of Computer, Qassim Private Colleges, (Saudi Arabia) for his support.