

Material-Based Object Segmentation Using Near-Infrared Information

N. Salamati and S. Süsstrunk

School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
`{neda.salamati, sabine.susstrunk}@epfl.ch`

Abstract

We present a framework to incorporate near-infrared (NIR) information into algorithms to better segment objects by isolating material boundaries from color and shadow edges. Most segmentation algorithms assign individual regions to parts of the object that are colorized differently. Similarly, the presence of shadows and thus large changes in image intensities across objects can also result in mis-segmentation. We first form an intrinsic image from the R, G, B, and NIR channels based on a 4-sensor camera calibration model that is invariant to shadows. The regions obtained by the segmentation algorithms are thus only due to color and material changes and are independent of the illumination. Additionally, we also segment the NIR channel only. Near-infrared (NIR) image intensities are largely dependent on the chemistry of the material and have no general correlation with visible color information. Consequently, the NIR segmentation only highlights material and lighting changes. The union of both segmentations obtained from the intrinsic and NIR images results in image partitions that are only based on material changes and not on color or shadows. Experiments show that the proposed method provides good object-based segmentation results on diverse images.

KEYWORDS

Object Segmentation, Near-Infrared Imaging, Material Classification, Illuminant Invariant Images, Mean Shift Segmentation.

INTRODUCTION

Object segmentation is important to many computer vision tasks and can facilitate image classification and image enhancement frameworks. However, the diversity in image content and lighting conditions of natural scenes makes accurate object segmentation a challenging task [1, 2, 3]. Inhomogeneities, such as shadows, specularities, changes of color within the object and variation in pigment density will introduce gradients in the image that can confound segmentation algorithms, resulting in multiple distinct segments being assigned to one single object (see Figure 1b).

The goal of this paper is to determine where in images material changes occur, i.e. to segment the image so that the segment boundaries correspond only to object boundaries. We want to avoid mis-segmentation due to variations in surface color and shadows. To do that, we incorporate a fourth image channel, which contains the near-infrared (NIR) scene information, into our segmentation framework. NIR radiation ($700\text{nm} - 1100\text{nm}$) generally penetrates deeper into an object's surface and can

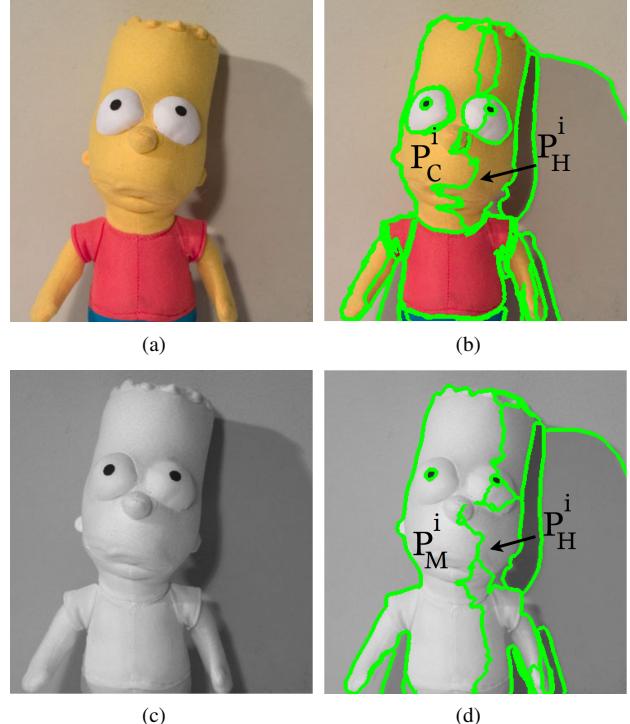


Figure 1: The mean shift segmentation result of visible (I_{RGB}) and NIR images (I_{NIR}). The first row shows the color image and its segmentation result, the second row is the NIR image and its segmentation. Note the oversegmentation resulting from changes in illumination P_H^i or colors P_C^i within the object.

reveal the underlying material characteristics [4, 5]. As such, changes in intensity in the NIR image are due to material and illumination changes, but not to color variations within the same material.

