

Article

Hybrid Fusion-Based Background Segmentation in Multispectral Polarimetric Imagery

Md Nazrul Islam , Murat Tahtali and Mark Pickering

School of Engineering and Information Technology, The University of New South Wales (UNSW@ADFA), Canberra, ACT 2610, Australia; m.tahtali@adfa.edu.au (M.T.); m.pickering@adfa.edu.au (M.P.)

* Correspondence: md.nazrul.islam@student.unsw.edu.au; Tel.: +61-401-952-817

Received: 9 May 2020; Accepted: 27 May 2020; Published: 1 June 2020



Abstract: Multispectral Polarimetric Imagery (MSPI) contains significant information about an object's distribution, shape, shading, texture and roughness features which can distinguish between foreground and background in a complex scene. Due to spectral signatures being limited to material properties, Background Segmentation (BS) is a difficult task when there are shadows, illumination and clutter in a scene. In this work, we propose a two-fold BS approach: multiband image fusion and polarimetric BS. Firstly, considering that the background in a scene is polarized by nature, the spectral reflectance and correlations and the textural features of MSPI are calculated and analyzed to demonstrate the fusion significance. After that, integrating Principal Component Analysis (PCA) with Fast Fourier Transform (FFT), a hybrid fusion technique is proposed to show the multiband fusion effectiveness. Secondly, utilizing the Stokes vector, polarimetric components are calculated to separate a complex scene's background from its foreground by constructing four significant foreground masks. An intensity-invariant mask is built by differentiating between the median filtering versions of unpolarized and polarized images. A strongly unpolarized foreground mask is also constructed in two different ways, through analyzing the Angle of Linear Polarization (AoLP) and Degree of Linear Polarization (DoLP). Moreover, a strongly polarized mask and a strong light intensity mask are also calculated based on the azimuth angle and the total light intensity. Finally, all these masks are combined, and a morphological operation is applied to segment the final background area of a scene. The proposed two-fold BS algorithm is evaluated using distinct statistical measurements and compared with well-known fusion methods and BS methods highlighted in this paper. The experimental results demonstrate that the proposed hybrid fusion method significantly improves multiband fusion quality. Furthermore, the proposed polarimetric BS approach also improves the mean accuracy, geometric mean and F1-score to 0.95, 0.93 and 0.97, respectively, for scenes in the MSPI dataset compared with those obtained from the methods in the literature considered in this paper. Future work will investigate mixed polarized and unpolarized BS in the MSPI dataset with specular reflection.

Keywords: polarimetric imagery; spectral reflectance; multiband fusion; background segmentation; statistical measurement

1. Introduction

The emerging significance of Multispectral Polarimetric Imagery (MSPI) has been actively pursued in diverse applications over the last few decades. Using advanced communication tools and techniques, multispectral images are currently composed of several sub-bands from both the visible and near infrared (NIR) wavelengths [1], with the latter material-dependent and capable of penetrating deeper into materials. Potential applications could investigate acquiring an imaging system that performs image denoising [2], image dehazing [3] and semantic segmentation [4]. Multispectral imaging is a mode commonly reported in the literature for enhancing color reproduction [5], illuminant estimation [6],

vegetation phenology [7,8], shadow detection [9] and background subtraction [10]. However, based on only multispectral information, it is sometimes unfeasible to segment the background from a complex scene, whereas polarimetric imagery is shown to extract meaningful information about surface features, shapes, shading and smoothness in optical sensing images [11]. Specific photoreceptors are responsible for polarized light vision, a phenomenon used by polarimetric imaging techniques in diverse applications, such as specular and diffuse separation [12], material classification [13], shape estimation [14], target detection [15–17], anomaly detection [18], man-made object separation [19] and camouflaged object separation [20].

Currently, computer vision-based Background Segmentation (BS) has potential application for shape detection, activity detection and recognition, behavioral understanding, surveillance, medical care and parenting [21]. However, an error in object detection can degrade the overall performance of a system. In terms of surveillance and monitoring systems, the foreground in a scene is described as a disruptor that breaks up the background. Existing approaches [22,23] based on this criterion first look for a region of interest to segment the background from a scene, with their performances reported to depend on the contrast between an object and its background. Although a frame-differencing approach may sometimes be beneficial for BS, selecting an appropriate threshold is a challenging task due to the foreground and background sometimes being mixed in a complex scene. Given fast processing power and a noise-handling capacity, an average image of a series of N bands can be calculated to predict a difference image. However, challenges, including low lighting conditions, illumination, occluded foreground and shadow effects, which can be camouflaged in an imaging modality, including the visible light spectrum, need to be considered.

Although integrating data captured from multiple spectra followed by different polarization filter orientations benefits a BS task, registering multispectral acquisitions is more challenging [24]. Combining six bands of multispectral images in a spectrum (420–1000 nm) and four different orientations of linear polarization filters (0, 45, 90, 135) can improve the accuracy of a BS technique. In this paper, to address the problem of the dissimilarity of multiband information, a multiband fusion framework with an algorithm is proposed. Firstly, the fusion algorithm decomposes each band into low- and high-frequency components and then the first Principal Component (PC) of each side is calculated. This proposed automatic approach does not require either prior knowledge of the foreground objects in a scene or a background model. After the polarimetric components are calculated, four different foreground masks are generated to distinguish between the foreground and background.

The main contribution of this research is two-fold. Firstly, a hybrid fusion technique for fusing information from multiple spectra with different polarimetric orientation images is proposed. This stage is composed of a Fast Fourier Transform (FFT) and Principal Component Analysis (PCA), with each spectrum initially decomposed into low and high frequencies. Then, the first Principal Component (PC) of each frequency is computed and combined to obtain a final fused image. Secondly, a polarimetric BS technique for segmenting a complex background from a scene is proposed. It uses polarimetric components to compute the significant foreground masks that are then combined through a morphological operation to produce the final background area of the scene. The proposed research aims to segment a background of a complex scene utilizing MSPI features. The performances of these approaches are evaluated and compared using a publicly available MSPI dataset to demonstrate the significance of this study.

This paper is organized as follows. In Section 2, the background to multiband fusion in MSPI is fully described. In Section 3, details of the MSPI dataset used are provided and the spectral reflectance, correlations and textural features for obtaining information dissimilarity are analyzed. In Section 4, a complete two-fold BS framework and algorithm are presented. A hybrid fusion algorithmic framework, polarimetric component processing and a polarimetric BS algorithm with significant steps and proper mathematical and logical explanations are also discussed. In Section 5, the performance of the proposed hybrid fusion method and accuracy of the proposed BS algorithm are evaluated and compared with those of existing approaches in this literature. An experiment on

computational time for the proposed method is also presented and compared with other approaches in this section. Finally, concluding remarks and suggested future directions are provided in Section 6.

2. Related Works

BS techniques usually assume that the intensity of background pixels varies from that of other pixels in multiple spectra according to [25]:

$$M_{x,y, \lambda, o_i} = \begin{cases} 1 & \text{if } d(I_{x,y, \lambda, o_i}, B_{x,y, \lambda, o_i}) > \tau_G \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where τ_G is a global threshold, M_{x,y, λ, o_i} the final background mask at pixel (x, y) of a fused spectrum (λ) image at a polarimetric orientation (o_i) and d is the distance between the pixel of the predicted background (B) and that of the fused image in spectrum $\lambda(I)$ at orientation o_i . In this paper, the literature related to BS is reviewed in terms of multispectral fusion and polarimetric BS.

2.1. Fusion

In general, multispectral images contain both spectral and spatial redundancies. The former is defined by a high correlation among multiple spectra and the latter by the correlation among neighboring pixels. Multispectral fusion aims to reduce both these redundancies while destroying as little as possible of the original image's information and is usually performed by transform compression. Individual techniques and their performances related to visible and infrared image fusion are discussed in recent literature [26–28]. Although, in a lossless image compression technique, images remain almost identical to their original versions, in a lossy one, a high compression ratio such as PCA [29–33], which is based on dimensionality reduction, is calculated according to the covariance matrix (C):

$$C = \frac{1}{N} \sum_{n=1}^N (X_n - \bar{X})(X_n - \bar{X})^T \quad (2)$$

where N is the number of spectral bands, X_n is a vector containing the pixel intensities of the n^{th} band image and \bar{X} is the mean of X_n . The eigenvectors and eigenvalues are then computed from C , after that PCA scores also calculated and multiplied by those of the original pixel intensities of the band images to obtain the first PC as a fused image [28]. Through applying PCA, Bavirisetti [34] has proposed edge-preserving Fourth-order Partial Differential Equations (FPDEs) to fuse multisensor images. Despite the advantage over energy compaction of the decorrelated spectral bands, the dependence on data to calculate the C among the spectral bands is a disadvantage of PCA.

Due to the energy balance and registration problem existing in a multispectral data acquisition system, the appearances of objects demonstrate low correlations in different spectral bands [35]. However, multispectral pixel-level and transform domain-based fusion can effectively present an overall dataset. The traditional pixel-level fusion averages (AVG) method [36] contains the information of multiple spectra and, although it is simple and fast, each band's images participate equally in the fused image without their information considered. In contrast, transform domain-based fusion methods first decompose the source images into a sequence of sub-band ones through a specific mathematical transform structure and then apply some fusion rules to determine the so-called sub-band coefficients [3]. Finally, a fused image is formed using the corresponding inverse transform. Mathematical transforms, which are well known as Multiscale Geometric Analyses (MGAs) include the Discrete Wavelet Transform (DWT) [37–39], Shearlet Transform (SHT) and Discrete Cosine Transform-based Laplacian Pyramid (DCT-LP). The DWT generates four different coefficients: approximation, horizontal detail, vertical detail and diagonal detail, with a mean–maximum rule applied to fuse those of multiple spectra. Finally, an inverse transform is applied on all the final coefficients [27]. Although the DWT

provides a higher signal-to-noise ratio (SNR) and better temporal resolution than other transforms, its performance depends on the number of decomposition levels.

