https://doi.org/10.2352/ISSN.2470-1173.2019.15.AVM-031


# Automatic shadow detection using hyperspectral data for terrain classification

***Christian Winkens, Veronika Adams, Dietrich Paulus, Active Vision Group, Institute for Computational Visualistics, University of Koblenz-Landau, Germany***

## Abstract

*Hyperspectral image classification has received more attention from researchers in recent years. Hyperspectral imaging systems utilize sensors, which acquire data mostly from the visible through the near infrared wavelength ranges and capture tens up to hundreds of spectral bands. Using the detailed spectral information, the possibility of accurately classifying materials is increased. Unfortunately conventional spectral cameras sensors use spatial or spectral scanning during acquisition which is only suitable for static scenes like earth observation. In dynamic scenarios, such as in autonomous driving applications, the acquisition of the entire hyperspectral cube in one step is mandatory. To allow hyperspectral classification and enhance terrain drivability analysis for autonomous driving we investigate the eligibility of novel mosaic-snapshot based hyperspectral cameras. These cameras capture an entire hyperspectral cube without requiring moving parts or line-scanning. The sensor is mounted on a vehicle in a driving scenario in rough terrain with dynamic scenes. The captured hyperspectral data is used for terrain classification utilizing machine learning techniques. A major problem, however, is the presence of shadows in captured scenes, which degrades the classification results. We present and test methods to automatically detect shadows by taking advantage of the near-infrared (NIR) part of spectrum to build shadow maps. By utilizing these shadow maps a classifier may be able to produce better results and avoid misclassifications due to shadows. The approaches are tested on our new hand-labeled hyperspectral dataset, acquired by driving through suburban areas, with several hyperspectral snapshot-mosaic cameras.*

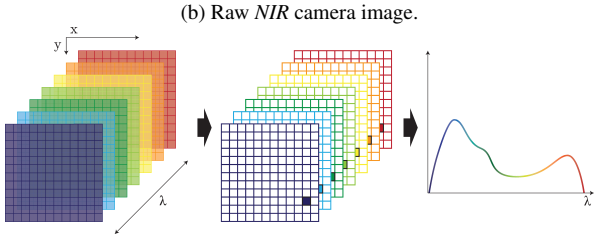
## Introduction

Environment perception and analysis is crucial for autonomous driving, especially in off-road scenarios. Given sensor data, the correct semantic interpretation of a scene is a key factor for successful autonomous navigation. The use of hyperspectral sensors brings an advantage, as it allows a more detailed view of the composition and surface of materials, plants and floor coverings than conventional cameras.

We make use of novel sensors hyperspectral snapshot mosaic sensors on unmanned land vehicles for drivability and scene analysis. Therefore, we utilize machine learning techniques to classify the captured spectral reflectances. As semantically segmenting an environment is an important goal in computer vision, it's also very complex. Even though we have seen a lot of progress in recent years by using advanced image descriptors and deep learning techniques, segmentation remains a challenge as it produces a negative impact on the accuracy and efficiency of classification results. While humans have no difficulty in performing semantic scene interpretation, computer vision systems are still struggling because of the complex interactions between light, objects and surface reflectances. Therefore, shadow detection has long been considered a crucial component of scene interpretation, but it still remains a challenging problem, particularly from a single image. For example, a dark pixel can result from either a dark surface under normal lighting conditions or a bright surface under shadows. So unwanted shadow boundaries may be detected together with true object boundaries, misleading classificators as they assume intensity or color constancy inside objects and material boundaries. Therefore, effective and robust shadow detection techniques are crucial for computer vision applications such as

(a) Raw *VIS* camera image.

(b) Raw *NIR* camera image.




(c) A schematic representation of a hypercube and an interpolated plot of a single data point.