Another source of mis-segmentation are shadows, which occur due to the shape of the object and/or the geometric arrangement of the object and the light source. Many algorithms have been proposed to correct the color within the objects so the edges corresponding to the shadows are not confounded with the object boundaries [6, 7]. Some shadow removal frameworks try to recover an image based on ratios of color bands, in which the absolute intensity variation over an object is reduced so that the result is invariant to shadows [8]. Inspired by the 4-sensor camera calibration model by Finlayson and Drew [9], we combine both visible RGB and NIR images to obtain an intrinsic image that

is independent of illumination. Different pixel values represent reflectance variations, thus color and material changes, but are shadow independent. The union of the NIR and intrinsic image segmentations results in segments that are only bound by material changes, but not by color and shadow variations within the object. Other image segmentation methods using data beyond the visible spectrum is primarily applied in remote sensing applications [10]. The hyper-spectral images are captured with visible/infrared spectrometers that measure the spectral reflectance of each surface in the scene. To reduce the dimensionality of the original dataset, feature extraction is done for each material. A classifier is then applied on the reduced-dimension dataset [11]. Although this scheme is appropriate for segmenting the hyper-spectral data in remote sensing applications, it requires having a high spectral resolution imaging spectrometer system and prior knowledge of the material characteristics that need to be classified. In our approach, we only use four channels, R , G , B , and NIR, and make no a-priori assumption of the materials we encounter.

In Section 2 we address the scope of the problem and the approach to the solution. Section 3 discusses the physics of the possible relations between two patches within an object and the properties of the color ratios across boundaries. Section 4 focuses on how we can form the images that are independent of illumination. Section 5 presents the segmentation framework and in Section 6 results are shown and discussed. Conclusions are presented in Section 7.

THE PROPOSED APPROACH

Fig. 1 shows the segmentation result of a visible (I_{RGB}) and its corresponding NIR image (I_{NIR}). As illustrated in Fig. 1 (b), different segments in the visible image $\{P^i_v \mid \bigcup_{i=1}^N P^i_v = I_{RGB}\}$ correspond to different parts of the object with different colors as well as spatial inhomogeneities resulting from shadows and shadings. Thus,

$$P_v \subset P_C \cup P_H \quad (1)$$

where P_C is the set of all the segments with different colors and P_H is the set of the segments in which the illumination is different. In NIR images, on the other hand, changes in material, shadings, and cast shadows are responsible for the different segments $\{P^i_n \mid \bigcup_{i=1}^M P^i_n = I_{NIR}\}$ (see Fig. 1 (d) for illustration). Thus,

$$P_n \subset P_M \cup P_H \quad (2)$$

where P_M is the set of all the segments with different materials. To get an accurate segmentation result, we need to eliminate different segments that are generated due to different illumination conditions or varying colors within the same material. Thus, an intrinsic image I_{int} , in which each segment $\{P^i_{int} \mid \bigcup_{i=1}^L P^i_{int} = I_{int}\}$ belongs to either P_M or P_C includes significant information. Incorporating this information along with NIR images, one can simply derive a segmented object by applying the " \cap " operator.

$$P_{int} \subset P_C \cup P_M \implies P_{int} \cap P_n \subset P_M \quad (3)$$

where P_{int} and P_n are the segments of the intrinsic and NIR images, respectively.

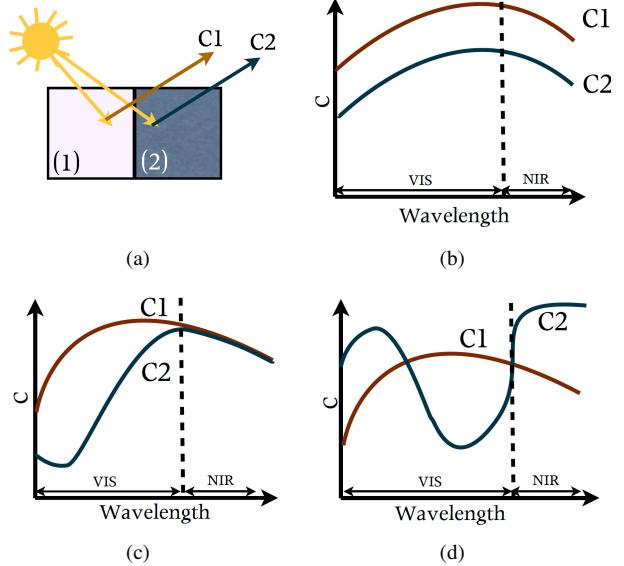


Figure 2: Three different relations that can hold between the color/NIR signals ($C(\lambda)$) of region (1) and (2) in the visible and NIR part of the spectrum. (a) Two different regions, (b) part (2) is under shadow, (c) (1) and (2) are of the same material but colored differently, and (d) a color and material change occurs.