In the SHT [20], two significantly different images (A and B) are selected from spectral bands which further decompose into some low- and many high-frequency coefficients (L_{iA} and H_{iA} , respectively). The mean of the low frequencies is the final low-frequency approximation coefficient (Equation (3)) and, by applying an area-based feature selection method (Equations (4) and (5)), the final high-frequency detail coefficients are calculated. Eventually, a fused image is obtained using an inverse SHT as follows.

$$L(x, y) = \frac{L_A(x, y) + L_B(x, y)}{2} \quad (3)$$

$$H_i(x, y) = \begin{cases} H_i^A(x, y), & S_i^A(x, y) \geq S_i^B(x, y) \\ H_i^B(x, y), & \text{otherwise} \end{cases} \quad (4)$$

where S is the weighted energy of the local window (Ω) calculated as:

$$S(x, y) = |f(x, y)|^2 \frac{1}{\text{Card}(\Omega)} \sum_{(m,n) \in \Omega} |f(m, n)|^2 \quad (5)$$

where $\text{Card}(\Omega)$ denotes the cardinality of a window (Ω) that is calculated by counting the number of distinct elements in the (m, n) coordinates. The advantage of the SHT is that it includes multiscale, shift-invariant and multidirectional image decompositions, but its computational complexity is significant.

In the DCT-LP [40], firstly, a Gaussian Pyramid (GP) is constructed with a reduced size of a band's image and then the DCT and inverse DCT are applied on it, with an LP built by expanding the size of the previous image and then applying the DCT and inverse DCT on it. After constructing a k -level LP for individual images ($P_k^i \rightarrow \{g_k^i, l_0^i, l_1^i, \dots, l_{k-1}^i\}$, where $i = 0, 1, \dots, N$), the fusion rule is

$$\text{at the } k\text{th level, } g_k^f = g_k^i / i \quad (6)$$

$$\text{and, at the } k-1 \text{ to } 0 \text{ levels, } g_{k-1}^f = l_{k-1}^f + E(g_k^f) \quad (7)$$

where $E(g_k^f)$ is an expanded function and $l_{k-1}^f = l_{k-1}^i$ if $|l_{k-1}^i| > |l_{k-1}^j|$, where $j \neq i$ and $j = 0, 1, \dots, N$. Despite its signal extension capability, the DCT suffers from poor image quality.

Therefore, considering the pros and cons of individual methods, we propose a holistic hybrid fusion approach for the lossy compression of multispectral images with high spatial resolutions by integrating an FFT with PCA.

2.2. Segmentation

Recent literature on multispectral polarized BS has focused mainly on binary (or foreground-background) segmentation [41] after fusion [42]. Without considering some constraints or assumptions, BS in multiband images seems to be very difficult due to an object appearing in some spectra exhibiting a low correlation [35]. Approaches based on monocular segmentation usually rely on the hypotheses of visual saliency (e.g., an object appearing in a scene is highly focused) or human oversight to achieve good results [43,44]. Given the additional assumptions that can be made about the motion of an object or scene, the same issue in the temporal domain (i.e., for image sequences) is easier to tackle. There are methods for motion clustering that depend on partitioning the optical flow or trajectory points to identify image regions that behave differently from their surroundings [45]. Also, the close correlation between motion partitioning and the segmentation of video objects has gained attention in the last decade [46,47].

Foreground-background separation is an easy task if the object of interest is available in multiple spectra. An example of the assumption of visual appearance is where common foreground

objects are presented, and the background demonstrates low correlations in multiple spectra [48–50]. In contrast, methods for mutual segmentation usually assume that the same object is observed from multiple viewpoints and optimize the geometric consistency of the extracted foreground area [51–53]. Earlier mutual segmentation approaches which focus on single-spectrum imagery [53–55], with the registration problem solved using depth cameras [56,57], offer a range-based solution for detecting objects in a scene [51,52,58].

Currently, researchers are concentrating on combining spatial, spectral and polarimetric information to effectively predict the background and foreground in a scene without creating any background model. In some studies, color space mapping [13,20,29] is applied on fused multispectral polarimetric components and then fuzzy C-means clustering on separate objects from the background. Lu [59] applied PCA on different polarized images and then used a clustering technique to segment objects. These methods do not require any previous background model to segment the foreground/object from the background and are robust to illumination conditions. The proposed BS method creates some significant foreground masks from different polarimetric components and combines them to segment a complex background from a scene.

3. Analysis of the MSPI Dataset

Multispectral polarimetric data contain rich sets of information for discriminating foreground objects from the background in a complex scene.

3.1. Description of the Dataset

The multispectral polarimetric image dataset used in this paper is obtained from a publicly available one [60], which contains polarimetric and multispectral images from visible and near infrared (NIR) bands in the range of 420–1000 nm. The imaging system generates 10 different scenes in six spectral channels at four different polarimetric filter orientation angles (0, 45, 90, 135). Of them, three scenes in which the background is polarized in nature are selected for our purpose. Samples of the dataset at 0 polarimetric orientation (I_0) are shown in Figure 1, where the liquid scene is composed of reflecting objects, such as water in a bottle and a doll in a jar, the food scene of an apple and a banana, and the leaf scene of real and fake leaves with a mixed background.

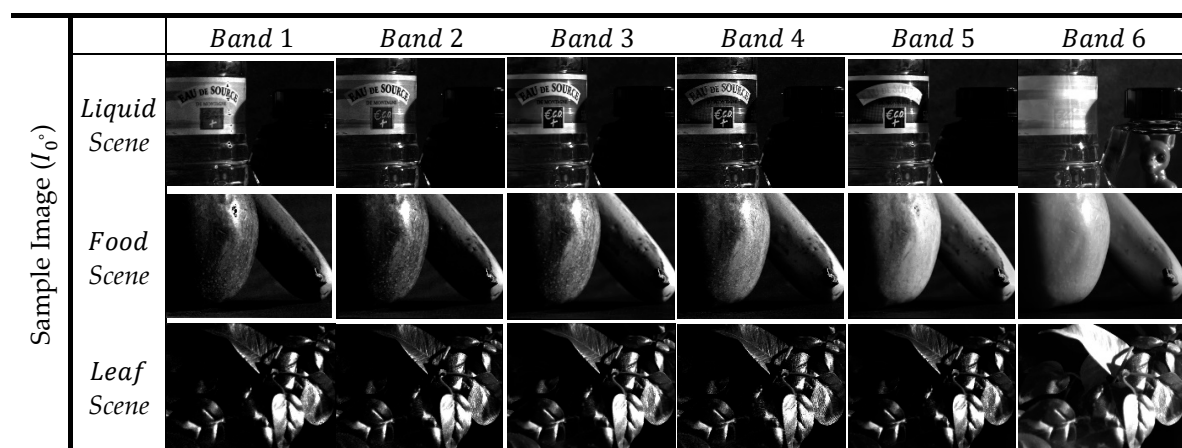


Figure 1. Samples of MSPI dataset scenes composed of six spectral channels and four polarimetric orientations [60].

3.2. Analysis of Spectral Reflectance

The camera response (ρ_i) of the i^{th} channel is described by [1] as:

$$\rho_i = \sum_{\lambda=420}^{\lambda=1000} E(\lambda) R(\lambda) O(\lambda) C_i(\lambda) \quad (8)$$

where $E(\lambda)$ is the spectral emission of the illuminant, $R(\lambda)$ the reactance, $O(\lambda)$ the global transmittance of all the optical elements and $C_i(\lambda)$ the spectral sensitivity of the i^{th} channel. After calculating the measurement matrix (M), the spectral reflectance (\hat{R}) is calculated as:

$$\hat{R} = M^{-1} \rho_i \quad (9)$$

The mean spectral reflectance of an area demonstrates variations in information among multiple bands. It is also predicted that the foreground areas of different scenes, such as the liquid (L-Bottle, L-Jar), food (F-Apple, F-Banana) and leaf (LF-Real, LF-Fake) ones, exhibit higher spectral reflectance than their backgrounds as shown in Figure 2.

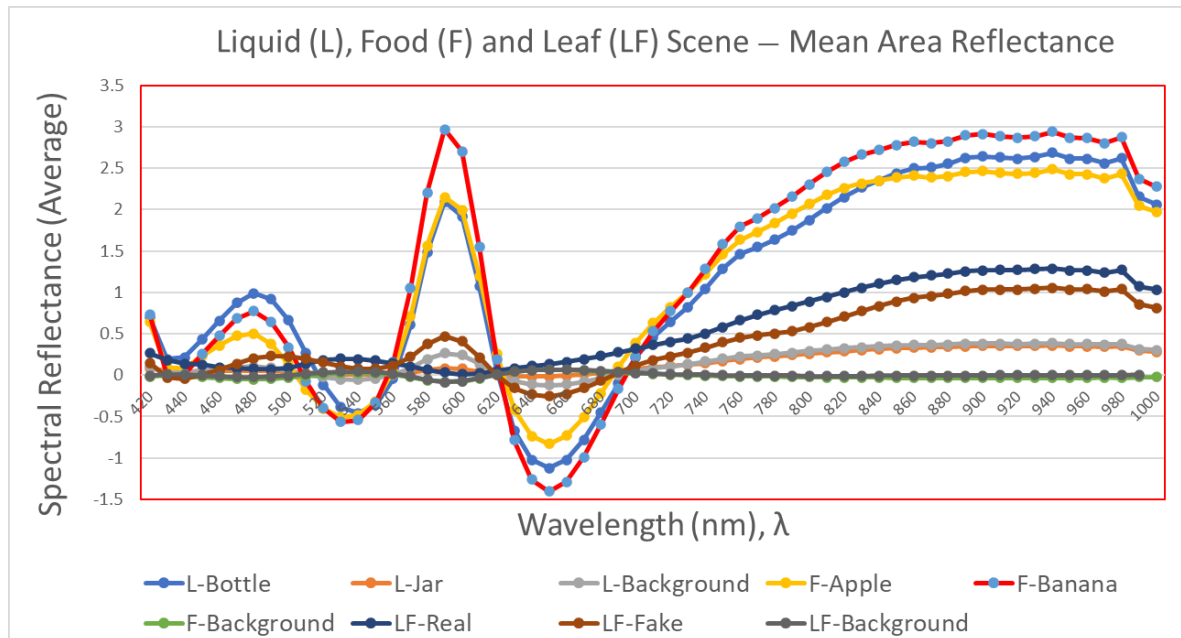


Figure 2. Multiband mean spectral reflectance measurements of the three scenes.