Figure 1: Raw images from *VIS* and *NIR* cameras. And a schematic representation of a hypercube.

| Name | Method | Data |
|---|---|---|
| Levine (2005) [Levine and Bhattacharyya, 2005] | SVM | RGB |
| Lalone (2010) [Lalonde et al., 2010] | Adaboost | RGB |
| Tian (2009) [Tian et al., 2009] | Image Formation Theory | RGB |
| Mustafa (2011) [Teke et al., 2011] | NIR-R-G Space | RGB + NIR |
| Rfenacht (2014) [Rüfenacht et al., 2014] | RGB-NIR Thresholding | RGB + NIR |

Table 1: Partial overview of state of the art in shadow detection algorithms.

environment perception and scene segmentation. To determine whether a region is in the shadows, we need to compare the region with others that have the same material and orientation. Approaches using user input or multiple images have yielded impressive results, but robust and automatic shadow detectiob from a single image remains unresolved. This is because the appearance and shapes of shadows depend on several hidden factors, such as color, direction and size of light sources, the geometry of the objects throwing shadows, and the shape and material properties of the objects on which shadows are casted. In this paper we study automatic shadow detection using hyperspectral information enhancing scene segmentation results. We believe that especially data captured in the near-infrared (*NIR*) range is relevant for shadow detection and semantic segmentation.

## Related Work

The standard procedure for image-based scene-segmentation is defined by capturing regular RGB images and trying to identify different classes, like Chetan et al. [Chetan et al., 2010] did. They used color information and local binary patterns (LBP) in combination with different supervised classifiers on RGB data. But, in recent years, hyperspectral classification has gained interest. Hyperspectral data allows for a more detailed insight into the composition and nature of objects and materials like plants and soil than standard RGB data. Given hyperspectral data, the goal of classification is to assign a unique label to each pixel vector so that it is well-defined by a given class. Most supervised classifiers suffer from the Hughes effect [Hughes, 1968], especially when dealing with high-dimensional hyperspectral data. To deal with this issue, Melgani et al. [Melgani and Bruzzone, 2004] and Camps-Valls et al. [Camps-Valls and Bruzzone, 2005] introduced support vector machines with adequate kernels for hyperspectral classifications. Most image based shadow detection algorithms can be divided into two categories, semi-automatic methods which require some kind of user-input and automatic methods, which make use of additional constraints. Finlayson et al. [Finlayson et al., 2006] introduced three different shadow-free image representations and discussed methods for shadow removal utilizing the canny detector. This approach was extended by Shiting et al. [Shiting and Hong, 2013] who used clustering methods for shadow detection. Sasi et al. [Sasi and Govindan, 2015] presented a shadow detection method using a fuzzy split and merge approach which follows a top down approach of recursively splitting images into homogeneous blocks, followed by a bottom up step which merges adjacent unique regions. Wang et al. [Wang et al., 2007] combined two methods of shadow detection using color based and model based methods. It detects moving regions of provided image data and makes use of background subtraction methods which is then utilized in the proposed shadow detection approach. The table 1 displays additional work in the area of automatic shadow detection. These algorithms rely almost exclusively on RGB data and often make assumptions and have limitations. This indicates that shadow detection is still a challenge which requires additional information. The last two algorithms additionally use one band from the near infrared range to enhance shadow detection accuracy. Which indicates that the near-infrared spectrum possesses some important properties that make shadow detection more accurate. As the reflectance of surfaces is greater there. So we can better distinguish dark objects from shadowed surfaces, which makes it easier to distinguish between dark objects and real shadows.