THE PHYSICAL PROPERTIES OF VISIBLE AND NIR SIGNALS

Visible and NIR image intensities depend on the interaction between the surface properties of the object, illuminants, and the camera. The sensor response I_k of a sensor $k : k \in \{R, G, B, NIR\}$ with sensitivity R_k is therefore expressed as:

$$I_k = \int_{\lambda=400}^{1100} S(\lambda) \times E(\lambda) \times R_k(\lambda) d\lambda \quad (4)$$

where $S(\lambda)$ is the reflectance of the surface and $E(\lambda)$ is the illuminant spectral power distribution. Fig. 2 depicts two different parts, (1) and (2), of an image, and three different relations that can hold between the color/NIR signals ($C(\lambda)$) of these regions.

$$C(\lambda) = S(\lambda) \times E(\lambda) \quad (5)$$

Depending on the location of the object with respect to the camera and the light source, a shadow can be cast. This shadow results in a reduction of measured intensity. We can describe the image intensity in the lit and shadow part of the object I_k^{lit} , I_k^{shade} as follows (see Fig. 2 (b) for illustration). If

$$I_k^{lit} = \int_{\lambda=400}^{1100} S(\lambda) \times E(\lambda) \times R_k(\lambda) d\lambda \quad (6)$$

then the shadow part of that object can be described as:

$$I_k^{shade} = \int_{\lambda=400}^{1100} S(\lambda) \times aE(\lambda) \times R_k(\lambda) d\lambda \quad (7)$$

where a represents a fraction of the light intensity ($0 \leq a \leq 1$). We assume that the above statement is true for both VIS and NIR

images, i.e., the ratio of the I_{RGB} to the I_{NIR} response across a material with certain color stays unaffected by changes in the illumination. Thus, the following relations can be formulated:

$$[I^{shade}_R, I^{shade}_G, I^{shade}_B] = a[I^{lit}_R, I^{lit}_G, I^{lit}_B] \quad \text{and} \\ I^{shade}_{NIR} = aI^{lit}_{NIR} \\ \Rightarrow \frac{[I^{lit}_R, I^{lit}_G, I^{lit}_B]}{I^{lit}_{NIR}} = \frac{[I^{shade}_R, I^{shade}_G, I^{shade}_B]}{I^{shade}_{NIR}} \quad (8)$$

Although this assumption does not hold in general, as the ambient illumination is not accounted for, it has been applied in other color correction models [12].

The second relation is when both regions belong to the same material, but are colorized differently (see Fig. 2 (c) for illustration). NIR imaging is ‘transparent’ to a number of colorants and dyes; it can see through the first layer to reveal the material surface underneath [4]. Thus, the NIR images reveal more about the material itself rather than its color.

The chemistry and the process of colorizing different materials makes each class of material have an affinity towards a certain class of colorants (chemistry and functional bond specific). Hence, even if the object colors are not transparent to the NIR, the NIR response is more probable to be the same ($I^{(1)}_{NIR} = I^{(2)}_{NIR}$) [6], because it is very likely that different colorants applied to colorizing the material came from the same chemistry. Consequently,

$$\frac{[I^{(1)}_R, I^{(1)}_G, I^{(1)}_B]}{I^{(1)}_{NIR}} \neq \frac{[I^{(2)}_R, I^{(2)}_G, I^{(2)}_B]}{I^{(2)}_{NIR}} \quad (9)$$

This assumption does not always hold, specifically when the color is too dark.

Fig. 2 (d) shows the change in color and material in two regions. In this case, the ratio of I_{RGB} to I_{NIR} is not constant for the two patches (see equation 9).

FORMING THE INTRINSIC IMAGE

Up to now, we have argued that R , G , and B to NIR ratio images are potentially able to present changes that correspond to either different materials or different colors within that material. Inspired by the physics of the NIR and color signal of a surface, we modify the algorithm by Finalyson and Drew [9]. This algorithm tries to find the coordinates in which the ratio image is invariant to both intensity and color of the illuminant. The color constancy at a pixel algorithm is based on the assumptions that Planck’s black body equation models the illuminant spectra and the sensors’ spectral response can be modeled by the Dirac delta function. With these assumptions, the logarithmic response of sensor k for an illuminant $E(\lambda, T)$ is:

$$\log(I_k) = \log(E(\lambda_k, T)S(\lambda_k)), \quad E(\lambda_k, T) = \simeq K_1 \lambda^{-5} e^{-\frac{K_2}{T\lambda_k}} \\ \log(I_k) = -\frac{1}{T} \underbrace{\left(\frac{K_2}{\lambda_k} \right)}_{E_k} + \underbrace{\log(K_1 \lambda^{-5} S(\lambda_k))}_{S_k} \quad (10)$$

where λ_k is the wavelength to which sensor k is sensitive to, S_k is the reflectance of the surface being imaged at wavelength λ_k , T