3.3. Analysis of Multiband the Dissimilarity Matrix

In MSPI, the data redundancies that can occur in spectral bands denote that one band's data can be partly or fully predicted from those of the others. Multiband dissimilarity is calculated in two different ways, by the Pearson correlation and Euclidean distance. We consider three different MSPI scenes to compute their dissimilarity levels in multiple spectra.

3.3.1. Pearson Correlation

Higher correlations among multiple spectra demonstrate highly redundant information. To measure an inherent correlation between two images (A_{mn} and B_{mn}), the Pearson correlation coefficient (r) is calculated as:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2) (\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (10)$$

where $\bar{A} = \frac{1}{N} \sum_{i=0}^{N-1} A_{mn}$ and $\bar{B} = \frac{1}{N} \sum_{i=0}^{N-1} B_{mn}$.

The correlation matrices in the MSPi scenes due to the multiple spectra and multiple polarimetric orientations are presented in Figure 3a–c [61]. As can be seen, band 6 in the liquid and leaf scenes is not correlated with the other bands, while band 6 in the food scene is correlated with only band 5 above a threshold of 0.90. The bands not strongly correlated are indicated in green.

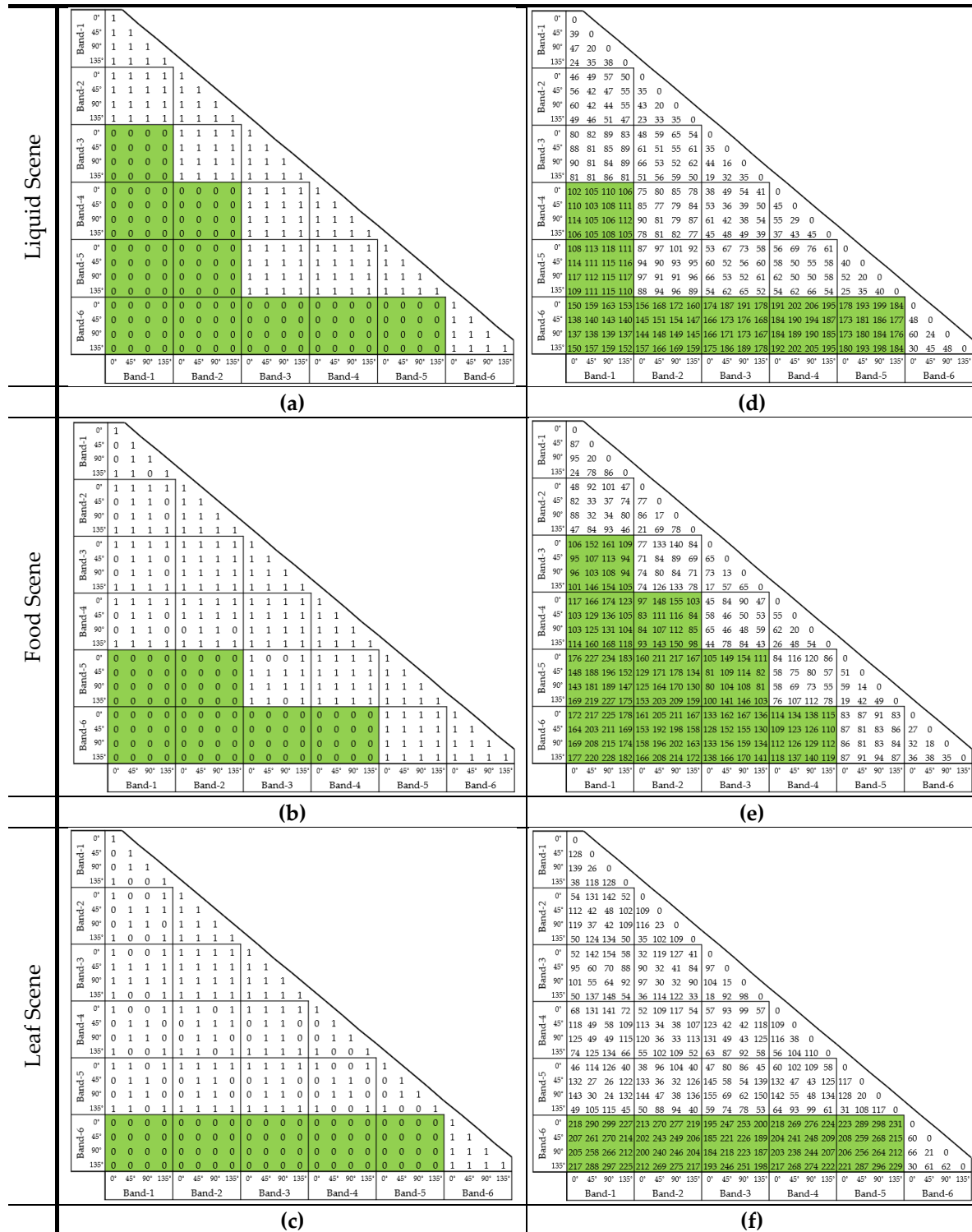


Figure 3. Multiband dissimilarity matrices for three scenes: (a–c) Pearson correlations; and (d–f) Euclidean distances.

3.3.2. Euclidean Distance

To measure the distances among multiple spectra, the classical method used is the Euclidean distance that, between two images (A_{mn} and B_{mn}), is calculated by:

$$d(A, B) = \sqrt{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (A_{mn} - B_{mn})^2} \quad (11)$$

and then normalized in the range (0–1). The distance matrices in the MSPI scenes due to the multiple spectra and different polarimetric orientations are presented in Figure 3d–f. As can be seen, band 6 is the most dissimilar to the others in the liquid and leaf scenes, while, in the food scene, bands 5 and 6 demonstrate similar information but are dissimilar to the others. The bands with large distances are indicated in green.

3.4. Analysis of Multiband Textural Features

In MSPI, textural features describe the local spatial variations in intensity as well as the homogeneity of images. They are divided into the two broad categories of first- and second-order statistical features, with the former dependent mainly on a histogram analysis of gray-level image representations. Considering z_i as the gray value of the i^{th} pixel, $P(z_i)$ is a normalized histogram and N the number of distinctive gray levels, with the first-order histogram defined as [50]:

$$P(z_i) = \frac{\text{number of pixels with value } I}{\text{number of total pixels}} \quad (12)$$

The first-order statistical features are defined as follows [62].

1. **Mean** is a measure of the spreading of the distribution from the mean value. $\text{FOSMen}, \mu = \sum_{i=0}^{N-1} z_i P(z_i)$ (13)

2. **Standard deviation** is used to sharpen edges as the intensity level changes by a large value at the edge of an image. $\text{FOSStd}, \sigma = \sqrt{\sum_{i=0}^{N-1} (z_i - \mu)^2 P(z_i)}$ (14)

3. **Energy** is a measure of the homogeneity of the histogram. $\text{FOSEng}, E = \sum_{i=0}^{N-1} (P(z_i))^2$ (15)

4. **Skewness** is a measure of the degree of the histogram's asymmetry around the mean. $\text{FOSSkw}, \mu_3 = \frac{\frac{1}{N} \sum_{i=0}^{N-1} (z_i - \mu)^3}{\left(\sqrt{\sum_{i=0}^{N-1} (z_i^2 - \mu^2)} \right)^3}$ (16)

5. **Kurtosis** is a measure of the histogram's sharpness, that is, whether the data are peaked or flat relative to a normal distribution. $\text{FOSKur}, \mu_4 = \frac{\frac{1}{N} \sum_{i=0}^{N-1} (z_i - \mu)^4}{\left(\sqrt{\sum_{i=0}^{N-1} (z_i^2 - \mu^2)} \right)^4}$ (17)

The second-order statistics or Gray-level Co-occurrence Matrix (GLCM) define a square matrix the size of which represents the probability of the gray value (g_1) being moved from a fixed spatial location relationship (size and direction) to another gray value (g_2). Assuming that $f(i_1, j_1)$ is a 2D gray-scale image, where S is a set of pixels with a certain spatial relationship in the region and P the GLCM, this is expressed as [63]:

$$P(i, j) = \frac{\#\{[(i_1, j_1), (i_2, j_2)] \in S | f(i_1, j_1) = g_1 \ \& \ f(i_2, j_2) = g_2\}}{\#S} \quad (18)$$

where # is used to define the quantity. The second-order statistical features are defined as follows.

1. **Contrast** is a measure of the local variations present in an image, that is, it reflects the depth and smoothness of the image's textural structure.

$$\text{SOSCon} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 P(i, j) \quad (19)$$
2. **Correlation** is a measure of the gray-level linear dependence between the pixels at specified positions relative to each other, that is, it reflects the similarity of an image's texture in the horizontal or vertical direction.

$$\text{SOSCor} = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\bar{x})(j-\bar{y})P(i, j)}{\sigma_x \sigma_y} \quad (20)$$

where, $\bar{x} = \sum_{i=0}^{N-1} i \sum_{j=0}^{N-1} P(i, j)$,
 $\bar{y} = \sum_{j=0}^{N-1} j \sum_{i=0}^{N-1} P(i, j)$,
 and $\sigma_x^2 = \sum_{i=0}^{N-1} (i-\bar{x})^2 \sum_{j=0}^{N-1} P(i, j)$,
 $\sigma_y^2 = \sum_{j=0}^{N-1} (j-\bar{y})^2 \sum_{i=0}^{N-1} P(i, j)$
3. **Angular Second Moment or Energy** is a measure of the global homogeneity of an image.

$$\text{SOSEng} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j)^2 \quad (21)$$
4. **Homogeneity** is a measure of the local homogeneity of an image.

$$\text{SOSHom} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i, j)}{1+|i-j|} \quad (22)$$
5. **Entropy** is a measure of a histogram's uniformity, that is, it reflects the complexity of the textural distribution.

$$\text{SOSEnt} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j) \log_2[P(i, j)] \quad (23)$$

The MSPI's textural features are presented in Figure 4. As can be seen, the kurtosis, skewness, energy and entropy values demonstrate high variations in the bands in the liquid, food and leaf scenes. Therefore, it is predicted that the information in MSPI is not inherently correlated among the bands.