## SENSOR SETUP

In this work we used the MQ022HG-IM-SM4X4-VIS (*VIS*) and the MQ022HG-IM-SM5X5-NIR (*NIR*) manufactured by Ximea with an image chip from IMEC [Geelen et al., 2014] utilizing a snapshot mosaic filter which has a per-pixel design. The filters are arranged in a rectangular mosaic pattern of $n$ rows and $m$ columns, which is repeated $w$ times over the width and $h$ times over the height of the sensor. These sensors are designed to work in a specific spectral range which is called the active range which is 470–620 nm for the current sensor. The *VIS* camera has a mosaic pattern with $n_{\text{VIS}} = 4, m_{\text{VIS}} = 4$. Ideally every filter has peaks centered around a defined wavelength spectrum with no response outside. However contamination is introduced into the response curve and the signal due to physical constraints. These effects can be summarized as a spectral shift, spectral leaking, and crosstalk and need to be compensated. Therefore the raw data captured by the camera needs a special preprocessing, described in [Winkens et al., 2017]. We need to construct a hypercube with spectral reflectances from the raw data. This step consists of cropping the raw-image to the valid sensor area, removing the vignette and converting to a three dimensional image, which we call a *hypercube*. Reflectance calculation is the process of extracting the reflectance signal from the captured data of an object. The purpose is to remove the influence of the sensor characteristics like quantum efficiency and the illumination source on the hyperspectral representation of objects. We define a hypercube as $\mathcal{H}\colon \mathcal{L}_x \times \mathcal{L}_y \times \mathcal{L}_\lambda \to \mathbb{R}$ where $\mathcal{L}_x$, $\mathcal{L}_y$ are the spatial domain and $\mathcal{L}_\lambda$ the spectral domain of the image. A visual interpretation of such a hypercube is displayed in figure 1c. The hypercube is understood as a volume, where each point $\mathcal{H}(x,y,\lambda)$ corresponds to a spectral reflectance. Derivated from the above definition a spectrum $\chi$ at $(x,y)$ is defined as $\mathcal{H}(x,y) = \chi$, where $\chi \in \mathbb{R}^{|\mathcal{L}_\lambda|}$ and $|\mathcal{L}_\lambda| = n \cdot m$. The image with only one wavelength, called a spectral band $\mathcal{H}(z) = \mathcal{B}_{\lambda=z}$, is defined as follows: $\mathcal{B}_\lambda : \mathcal{L}_x \times \mathcal{L}_y \to \mathbb{R}$. This image contains $\mathrm{p} = (x,y)$ the wavelength sensitivity $\lambda$ for each coordinate.

## Shadow Detection

### *Intuitive Approach*

Shadows occur when an object is in a ray of light. Thus in an image dark areas are more likely to be shadows than very bright areas. If not only a RGB image is available as an impression of a scene, but also different spectral bands, it seems intuitive to compare these bands with each other in order to draw conclusions about shadow regions. The cameras used in this work show a variety of spectra, so the 25 spectral bands captured by the *NIR* camera provide an insight into the shadow regions. One hyperpixel of the *NIR* camera thus contains 25 values for one pixel of

Raw Hyperimage (NIR) Maximum intensity image Shadow map

Figure 2: Examples of shadow detection with our intuitive approach.

the scene. Our assumption is that if all values of this hyperpixel have a low intensity, this pixel should be part of shadow region. However, if in this hyperpixel there are also values that represent a high intensity it's not likely it's a shadow region. This means that dark objects with a higher reflectance in only a few wavelength ranges are not recognized as shadows. For example, a dark green tree, which would be recognized as a shadow in the visible spectrum due to its low reflectance, has high reflectance values in the near-infrared range and would therefore not recognized as shadow. Thus, an image containing the highest value of the hyperpixel $p^{\lambda}_{(x,y)}$ for each pixel $p_{(x,y)}$ could act as a shadow candidate map:

$$p_{(x,y)} = \max_{\lambda}(p^{\lambda}_{(x,y)}) \tag{1}$$

A decision as to whether a shadow is involved can then be made utilizing a threshold value.

$$p^{bin}_{(x,y)} = \begin{cases} 1, & \text{IF } p_{(x,y)} \leq \theta \\ 0, & \text{ELSE} \end{cases} \tag{2}$$

The threshold value $\theta$ is calculated using the Otsu [Otsu, 1979] method. By applying the threshold a shadow mask is created, as it is usual in literature.

### *Adaption of Established Method*

An already existing method for shadow detection by Rüfenacht et al. is presented in [Rüfenacht et al., 2014]. It operates on an RGB image combined with a near-infrared image as input data. The figure 3 shows the process of this algorithm. First a shadow candidate map is calculated (formulas 3 to 6). Therefore the average of the three bands of the RGB image is calculated first. Then a non-linear projection is applied to the average and to the near-infrared image and then the result is inverted. The results are then multiplied with each other to form the shadow candidate map. This is shown in figure 3 as a green path. In order to refine the shadow candidate map and thus achieve an improvement on the method by Teke et al. [Teke et al., 2011], a relationship map 7 and 8) between the RGB image and the near-infrared image is created next. This ensures that shadow candidates are real shadows and not just dark objects. This is shown with the orange path in figure 3. Multiplying the shadow candidate map with the ratio map yields a shadow map (formulas 9 and 10), which is finally binarized by a threshold value. The purple path in figure 3 represents this. Since this procedure is extended to hyperspectral data in this work, some changes have to be considered. Hereinafter, the approach is adapted to suit our hyperspectral data.