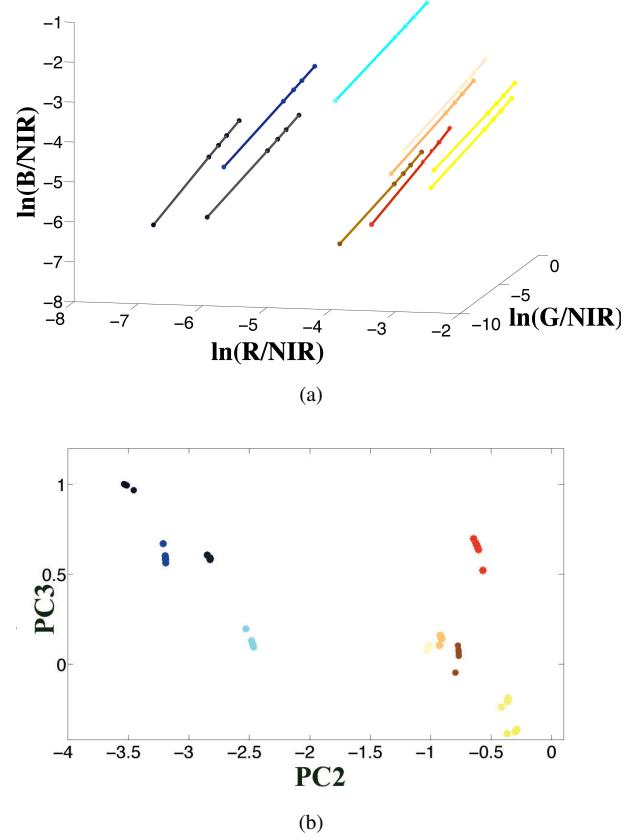


Figure 3: (a) The log-ratio of 10 samples under different light sources/shadows. The intensity ratio of all the samples under different lights lies along a single direction, (b) The chromaticity space given by the projection onto the second and third principle eigenvectors.

is the color temperature of the light, and K_1 and K_2 are constants. The first term in this equation E_k depends on the illuminant’s color temperature and the last part S_k depends on the surface reflectance. Given 4 sensors $k \in \{k_1, k_2, k_3, k_4\}$, subtracting the response of one logarithmic sensor from those of the other 3 sensors gives us the equation of a line in 3-dimensional space in which the reflectance dependent part appears as the intercept and the illuminant dependent part is the slope of the line.

$$\log\left(\frac{I_{k_1}}{I_{k_4}}\right) = \log(I_{k_1}) - \log(I_{k_4}) = S_{k_1} - S_{k_4} - \frac{1}{T}(E_{k_1} - E_{k_4}) \\ \log\left(\frac{I_{k_2}}{I_{k_4}}\right) = \log(I_{k_2}) - \log(I_{k_4}) = S_{k_2} - S_{k_4} - \frac{1}{T}(E_{k_2} - E_{k_4}) \\ \log\left(\frac{I_{k_3}}{I_{k_4}}\right) = \log(I_{k_3}) - \log(I_{k_4}) = S_{k_3} - S_{k_4} - \frac{1}{T}(E_{k_3} - E_{k_4}) \quad (11)$$

Thus, adjusting the color temperature of the light source T changes the log-ratio of the sensor responses along a single direction, on which the location of the sample’s log-ratio depends on T . The 3-dimensional space can be projected onto a 2-dimensional