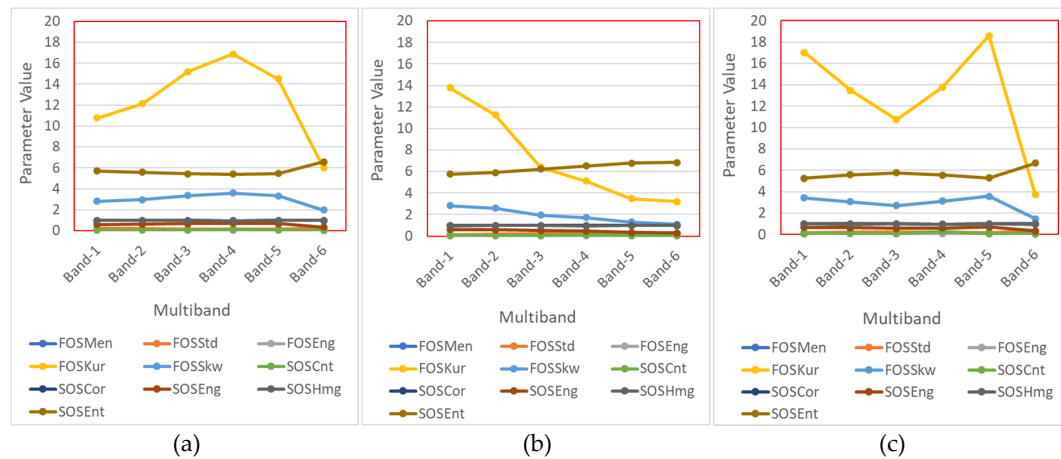


Figure 4. Analyses of multiband first- and second-order textural features of scenes: (a) liquid; (b) food; and (c) leaf.

4. Proposed Two-Fold BS

Our approach can be described as two-fold as it uses a low-level fusion of multiple spectra and polarimetric component-based BS.

4.1. Overall Framework and Algorithm

As, in our proposed method, the multiple spectra of each polarimetric filter orientation are fused to solve intricate situations and benefit from the complementarity of the information among multiple bands, it consists of multiband polarimetric fusion and BS. The whole process is decomposed into five main steps: in the first, multiband polarimetric data are captured and analyzed; in the second, a combination of a FFT and PCA is applied on multiple spectra to calculate the fused imagery/image of each polarimetric filter orientation; in the third, the polarimetric components are calculated; in the fourth, four different significant masks are obtained; and, in the last, all of the foreground masks are

combined and applied in a morphological operation to generate the final background mask. Figure 5 demonstrates this proposed two-fold BS framework.

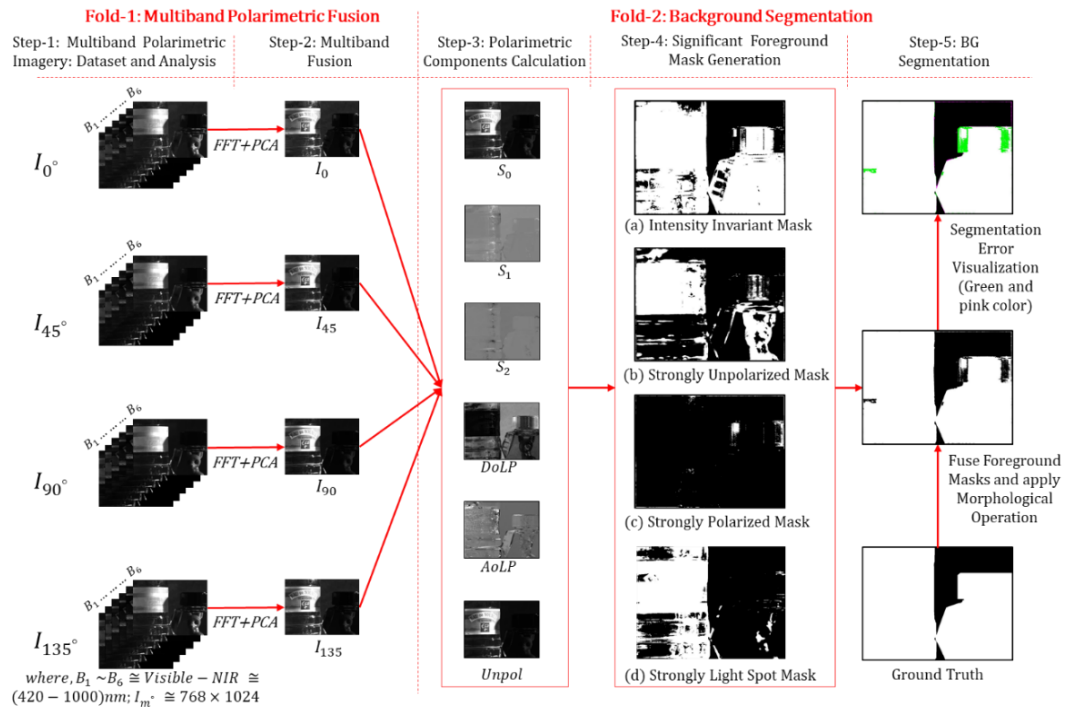


Figure 5. Proposed two-fold Background Segmentation (BS) framework.

The aim of this research is the joint use of spatial, spectral and polarimetric information to segment a polarized background from a complex scene, with Algorithm 1 describing the steps for multiband fusion and BS. In it, the significance of fusion is analyzed by calculating the spectral reflectance, correlation and textural features, and then the proposed fusion algorithm (Algorithm 2) is applied to determine the polarimetric components. Finally, Algorithm 3 is used to segment the background from a complex scene.

Algorithm 1. Multiband Fusion and BS

Requires:

Multiband Polarimetric Image Dataset

Ensures:

Background (BG) segmentation through Multiband Polarimetric Fusion

- 1: *Multiband Polarimetric Image Dataset Analysis*
 - 2: *Calculate Spectral Reflectance of the Mean Foreground (FG) and BG Area*
 - 3: *Calculate Correlation Among Bands and Polarimetric Orientation*
 - 4: *Calculate First Order and Second Order Texture Features*
 - 5: **if** *Information Differs Significantly among Bands*, **then**
 - 6: *Polarimetric Orientation-wise Multiband Fusion (Algorithm 2)*
 - 7: *Evaluate and Performance of the Fusion Method Statistically*
 - 8: **end if**
 - 9: **if** $I_{0^\circ}, I_{45^\circ}, I_{90^\circ}, I_{135^\circ}$ **exist then**
 - 10: *Compute Stokes Vector: $S_0 - S_2$*
 - 11: *Compute Polarimetric Components: DoLP, AoLP and Unpol*
 - 12: **end if**
 - 13: *BS in Polarimetric Imagery (Algorithm 3)*
 - 14: *Evaluate and Compare Performance of the Proposed Method Statistically*
-

4.2. Hybrid Fusion Framework and Algorithm

An analysis of MSPI in terms of spectral reflectance, correlation and textural features can predict the significance of multiband pixel-level fusion. The proposed fusion method is based on 2D neighborhood information, such as gradients or edge magnitudes and orientations, with the source images multiple spectral ones (Im_i) and the fused image (Im_f) based on different polarimetric filter orientations. This method is hybrid in nature as it combines two different techniques, an FFT and PCA, as shown in Figure 6. Firstly, by applying a Gaussian low-pass filter, each spectral image is decomposed into low- and high-frequency components (LPF_i and HPF_i , respectively). Then, a Discrete Fourier Transform (DFT_i) of each band is calculated with the help of the FFT. Multiplying the DFT_i by the LPF_i and HPF_i , the final FFTs (FLP_i and FHP_i) are obtained. Then, applying an inverse FFT, the results are converted to the spatial domain as the final low and high frequencies of each band as:

$$DFT_i = fft(band_i) \quad (24)$$

$$FLP_i = LPF_i \times DFT_i, \quad FHP_i = HPF_i \times DFT_i \quad (25)$$

$$LP_i = ifft(FLP_i), \quad HP_i = ifft(FHP_i) \quad (26)$$

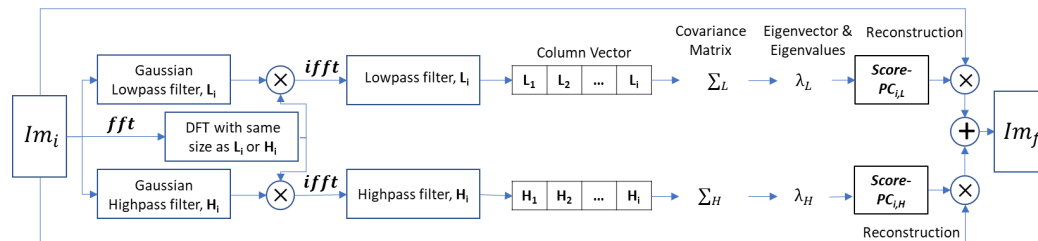


Figure 6. Proposed framework of hybrid fusion.

Then, the covariance matrix, eigenvector and eigenvalue of each low- and high-frequency component are calculated and the transformed image reconstructed by determining their first PCs. Finally, these components are combined to obtain the fused image; Figure 6 illustrates the framework of this method. MSPI often presents complementary information about the foreground and background in a scene efficiently combined by pixel-level fusion. The process for polarimetric orientation-wise multiband fused imagery is described in Algorithm 2.