**Shadow Candidates** As already mentioned, the first step is to calculate a brightness image $L^{VIS}$ of the RGB image. This can easily be extended to the 16 channels provided by the *VIS* cam.

$$L^{VIS}_{(x,y)} = \frac{\Sigma^{\lambda} p^{\lambda}_{x,y}}{\mathcal{N}_{\lambda}} \tag{3}$$

The original method had only a near-infrared image with one channel available. The *NIR* cam used in this work provides 25 channels, which must also combined to a brightness image $L^{NIR}$. This is done analogous to the *VIS* cam data. Afterwards the brightness images $L^{VIS}$ and $L^{NIR}$ with the nonlinear projection $f$ are mapped and inverted.

$$f(L) = \frac{1}{1 + e^{-\alpha(1-l^{\frac{1}{\gamma}}-\beta)}} \tag{4}$$

$$D^{VIS}_{(x,y)} = f(L^{VIS}); \quad D^{NIR}_{(x,y)} = f(L^{NIR}) \tag{5}$$

Where $D^{VIS}$ and $D^{NIR}$ define the dark maps. For the projection $f$ the value $\alpha = 14$ was recommended. Furthermore, the turning point $\beta$ is placed at $\beta = 0.5$. A $\gamma(> 1.0)$ allows to stretch the histogram in dark areas with $\gamma = 2.2$ as a recommendation. Since shadows are characterized by dark areas in both the *VIS* and *NIR* data, the shadow candidate map $D$ is now calculated as the product of the two dark maps.

$$D_{(x,y)} = D^{VIS}_{(x,y)} \cdot D^{NIR}_{(x,y)} \tag{6}$$

**Ratio Map** In the next step, the ratio between the *VIS* and *NIR* data is calculated. To calculate the ratio map $T$, the brightness image $L^{NIR}$ of the *NIR* cam and the individual channels B of the *VIS* cam are required. Each of these channels $B^{VIS}$ is divided pixel by pixel by the brightness image $L^{NIR}$.

$$t^{\lambda}_{(x,y)} = \frac{B^{\lambda}_{(x,y)}}{L^{VIS}_{(x,y)}} \tag{7}$$

After that the 16 calculated maps $T^{\lambda}$ are reduced to a ratio map $T$ containing only the highest values per pixel. However, sharing the images can result in very high values. While most values are between 0 and 1, outliers can disproportionately stretch the value range and must therefore be limited. This is done by the following function:

$$t_{x,y} = \frac{1}{\tau} \min(\max_{\lambda}(t^{\lambda}_{x,y}), \tau) \tag{8}$$

Where $\tau$ represents the upper limit of the values. In our work $\tau = 4$ has been shown to be suitable.

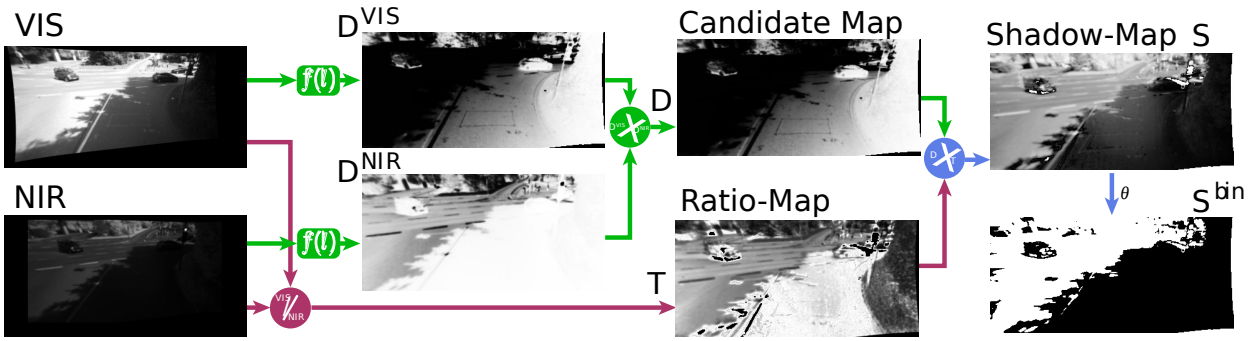


Figure 3: Step-by-step procedure for shadow detection according to [Rüfenacht et al., 2014].