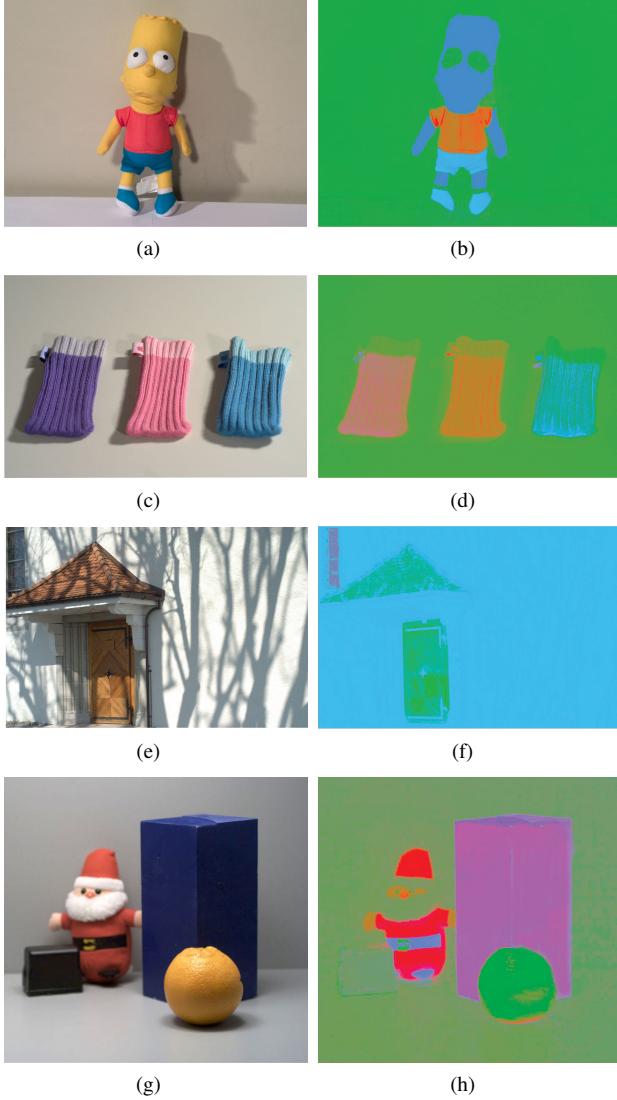


Figure 4: (First Column) visible images and (second column) the illuminant-independent representation. To visualize images in the new space, we present PC_2 and PC_3 as a and b values in the CIELAB color space. Lightness value is chosen to be 60 for all the intrinsic images.

space where illuminant induced variation is minimized, i.e., the new 2-dimensional representation of any image will be independent of the illuminant.

To employ this framework's ability to extract descriptors, which are independent of the lightning condition, we use R , G , B and NIR camera responses ($R, G, B = k_1, k_2, k_3$ and $NIR = k_4$). In order to find the new coordinates, pertinent to our camera, to map the ratio image onto the corresponding illumination invariant image, we measured the reflectances of 50 objects in the range of the visible and NIR spectrum. Samples' RGB and NIR intensities are calculated under different Planckian light sources with the temperature 3000, 5000, 6000, 6500 as well as Equi-Energy.

The log-ratio, R_{IR} , B_{IR} and G_{IR} , are computed as follows:

$$R_{IR} = \ln\left(\frac{R}{NIR}\right), \quad G_{IR} = \ln\left(\frac{G}{NIR}\right), \quad B_{IR} = \ln\left(\frac{B}{NIR}\right) \quad (12)$$

Fig. 3 (a) shows the log ratio of 10 samples under 6 different illuminants. All the intensity ratios of the samples under different lights roughly lie along a single direction. For all the samples under different light sources, the covariance matrix can be computed as:

$$\text{cov} = \begin{bmatrix} \text{cov}(R, R) & \text{cov}(R, G) & \text{cov}(R, B) \\ \text{cov}(G, R) & \text{cov}(G, G) & \text{cov}(G, B) \\ \text{cov}(B, R) & \text{cov}(B, G) & \text{cov}(B, B) \end{bmatrix} \quad (13)$$

The best coordinates can be calculated as the overall eigenvectors of the covariance matrix. For our database, the eigenvectors are:

$$\mathbf{C} = \begin{bmatrix} 0.378 & 0.89 & 0.23 \\ 0.54 & 0.00 & -0.84 \\ 0.75 & -0.44 & 0.49 \end{bmatrix}$$

Samples' log-ratio are projected onto the two eigenvectors with smaller eigenvalues using the following equation:

$$\begin{bmatrix} PC_2 \\ PC_3 \end{bmatrix} = \begin{bmatrix} 0.89 & 0.00 & -0.44 \\ 0.23 & -0.84 & 0.49 \end{bmatrix} \times \begin{bmatrix} R_{IR} \\ G_{IR} \\ B_{IR} \end{bmatrix} \quad (14)$$

Fig. 3 (b) shows the samples in the database in the new space. In this space, each sample under a specific light source appears as a dot and that sample under other light sources approximately project to the same position.