Algorithm 2. Polarimetric Orientation-wise Multiband Fusion

Requires:

MSPI Scene Dataset

Ensures:

Polarimetric Orientation – wise Fused Imagery

- 1: **for** all polarimetric orientations_k **do**
 - 2: **for** all band_i **do**
 - 3: Create a Gaussian Low Pass Filter (LPF_i) and Gaussian High Pass Filter (HPF_i)
 - 4: Calculate a Discrete Fourier Transform DFT_i
 - 5: Multiply DFT_i with LPF_i and HPF_i
 Convert the Result to the Spatial Domain by inverting Multiplication result
 - 6: Apply Inverse Fast Fourier Transformation to the multiplication results which produce band-wise final results as $LP_{k,i}$ and $HP_{k,i}$
 - 7: **end for**
 - 8: Calculate the covariance and eigenvector of LP_k and HP_k
 - 9: Calculate the first Principal Component of LP_k and HP_k
 - 10: Calculate the fused imagery by adding both Principal Components
 - 11: **end for**
-

4.3. Calculation of Polarimetric Components

The fused images calculated at different polarimetric orientations (0° , 45° , 90° , 135°) are labeled I_{0° , I_{45° , I_{90° and I_{135° , respectively, to determine their polarimetric components. The Stokes parameters (S_0 – S_2) [64] describe the linear polarization characteristics using a three-element vector (S), as shown in Equation (27), where S_0 represents the total intensity of light, S_1 the difference between horizontal and vertical polarizations and S_2 the difference between linear $+45^\circ$ and -45° ones.

$$S = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} I_{0^\circ} + I_{90^\circ} \\ I_{0^\circ} - I_{90^\circ} \\ I_{45^\circ} - I_{135^\circ} \end{bmatrix} \quad (27)$$

The *DoLP* is a measure of the proportion of the linear polarized light relative to the light's total intensity, the *AoLP* the orientation of the major axis of the polarization ellipse, which represents the polarizing angle where the intensity should be the strongest, and the *Unpol* a measure of the unpolarized light according to the dataset's author. The *DoLP*, *AoLP* and *Unpol* are derived from the Stokes vector, respectively, as:

$$DoLP = \frac{I_{pol}}{I_{tot}} = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \quad (28)$$

$$AoLP = \frac{1}{2} \tan^{-1} \left(\frac{S_2}{S_1} \right) \quad (29)$$

$$Unpol = S_0 - \frac{1}{2} \sqrt{S_1^2 + S_2^2} \quad (30)$$

4.4. Proposed BS Algorithm

The proposed BS algorithm, which involves five different steps for generating four significant foreground masks through processing the polarimetric components, is presented in Algorithm 3.

Algorithm 3. BS in Polarimetric Imagery

Requires:

Stokes Vector and Polarimetric Components

Ensures:

Segmented Background from a scene

- 1: **if** $S_0 - S_2$, *DoLP*, *AoLP*, *UoLP* **exist then**
 - 2: *Significant Foreground Mask Generation*
 - 3: Construct an intensity invariant mask through differentiating the median filtering version of unpolarized and polarized imagery
 - 4: Calculate a strongly unpolarized foreground mask in two different ways utilizing *AoLP* and *DoLP*
 - 5: Calculate a strongly polarized foreground mask
 - 6: Calculate a strong light intensity mask based on azimuth angle and S_0 .
 - 7: Combine steps 3–6 and apply a morphological operation to segment the total background area of a scene.
 - 8: **end if**
-

- Step 1. Construction of intensity-invariant mask

The proportion of the unpolarized to total intensity is defined as the degree of unpolarized intensity (Equation (31)). A *DoLP* modulation coefficient can be used to stretch the contrast and, thereby, the separability of the foreground/background areas in a polarimetric scene [29]. This function should be continuous in the field of the *DoLP* and monotonically increase with it. A logarithmic function is a natural choice for this purpose, as pointed out by Shannon [65], and the *DoLP* modulation coefficient (M_{DoLP}) is defined in Equation (32). The differentiation between the median filtering

versions of the degree of the unpolarized image (D_{Unpol}) and M_{DoLP} in Equation (33) produces an intensity-invariant mask as:

$$D_{Unpol} = Unpol \div S_0 \quad (31)$$

$$M_{DoLP} = DoLP \cdot \log(1 + DoLP) \quad (32)$$

$$Mask_{IntInv} = |D_{Unpol} - M_{DoLP}| > Threshold \quad (33)$$

- Step 2. Construction of strongly unpolarized mask

An indirect unpolarized mask is calculated through a *DoLP* analysis. Firstly, due to the dark offset and spatial nonuniformity of light, a *DoLP* image may contain noise. Therefore, a FFT with a Gaussian low-pass filter is applied to smooth it and reduce the noise level (Equation (34), where $D(u, v)$ is the distance from a point (u, v) to the center of its frequency in the *DoLP*, and D_0 is the cutoff frequency) and the FFT of the *DoLP* image with padding obtained by Equation (35). The transformed image is multiplied by the filter and then an inverse FFT applied to obtain a noise-free smooth image (Equation (36)). The calculated low-pass filtered image of the *DoLP* (*Smooth*) is multiplied by the D_{Unpol} to obtain a lower intensity (lower global threshold) of the polarized image and then binarized to calculate a noise-free smooth unpolarized mask (Equation (37)).

$$L(u, v) = e^{-D^2(u, v)/2D_0^2} \quad (34)$$

$$F(u, v) = \text{fft}(DoLP) \quad (35)$$

$$S_{DoLP} = \text{ifft}(L(u, v) \times F(u, v)) \quad (36)$$

$$Mask_{IndUnp} = \sim \text{imbinarize}(S_{DoLP} \times D_{Unpol}) \quad (37)$$

A direct unpolarized mask is calculated using an *AoLP* textural analysis. A desultory value of the *AoLP* denotes a rough surface that indicates the unpolarized part of a scene and a continuous one a smooth surface, which indicates the polarized part of a scene [20]. To segment the unpolarized part, firstly, a textural image is created by applying entropy filtering around a 9-by-9 neighborhood of the *AoLP*. The aim is to obtain an output image ($Mask_{DirUnp}$) in which each output pixel contains the entropy value of the 9-by-9 neighborhood around the corresponding pixel in the input image of the *AoLP* as:

$$Mask_{DirUnp} = \text{imbinarize}(\text{entropyfilt}(AoLP)) \quad (38)$$

A final unpolarized mask is generated through a bit-AND operation. Because the contrast, illumination and energy levels are not uniform among the scenes in the bands, some portions of an individual scene gain advantages during this process, with some considering an *AoLP* textural analysis and others a *DoLP* one. Applying this operation on both sides obtains an exact unpolarized mask as:

$$Mask_{StUnp} = Mask_{IndUnp} \& Mask_{DirUnp} \quad (39)$$

- Step 3. Calculation of strongly polarized mask

Based on Otsu's method for calculating the global threshold, the *DoLP* image is converted to a binary one to obtain a strongly polarized mask of a scene as:

$$Mask_{StPol} = \text{imbinarize}(DoLP) \quad (40)$$

- Step 4. Calculation of strong light intensity mask

The inclination or azimuth angle (β) of the polarized light is calculated. The electric field vector (E) is divided into two orthogonal plane waves (E_x and E_y) in the direction of the x- and y-axes,

respectively. The calculations of E_x and E_y depend on S_0 , S_1 and S_2 , as in Equations (41) and (42), and the azimuth angle is defined in Equation (43). Finally, as the β is calculated from the $DoLP$, a noise-free version of it is determined using Equations (41)–(43) and renamed $S_{Azimuth}$.

$$E_x = \sqrt{(S_0 \times DoLP) + S_1} \quad (41)$$

$$E_y = \sqrt{(S_0 \times DoLP) - S_1} \quad (42)$$

$$\beta = \tan^{-1}(E_y/E_x) \quad (43)$$

The total intensity image (S_0) is divided by the logarithmic function of β to calculate the area of the maximum strong light intensity. The purpose of using this natural logarithmic function of β is to stretch the contrast and, thus, the separability of the strong light intensity areas in polarimetric images by:

$$Mask_{StLight} = imbinarize\left(\frac{S_0}{\log(1 + S_{Azimuth})}\right) \quad (44)$$

- Step 5. Combination of foreground masks and application of morphological operation

All the foreground areas, such as the intensity-invariant, strongly unpolarized, strongly polarized and strong light intensity masks, are calculated and fused to obtain the final foreground mask (Equation (45)). Then, a morphological structuring operation based on a dilation followed by an erosion to boost the segmentation performance is applied on the combined background mask. The final background area in the corresponding scene is denoted in black and it can be seen that some object parts are mixed with the background due to the presence of polarization.

$$Mask_{Foreground} = Mask_{IntInv} \cup Mask_{StUnp} \cup Mask_{StPol} \cup Mask_{StLight} \quad (45)$$

5. Experimental Results

In this section, performance evaluations and comparisons of the proposed two-fold BS and other approaches using different metrics in terms of multiband fusion and polarimetric BS are discussed. Also, computational time analyses of individual methods are conducted.

5.1. Performance Evaluation of MSPI Fusion

5.1.1. Selection of Fusion Metric

Currently, the quality of a fused image can be quantitatively evaluated using the metrics [66] Mean Absolute Percentage Error (MAPE), Peak Signal-to-Noise Ratio (PSNR), Pearson Correlation Coefficient (PCOR) and Mutual Information (MI). The MAPE is a measure of the closeness between predictions of the reference and fused images. The PSNR block computes the PSNRs of the reference and fused images. A lower value of MAPE and higher value of PSNR indicate a better quality of the reconstructed or fused image. The PCOR computes the linear correlation coefficient between the reference and fused images, while MI indicates the mutual dependence between them, with the higher their values, the better the correlation. Considering two images A and B , the mathematical evaluation measures of the aforementioned metrics are:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \left(\frac{A_i - B_i}{B_i} \right) \times 100 \right| \quad (46)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (47)$$

$$PCOR = \frac{\sum_m \sum_n (A_{mn} - \bar{A}) (B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2) (\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (48)$$

$$MI = \sum_{a \in A} \sum_{b \in B} P(a, b) \log \frac{P(a, b)}{P(a) P(b)} \quad (49)$$

5.1.2. Observation of Fusion Quality

To evaluate the performance of fusion, the median image of each band is considered the reference one and the fused image the resultant one. The former contains the central value of the bands and tends to retain more detailed information. Table 1 demonstrates that the average PSNR value of the proposed method is lower than other methods; however, the average MAPE value is significantly better than some references. The average metric value of the three scenes demonstrates that the fusion quality of our proposed hybrid fusion-based approach is superior in terms of its higher PCOR and MI compared to those in the current literature.