**Shadow Map** To create the final shadow map $S$, the shadow candidate map $D$ and the ratio map $T$ are multiplied.

$$s_{x,y} = (1 - d_{x,y}) \cdot (1 - t_{x,y}) \tag{9}$$

Subsequently, the gray value map can be reduced to a binary shadow map by means of a threshold value method.

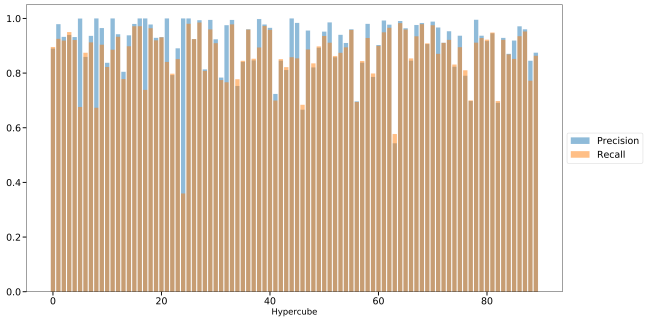
$$s^{bin}_{x,y} = \begin{cases} 1, & \text{IF } s_{x,y} \leq \theta \\ 0, & \text{ELSE} \end{cases} \tag{10}$$

## Experiments

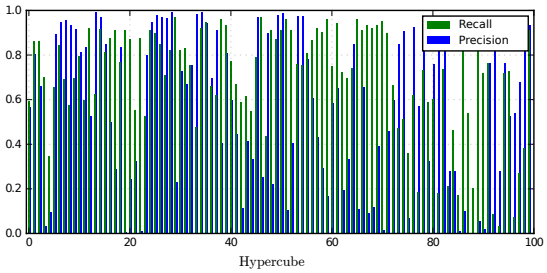
The *VIS* and *NIR* camera were mounted on a truck to collect a data set for evaluation. Both cameras are aligned parallel, forward and slightly downward to provide a good field of view over the road ahead. LM5JC10M 2/3 lenses from Kowa were mounted on the cameras to enlarge the field of view. These provide an angle of view of 82.2° horizontally, 66.5° vertically, and 95.4° diagonally. The dataset selected for the experiments consists of a total of 100 hypercubes of the *VIS* and 100 hypercubes of the *NIR* cam, which were synchronized using a hardware trigger. The individual hypercubes were annotated by hand and in each hypercube the shadow areas were annotated. Two methods for the detection of shadows have previously been presented, which are now being compared. Each method is applied to the available dataset and provides a binary shadow map that is compared to the ground truth. The annotations from the dataset are also extracted as a binary shadow map. Some examples of shadow detection using our method are displayed in figure 5. The results of both methods are displayed in figure 4. Based on our dataset, we have achieved a precision of 90% percent and an recall of about 86% with our intuitive method. Compared to our approach, the method from Ruefenacht et al. leads to varying results on our dataset, achieving a precision of 58% with a recall of 71%. However, we must keep in mind that this method was developed based on other data. Therefore one cannot generally say that our method is better than the established methods as the database is very different. But our method delivers good results based on the hyperspectral data generated by these new type cameras.

## Conclusions

We investigate the use of hyperspectral data for autonomous driving and scene analysis. For this purpose we use new sensor technology which can record a complete hypercube with 16 or 25 channels in one image frame at one discrete point in time. In this work we presented a simple algorithm for shadow detection.