To investigate the algorithm on real images, the log-ratio image I_{int} is calculated and projected on the "illuminant-independent" coordinates (applying Eq. 12-14 at each pixel position). Fig. 4 shows the illuminant-independent representation of some images. Fig 4 (c) and (e) are taken under an unknown illuminant. Their results, however, are fairly invariant to the light source's intensity. This can be explained by the object reflectances in the NIR part of the spectrum, as discussed in Section 3.

The primary drawback associated with this approach is its inability to differentiate dark plastic objects situated close to brighter objects (as illustrated in Fig. 4 (h) where a black object is placed in front of the grey background or the white parts of the doll). Carbon black is used as a pigment in rubber and dark plastic products (polymers in general). This pigment reflects almost no light in the visible or the NIR part of the spectrum and appears dark in the visible and NIR images. Thus, the shadow relation can hold between the black part and the brighter grey parts, and these two parts will be mapped onto the same value.

THE SEGMENTATION PROCEDURE

The idea is to segment the illuminant-independent images as well as the NIR images. As it has already been formulated in Eq. 1-2, segments P_n in the NIR images are formed due to changes in material or illuminant and segments P_m in the intrinsic images are formed due to changes in material or color. Thus, logically, the physical object boundaries are the ones present in both images. To segment the images, the mean shift algorithm is applied to

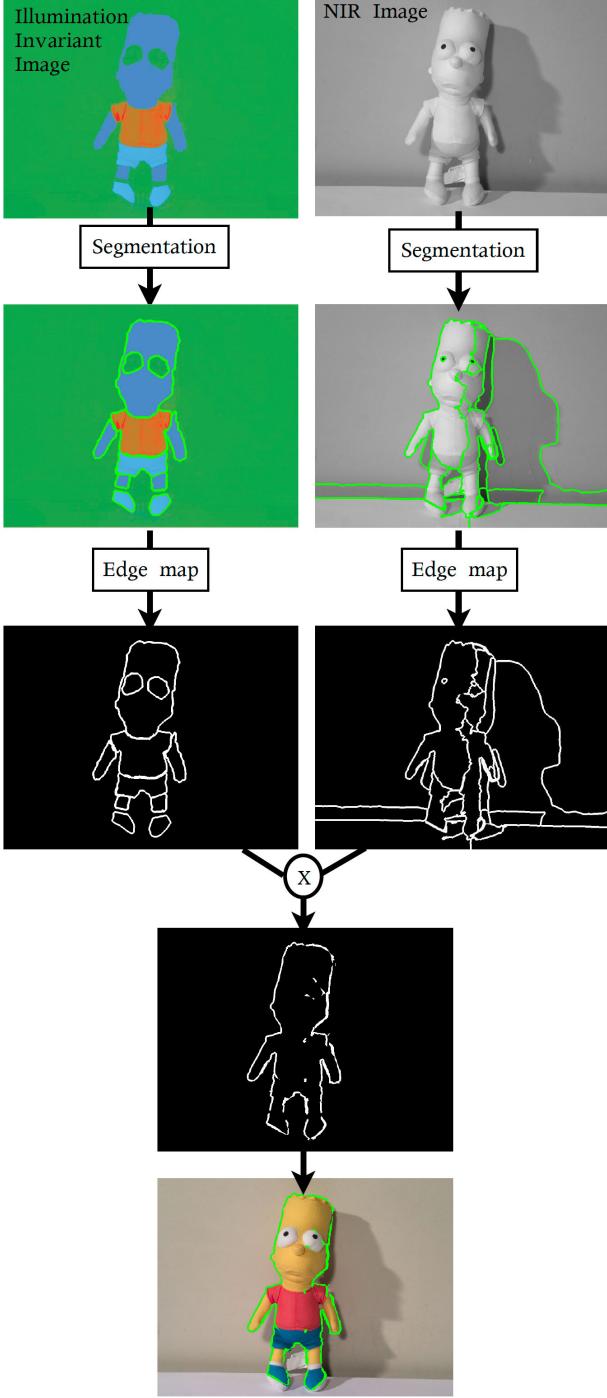


Figure 5: The flowchart detailing the segmentation framework.

both intrinsic and NIR images. Mean shift is an image clustering method based on color and spatial features [13]. The main idea behind the algorithm is to compute for every single pixel a series of mean values in feature space. The mean is shifted towards more densely populated regions in the feature space. Each segment contains all data points in the attraction basin of a convergence point. This approach does not require a priori knowledge of the number of segments.

The feature space for segmenting NIR images was chosen to be the pixel intensity, and for illuminant-invariant images, $PC2$ and $PC3$ coordinates formed the feature vector.