Table 1. Mean Absolute Percentage Error and Peak Signal-to-Noise Ratio error metrics (MAPE, PSNR) and Pearson Correlation Coefficient and Mutual Information correlation metrics (PCOR, MI) of different fusion methods.

		MAPE				PSNR				PCOR				MI			
		0°	45°	90°	135°	0°	45°	90°	135°	0°	45°	90°	135°	0°	45°	90°	135°
Liquid Scene	DWT [37]	26.28	33.42	41.25	31.74	31.94	32.96	31.73	32.28	0.99	0.99	0.99	0.98	2.32	2.42	2.32	2.18
	PCA [33]	19.23	26.47	34.69	22.54	29.05	28.88	28.86	28.86	0.99	0.99	0.99	0.99	2.62	2.72	2.64	2.49
	DCT-LP [40]	146.25	167.14	189.36	150.89	23.19	24.57	24.12	22.99	0.81	0.83	0.84	0.82	0.93	1.09	1.01	0.92
	SHT [20]	93.80	123.43	164.35	116.91	19.60	19.32	20.01	19.46	0.86	0.88	0.88	0.86	1.32	1.48	1.37	1.25
	FPDE [34]	16.79	21.53	27.83	19.28	30.48	31.08	31.24	30.51	0.99	0.99	0.99	0.99	2.73	2.87	2.77	2.63
	Proposed	30.86	39.55	52.80	35.23	18.90	19.37	19.49	19.10	0.99	0.99	0.99	0.99	2.89	2.97	2.84	2.75
Food Scene	DWT [37]	11.13	16.33	15.08	14.22	32.56	32.28	31.37	32.04	0.99	0.99	0.99	0.99	2.68	2.60	2.57	2.76
	PCA [33]	5.06	8.81	8.56	6.64	30.19	28.13	27.53	30.00	0.99	0.99	0.98	0.99	3.02	2.76	2.68	3.00
	DCT-LP [40]	93.21	156.36	157.00	122.33	21.64	21.59	21.24	22.05	0.84	0.81	0.81	0.83	1.17	1.04	1.02	1.16
	SHT [20]	20.39	30.40	30.56	22.83	20.47	19.21	18.38	20.01	0.95	0.92	0.92	0.94	1.87	1.70	1.62	1.85
	FPDE [34]	19.79	31.96	30.23	25.02	20.11	18.76	19.16	19.93	0.90	0.80	0.81	0.89	2.21	1.77	1.77	2.20
	Proposed	10.14	15.66	15.08	12.22	16.90	18.15	18.32	17.14	1.00	0.99	0.99	1.00	3.32	3.07	3.05	3.43
Leaf Scene	DWT [37]	15.00	14.45	14.52	14.95	26.14	26.63	26.60	26.29	0.98	0.97	0.97	0.98	2.20	2.07	2.07	2.22
	PCA [33]	5.32	8.64	9.51	5.37	26.95	22.53	21.97	26.60	0.99	0.94	0.93	0.99	2.78	2.35	2.30	2.75
	DCT-LP [40]	55.15	63.95	62.89	54.54	14.87	14.53	14.61	14.82	0.84	0.76	0.75	0.84	1.17	1.10	1.12	1.18
	SHT [20]	26.72	32.41	36.56	26.46	13.19	11.96	11.89	13.04	0.86	0.76	0.74	0.85	1.58	1.57	1.56	1.60
	FPDE [34]	9.51	24.55	32.66	10.82	25.41	20.43	18.84	24.64	0.98	0.89	0.86	0.97	3.43	2.01	1.48	3.28
	Proposed	10.41	14.39	15.58	10.30	10.54	12.14	12.22	10.75	0.99	0.96	0.95	0.99	2.89	2.52	2.47	2.88
Average	DWT [37]	17.47	21.40	23.62	20.30	30.21	30.63	29.90	30.20	0.99	0.98	0.98	0.99	2.40	2.36	2.32	2.39
	PCA [33]	9.87	14.64	17.59	11.52	28.73	26.51	26.12	28.49	0.99	0.97	0.97	0.99	2.81	2.61	2.54	2.74
	DCT-LP [40]	98.20	129.15	136.42	109.25	19.90	20.23	19.99	19.95	0.83	0.80	0.80	0.83	1.09	1.08	1.05	1.09
	SHT [20]	46.97	62.08	77.16	55.40	17.75	16.83	16.76	17.50	0.89	0.85	0.85	0.88	1.59	1.58	1.52	1.57
	FPDE [34]	15.36	26.01	30.24	18.37	25.33	23.42	23.08	25.03	0.95	0.89	0.88	0.95	2.79	2.22	2.01	2.70
	Proposed	17.14	23.20	27.82	19.25	15.44	16.55	16.68	15.66	0.99	0.98	0.98	0.99	3.03	2.86	2.79	3.02

5.1.3. Comparison of Performances of Fusion Methods

The textural features of the fused images are presented in Figure 7. It can be seen that the kurtosis, skewness and energy of the liquid, food and leaf scenes demonstrate lower values for the proposed hybrid fusion method than for the DWT [37], PCA [33], AVG [36], DCT-LP [40], SHT [20] and FDPE [34] while, in contrast, its entropy is predicted to be higher than those of the DWT, PCA and AVG methods. Therefore, it is considered that, overall, the proposed hybrid pixel-level fusion method generates the maximum amount of information among multiple spectra.

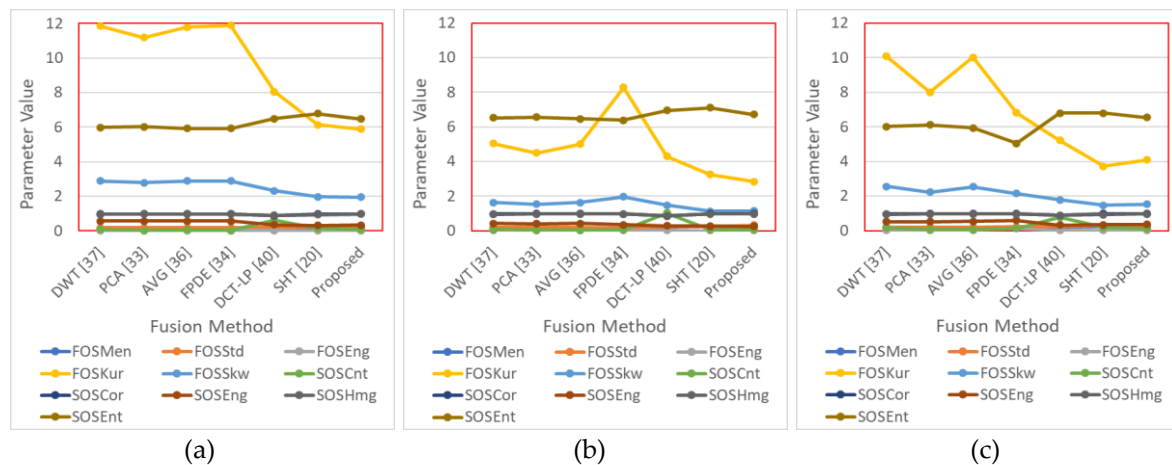


Figure 7. Analysis of multiband fusion-based first- and second-order textural features of scenes: (a) liquid, (b) food and (c) leaf.

5.1.4. Visualization of Fusion Performance

It is obvious that multiband fusion methods demonstrate better BS accuracy than direct processing [67,68]. The multiband fusion results shown in Figure 8 demonstrate that the proposed hybrid fusion approach has an advantage over the other methods as it appears strongly at the edge of an object, which will enable foreground objects to be further separated from a scene.

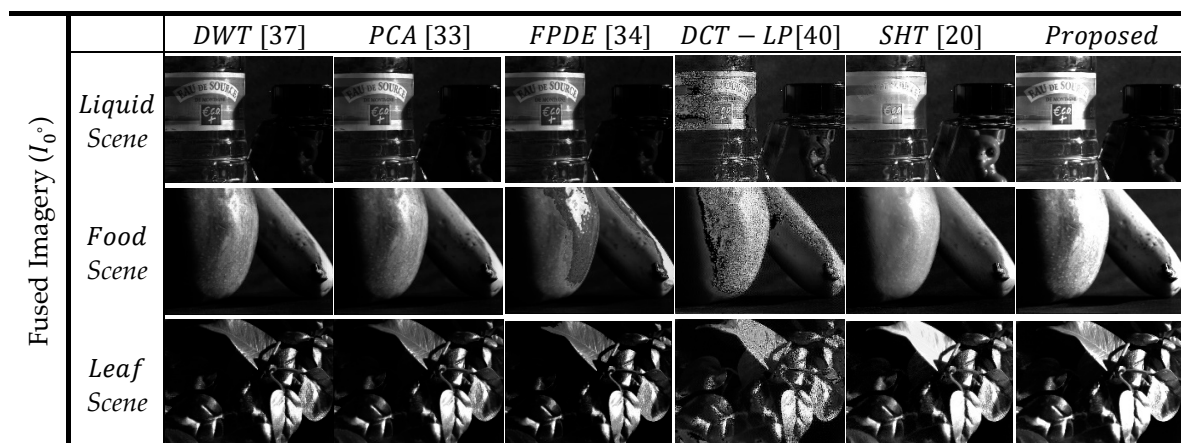


Figure 8. Samples of results of fusion methods across multiple bands at 0° polarimetric orientation.

5.2. Calculation of Polarimetric Component

Based on different fusion methods, the polarimetric components S_0 and $DoLP$ are calculated and shown in Figure 9. Analyzing the $DoLP$ values, it is predicted that, in the scenes, the backgrounds are polarized, whereas the foregrounds have mixed polarized and unpolarized intensities.