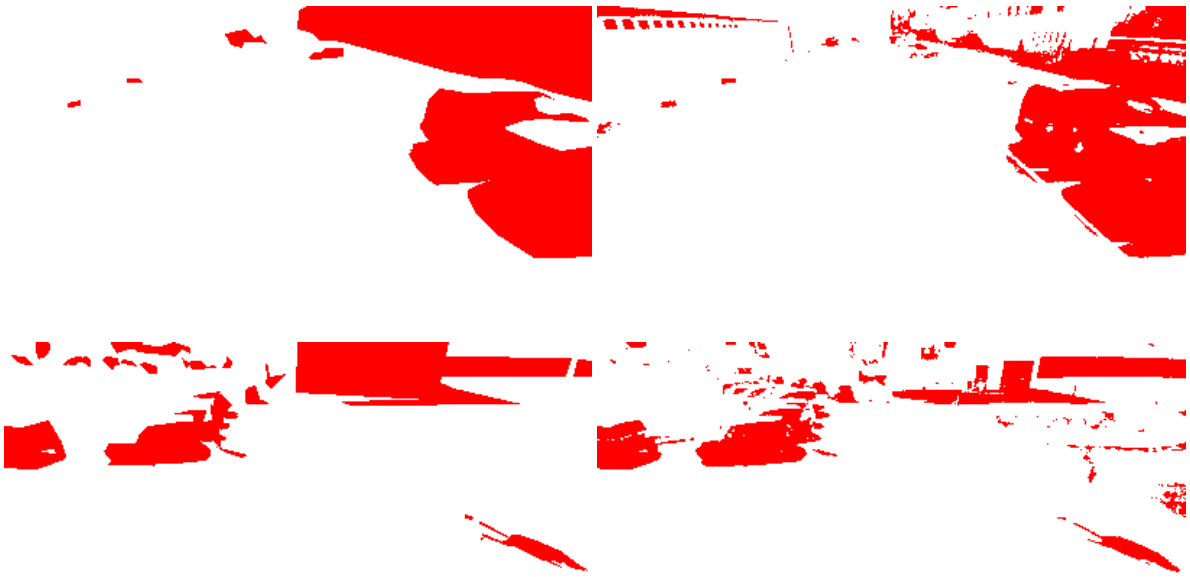
(a) Results using our approach.



(b) Results for [Rüfenacht et al., 2014].

Figure 4: Results of both methods on our own dataset.

It is a simple and intuitive approach that detects shadows in a hypercube with maximum values of spectral channels per pixel position. This approach provided suitable results on our dataset and performed better than an established method, which we have adapted in order to be used with our data. These are only the first results and we will continue to work on this topic. In the future we will further investigate the simultaneous use of camera data from both cameras and try to make shadow detection more stable. Furthermore we will try to integrate the shadow information into the classification process.



Ground truth Prediction

Figure 5: Some results of our shadow detection approach. The ground truth is displayed on the left an the prediction on the left.

## References


[Camps-Valls and Bruzzone, 2005] Camps-Valls, G. and Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1351–1362.

[Chetan et al., 2010] Chetan, J., Krishna, M., and Jawahar, C. (2010). Fast and spatially-smooth terrain classification using monocular camera. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 4060–4063. IEEE.

[Finlayson et al., 2006] Finlayson, G. D., Hordley, S. D., Lu, C., and Drew, M. S. (2006). On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):59–68.

[Geelen et al., 2014] Geelen, B., Tack, N., and Lambrechts, A. (2014). A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic. In *Spie Moems-Mems*, pages 89740L–89740L. International Society for Optics and Photonics.

[Hughes, 1968] Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63.

[Lalonde et al., 2010] Lalonde, J.-F., Efros, A. A., and Narasimhan, S. G. (2010). Detecting ground shadows in outdoor consumer photographs. In *European conference on computer vision*, pages 322–335. Springer.

[Levine and Bhattacharyya, 2005] Levine, M. D. and Bhattacharyya, J. (2005). Removing shadows. *Pattern Recognition Letters*, 26(3):251–265.

[Melgani and Bruzzone, 2004] Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8):1778–1790.

[Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

[Rüfenacht et al., 2014] Rüfenacht, D., Fredembach, C., and Süsstrunk, S. (2014). Automatic and accurate shadow detection using near-infrared information. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1672–1678.