A key feature to the implementation is that NIR images are needed to be a bit over-segmented so that detecting all the boundaries corresponding to the material changes is guaranteed. Thus, for the NIR image segmentation, the resolution parameters as the input of the mean shift algorithm were chosen 10% larger than the ones for the intrinsic images.

After segmentation, the boundaries of each segment formed a binary edge map. Since the resolution parameters of the segmentation algorithm are different for the intrinsic and NIR images, corresponding segments' edges of these two images may not intersect. Thus, the binary edge map was dilated by a structural element of size 3.

The segmented object is the result of applying the " \cap " operator on the edge map of the segmented NIR and illumination-inavriant images. Fig. 5 exhibits the flowchart detailing the segmentation framework.

RESULTS

All the images were photographed in the visible and in the NIR range of the spectrum. The camera we used is a modified Canon EOS 300D [14].

We compare the mean shift visible/NIR segmentation with mean shift on visible images only. The results are shown in Fig. 6. This comparison provides useful insights into how much more accurately our segmentation procedure is able to predict the physical object's boundaries.

Comparing the result with the visible-only segmentation, one can notice that regions in which a small gradient of illumination or color exists were outlined as a single region. For instance, the orange in Fig. 6 (e) and (i) is divided into different segments, using just visible information, because of the changes in illumination, while in Fig. 6 (f) and (j) the actual physical boundary of the orange is detected as a single object.

As can be seen in Fig. 6 (i) and (j) on the green object on the wall, precision in object boundary retrieval is higher using the proposed framework. Variation of the illumination across the scene makes visible-only segmentation results more sensitive to the resolution parameters in the mean shift algorithm. In visible-only segmentation there is always a trade-off between the resolution value and the segments detected within an object. In other words, when the resolution parameter is increased many segments are identified within the object (i.e. the object is over-segmented) but the object boundaries are detected with precision. At a lower resolution, however, the segments within the object are merged and hence fewer segments are identified, risking the detection of the actual boundary of the object (i.e. the tail of the object may be identified as a part of the background). Our approach, on the other hand, results in more accurate object boundaries, as we can increase the resolution parameter to obtain the exact boundaries in both NIR and intrinsic image segmentations. By applying the " \cap " operator, all the undesired segments are removed.

The drawback of our approach is losing the dark plastic objects in segmentation when they have a similar boundary as a lighter object (see Fig.6 (f)). Another drawback of this method is the