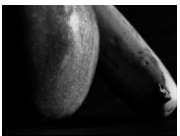





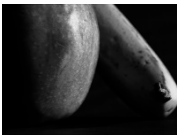
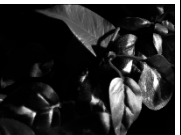
























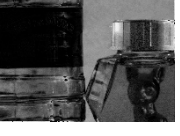




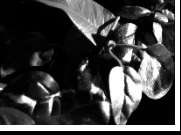
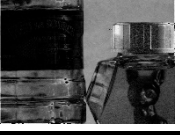
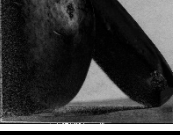

	S_0			$DoLP$		
	Liquid	Food	Leaf	Liquid	Food	Leaf
DWT [37]						
PCA [33]						
AVG [36]						
DCT-LP [40]						
SHT [20]						
FPDE [34]						
Proposed						

Figure 9. Polarimetric components calculated using different fusion methods.

5.3. Performance Evaluation of MSPI BS

5.3.1. Selection of BS Metric

The BS method is evaluated at the pixel level of a binarized scene in which the foreground and background are white and black, respectively. Its performance can be divided into four pixel-wise classification results: true positives (T_p), which mean correctly detected foreground pixels; false positives (F_p), that is, background pixels incorrectly detected as foreground ones; true negatives (T_n), which indicate correctly detected background pixels; and false negatives (F_n), that is, foreground pixels incorrectly detected as background ones. The binary classification metrics used in this paper are accuracy, specificity, sensitivity, Geometric mean (G-mean), precision, recall and the F1-score. Accuracy is measured as the proportion of true results obtained, either T_n or T_p ; specificity (a T_n fraction) as the proportion of actual negatives predicted as negatives; sensitivity (a T_p fraction) as the proportion of actual positives predicted as positives; the G-mean, the root of the product of specificity and sensitivity; precision, the number of foreground pixels detected that are actually foreground ones; recall, the number of foreground pixels detected from the actual foreground (recall and sensitivity are similar); the F1-score (a boundary F1 measure) the harmonic mean of the precision and recall values, which measures how closely the predicted boundary of an object matches its ground-truth and is an

overall indicator of the performance of binary segmentation. The mathematical evaluation measures of the aforementioned metrics follow [69]:

$$Accuracy (AC) = \frac{T_p + T_n}{T_p + F_n + T_n + F_p} \quad (50)$$

$$Specificity (SP) = \frac{T_n}{T_n + F_p} \quad (51)$$

$$Sensitivity (SN) \text{ or } Recall (RC) = \frac{T_p}{T_p + F_n} \quad (52)$$

$$Geometric - Mean (GM) = \sqrt{Specificity \times Sensitivity} \quad (53)$$

$$Precision (PR) = \frac{T_p}{T_p + F_p} \quad (54)$$

$$F1 - Score (F1S) = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (55)$$

5.3.2. Generation of Ground Truth

Each ground truth is generated manually by an expert with the maximum possible background among bands covered. In Figure 10, for the scenes' binary ground truths, the black and white pixels indicate their backgrounds and foregrounds, respectively.

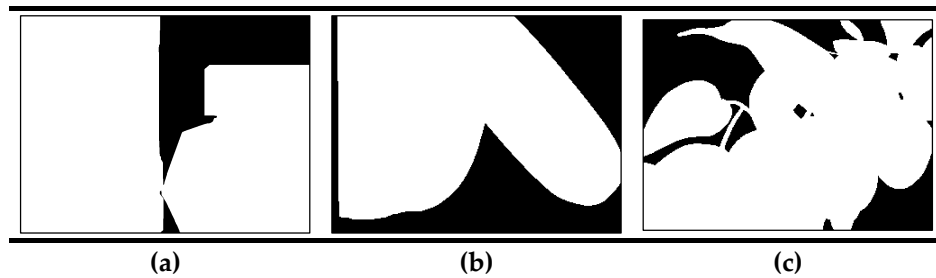


Figure 10. Ground truths of the MSPI dataset for: (a) liquid scene, (b) food scene, (c) leaf scene.

5.3.3. Comparison of BS Accuracy: Direct vs. Fusion

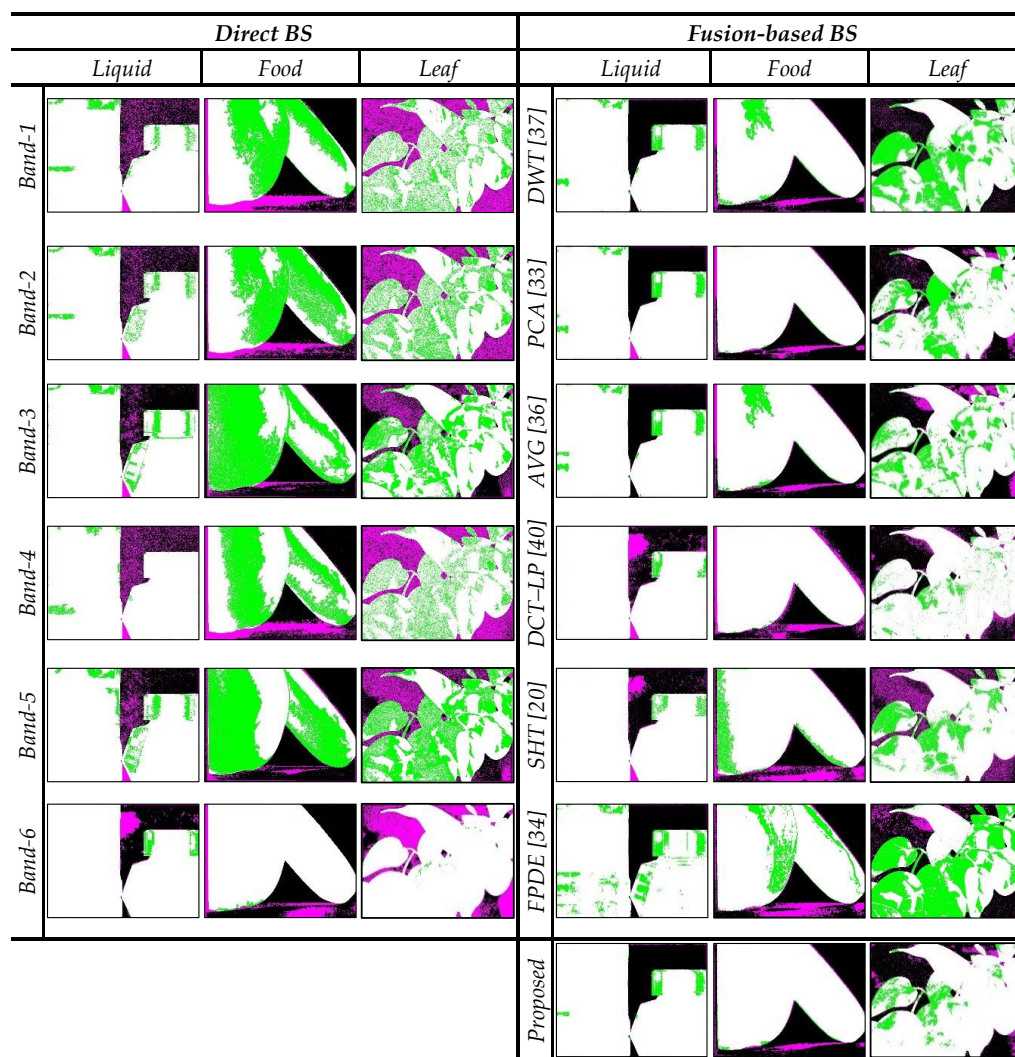
In Table 2, the performances of BS in terms of the metrics for direct and fusion-based approaches are compared, where B-1 to B-6 denote Bands 1 to 6, respectively, in the range of 400–1000 nm. As can be seen, the mean of an individual evaluation metric is significantly higher for a fusion-based BS method than for a direct band-wise BS one. Also, it can be ascertained that the mean accuracy, G-mean and F1-score are 0.88, 0.86 and 0.91, respectively, for the former, and 0.74, 0.67 and 0.80 for the latter.

5.3.4. Visualization of Performance of BS

The polarimetric BS errors are shown in Figure 11 in which purple and green denote the error areas of the segmentation methods present in the ground truths' black and white portions, respectively. These errors are lower for the fusion-based methods than for the direct band-wise ones, with our proposed hybrid fusion-based polarimetric BS approach superior to all the others.

Table 2. Evaluation of direct- vs. fusion-based BS performances for three scenes in the MSPI dataset.