[Sasi and Govindan, 2015] Sasi, R. K. and Govindan, V. (2015). Fuzzy split and merge for shadow detection. *Egyptian Informatics Journal*, 16(1):29–35.

[Shiting and Hong, 2013] Shiting, W. and Hong, Z. (2013). Clustering-based shadow edge detection in a single color image. In *Mechatronic Sciences, Electric Engineering and Computer (MEC), Proceedings 2013 International Conference on*, pages 1038–1041. IEEE.

[Teke et al., 2011] Teke, M., Başeski, E., Ok, A. Ö., Yüksel, B., and Şenaras, Ç. (2011). Multi-spectral false color shadow detection. In *Photogrammetric Image Analysis*, pages 109–119. Springer.

[Tian et al., 2009] Tian, J., Sun, J., and Tang, Y. (2009). Tricolor attenuation model for shadow detection. *IEEE Transactions on image processing*, 18(10):2355–2363.

[Wang et al., 2007] Wang, S.-K., Qin, B., Fang, Z.-H., and Ma, Z.-S. (2007). Fast shadow detection according to the moving region. In *Machine Learning and Cybernetics, 2007 International Conference on*, volume 3, pages 1590–1595. IEEE.

[Winkens et al., 2017] Winkens, C., Sattler, F., Adams, V., and Paulus, D. (2017). Vorverarbeitung hyperspektraler bilddaten von snapshot-mosaik kameras. In *23. Workshop Farbbildverarbeitung*. Der Andere Verlag.

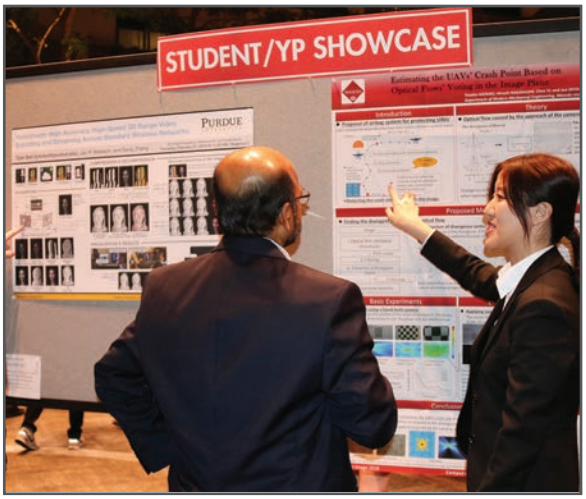
# JOIN US AT THE NEXT EI!

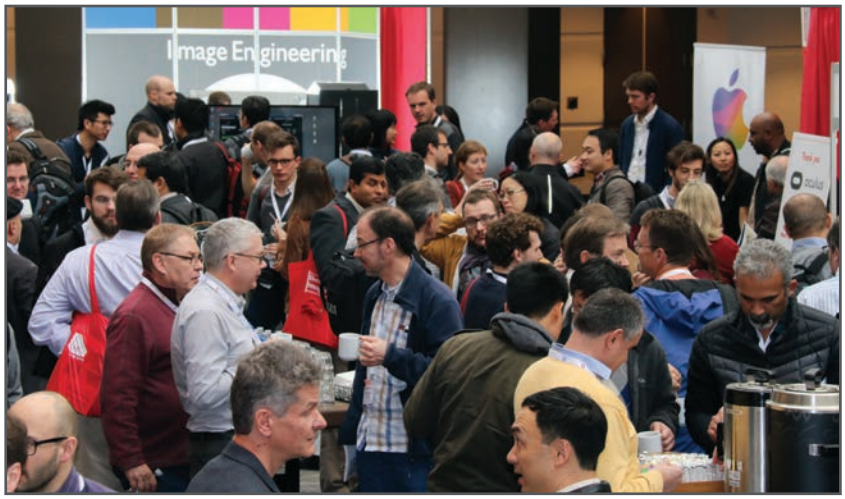
IS&T International Symposium on

# Electronic Imaging

SCIENCE AND TECHNOLOGY

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





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

**www.electronicimaging.org**

IS&T
imaging.org