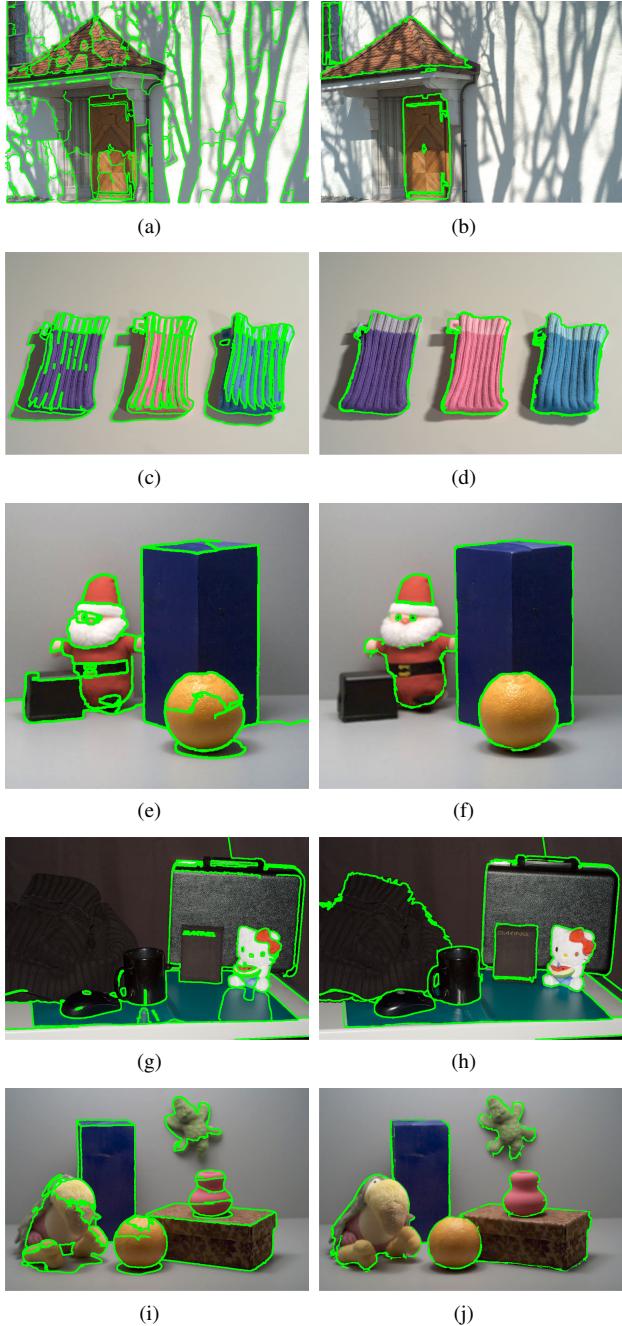


Figure 6: (First Column) visible-only image segmentation result, (second column) segmentation result using joint information.

existence of some edges in the result that do not correspond to any changes of material (see Fig.5: the segmentation result of the doll around the eye and the collar). This mis-segmentation occurs when there are many variations in the channel intensities that are due to the illumination. In this case there will be so many edges in the NIR segmentation corresponding to the changes of illumination that may intersect with the changes of color in the intrinsic image and form new segments.

CONCLUSION

We have presented a method that accurately detects physical object boundaries in images using visible RGB and NIR information. In order to discard the segments corresponding to color changes within an object, we propose to use NIR image segmentation only. By combining the NIR information as the forth channel along with *RGB* values to form an illuminant independent image, we can achieve a shadow-free representation of the scene. The union of the two segmentation results produce segments which are only material dependent. By applying the proposed framework on real images, we show that segmentation using NIR information as well as visible images yields more accurate results in detecting physical object boundaries.

Acknowledgment

The work presented in this paper was supported by the Swiss National Science Foundation under grant number 200021-124796/1.

References

- [1] Z. Li, P. Jiang, H. Ma, J. Yang, and D. Tang. A model for dynamic object segmentation with kernel density estimation based on gradient features. *Image and Vision Computing* vol. 27, pp. 817-823, 2009.
- [2] M. Mushrif, and A. Ray. Color image segmentation: Rough-set theoretic approach. *IPattern Recognition Letters* vol. 29, pp. 483-493, 2008.
- [3] Y. Deng, B. Manjunath, and H. Shin. Color image segmentation. *In Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)* vol. 11, pp. 71-82, Fort Collins, 1999.
- [4] N. Salamat, C. Fredembach, and S. Süsstrunk. Material classification using color and NIR images. *In Proc. of the IS& T/SID Color Imaging Conference* New Mexico, 2009.
- [5] D.A. Burns, and E.W. Ciurczak. *Handbook of near-infrared analysis* Marceld Ekkerin,Inc., 2001.
- [6] M.H. Brill. Image segmentation by object color: a unifying framework and connection to color constancy. *Optical Society of America* vol. 7 (10), pp. 2041-2047, 1990.
- [7] G.D. Finlayson, S.D. Hordley, M. S. Drew, and C. Lu. On the removal of shadows from images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* vol. 28 (1), pp. 59-68, 2006.
- [8] B.V. Funt, and G.D. Finlayson. Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence* vol. 17(5), pp. 522-529, 1995.
- [9] G.D. Finlayson, and M.S. Drew. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. *in Proc. of International Conference on Computer Vision (ICCV)* Vancouver, 2001.
- [10] C. Pohl, and J. L. Van Genderen. Review article multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing* vol. 19 (5), pp. 823854, 1998.
- [11] X. Wang, B. Waske, and J.A. Benediktsson. Ensemble methods for spectral-spatial classification of urban hyperspectral data. *Geoscience and Remote Sensing Symposium* vol. 4, pp. 944-947, 2010.
- [12] M. D. Levine, and J. Bhattacharyya. Removing shadows. *Pattern Recognition Letters* vol. 26 (3), pp. 251-265, 2005.
- [13] D. Comanicu, and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence* vol. 24 (5), pp. 603-619, 2002.
- [14] C. Fredembach, and S. Süsstrunk. Colouring the near infrared. *In Proc. of the IS& T/SID Color Imaging Conference* Portland, 2008.

Author Biography

Neda Salamat received her BSc and MSc in Chemical (Textile) Engineering from Tehran Polytechnic University. Since 2008, she is a Research Assistant in Color Imaging, pursuing a PhD degree in the Audio-visual Communications Laboratory, EPFL.