		Direct BG Separation								Fusion-based BG Segmentation						
		AC	SP	SN	GM	PR	RC	FIS		AC	SP	SN	GM	PR	RC	FIS
<i>Liquid Scene</i>	B-1	0.91	0.68	0.96	0.80	0.93	0.96	0.94	DWT [37]	0.96	0.97	0.96	0.97	0.99	0.96	0.98
	B-2	0.91	0.74	0.95	0.84	0.94	0.95	0.95	PCA [33]	0.96	0.95	0.96	0.95	0.99	0.96	0.97
	B-3	0.92	0.84	0.94	0.89	0.96	0.94	0.95	AVG [36]	0.96	0.98	0.95	0.97	1.00	0.95	0.97
	B-4	0.91	0.61	0.98	0.77	0.92	0.98	0.95	DCT-LP [40]	0.94	0.73	0.98	0.85	0.94	0.98	0.96
	B-5	0.91	0.78	0.94	0.86	0.95	0.94	0.94	SHT [20]	0.95	0.81	0.98	0.89	0.96	0.98	0.97
	B-6	0.95	0.81	0.98	0.89	0.96	0.98	0.97	FPDE [34]	0.92	0.98	0.91	0.94	1.00	0.91	0.95
									Proposed	0.97	0.98	0.97	0.97	1.00	0.97	0.98
<i>Food Scene</i>	B-1	0.62	0.69	0.59	0.64	0.81	0.59	0.68	DWT [37]	0.91	0.85	0.93	0.89	0.93	0.93	0.93
	B-2	0.59	0.76	0.51	0.63	0.83	0.51	0.63	PCA [33]	0.95	0.85	0.99	0.92	0.94	0.99	0.96
	B-3	0.54	0.88	0.39	0.59	0.88	0.39	0.54	AVG [36]	0.92	0.87	0.94	0.90	0.94	0.94	0.94
	B-4	0.60	0.67	0.56	0.61	0.79	0.56	0.66	DCT-LP [40]	0.92	0.75	0.99	0.86	0.90	0.99	0.94
	B-5	0.53	0.87	0.37	0.57	0.86	0.37	0.52	SHT [20]	0.85	0.73	0.91	0.82	0.88	0.91	0.89
	B-6	0.93	0.80	0.99	0.89	0.92	0.99	0.95	FPDE [34]	0.84	0.79	0.86	0.82	0.90	0.86	0.88
									Proposed	0.96	0.90	0.99	0.94	0.96	0.99	0.97
<i>Leaf Scene</i>	B-1	0.65	0.25	0.77	0.44	0.77	0.77	0.77	DWT [37]	0.71	0.90	0.65	0.76	0.95	0.65	0.77
	B-2	0.63	0.29	0.73	0.46	0.77	0.73	0.75	PCA [33]	0.80	0.82	0.79	0.80	0.93	0.79	0.86
	B-3	0.67	0.63	0.68	0.66	0.86	0.68	0.76	AVG [36]	0.81	0.80	0.82	0.81	0.93	0.82	0.87
	B-4	0.68	0.24	0.81	0.44	0.78	0.81	0.79	DCT-LP [40]	0.92	0.81	0.95	0.88	0.94	0.95	0.94
	B-5	0.55	0.52	0.57	0.54	0.79	0.57	0.66	SHT [20]	0.76	0.55	0.82	0.67	0.86	0.82	0.84
	B-6	0.82	0.26	1.00	0.50	0.81	1.00	0.89	FPDE [34]	0.58	0.83	0.50	0.65	0.91	0.50	0.65
									Proposed	0.89	0.84	0.90	0.87	0.95	0.90	0.92
Average	Direct	0.74	0.63	0.76	0.67	0.86	0.76	0.80	Fusion	0.88	0.84	0.89	0.86	0.94	0.89	0.91

**Figure 11.** Direct vs. fusion-based segmentation errors of three scenes in the MSPI dataset.

5.3.5. Comparison of BS Accuracy: Proposed Method vs. Those in Literature

It is worth mentioning that the performances of these existing BS methods are not exactly comparable, as each reports its accuracy for a specific MSPI database. Also, the recognition accuracy values obtained from the fusion methods and color-mapping techniques used for segmentation vary.

By integrating an FFT with PCA for a fusion approach and then generating foreground masks for the segmentation task, the proposed BS method obtains higher accuracy, G-mean and F1-score values for the liquid (0.97, 0.97 and 0.98, respectively), food (0.96, 0.94 and 0.97, respectively) and leaf (0.89, 0.87 and 0.92, respectively) scenes in the MSPI dataset than those in the existing literature on foreground separation methods, as presented in Table 3.

Table 3. Comparison of performances of BS methods for three scenes in the MSPI dataset.

	Liquid Scene							Food Scene							Leaf Scene						
	AC	SP	SN	GM	PR	RC	F1S	AC	SP	SN	GM	PR	RC	F1S	AC	SP	SN	GM	PR	RC	F1S
Wolf [13]	0.69	0.00	0.84	0.06	0.79	0.84	0.82	0.57	0.99	0.38	0.61	0.99	0.38	0.55	0.65	0.40	0.73	0.54	0.80	0.73	0.76
Zhao [29]	0.95	1.00	0.94	0.97	1.00	0.94	0.97	0.81	0.38	1.00	0.62	0.78	1.00	0.88	0.44	0.93	0.28	0.51	0.93	0.28	0.44
Zhou [20]	0.67	1.00	0.60	0.78	1.00	0.60	0.75	0.70	0.85	0.62	0.73	0.90	0.62	0.74	0.58	0.40	0.64	0.51	0.78	0.64	0.70
Lu [59]	0.75	1.00	0.69	0.83	1.00	0.69	0.82	0.89	0.72	0.96	0.83	0.88	0.96	0.92	0.51	0.62	0.48	0.54	0.80	0.48	0.60
Proposed	0.97	0.98	0.97	0.97	1.00	0.97	0.98	0.96	0.90	0.99	0.94	0.96	0.99	0.97	0.89	0.84	0.90	0.87	0.95	0.90	0.92

Figure 12 shows the average performances of individual methods for scenes in the MSPI dataset. It is clear that the proposed method performs better than the others with a mean accuracy, G-mean and F1-score of 0.95, 0.93 and 0.97, respectively.

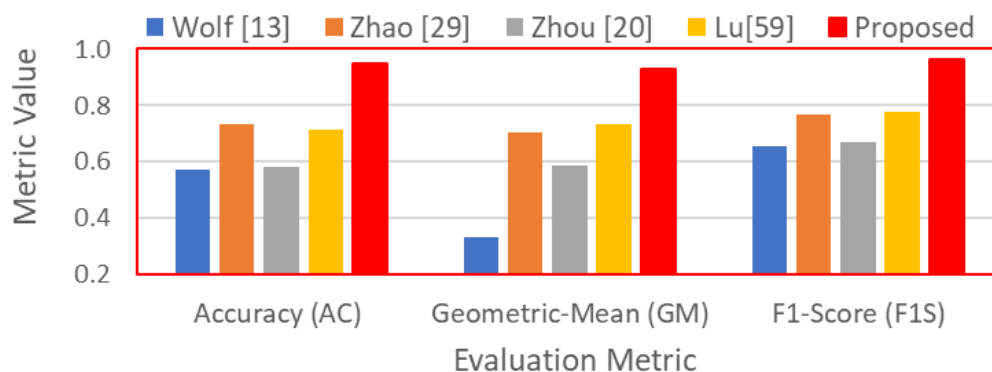


Figure 12. Average performances of individual methods for scenes in the MSPI dataset.

Figure 13 presents the error areas (purple and green) obtained from the individual methods for different scenes in the MSPI dataset. As can be seen, the proposed system reports fewer BS errors than the others. In the liquid scene, however, the background is completely separated by the proposed method, due to weak polarization in the jar cap's area; it has some errors. In the food (apple and banana) and leaf scenes, although their foregrounds are completely separated by the proposed method, there are some errors due to strong unpolarization in their background areas.

5.4. Computational Time Analysis

The experiments are carried out using MATLAB on a desktop with a hardware configuration of 16 GB RAM and 3.4 GHz Intel Core i7 CPU. Table 4 presents the running times (in seconds) of the individual methods for multiband fusion and polarimetric BS approach. As can be seen, the proposed system incurs slightly longer computational times than the PCA-based multiband fusion method, but shorter ones than the polarimetric BS, with its shortest mean running time demonstrating its effectiveness in terms of BS in MSPI.

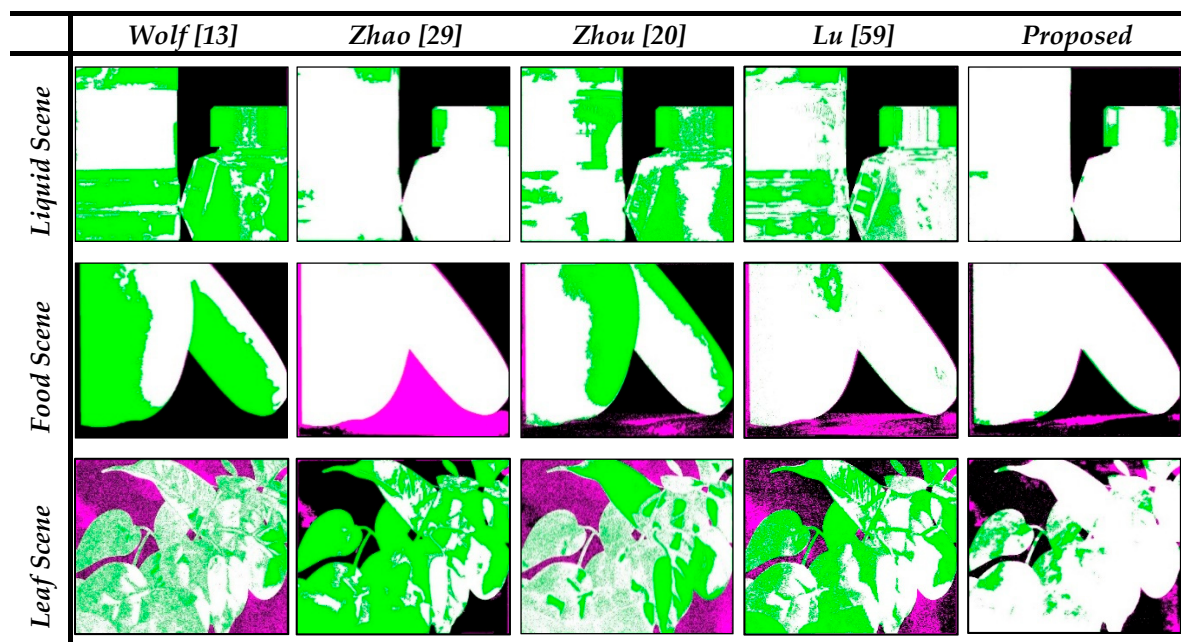


Figure 13. Comparison of segmentation errors of three scenes in the MSPI dataset.

Table 4. Analysis of computational times of different techniques for three scenes in the MSPI dataset.

	Running Time (Seconds) – Multiband Fusion					Running Time (Seconds) – Polarimetric BS			
	Liquid	Food	Leaf	Average		Liquid	Food	Leaf	Average
DWT [37]	55.76	57.01	57.51	56.76	Wolf [13]	10.74	10.14	11.35	10.74
PCA [33]	38.70	38.19	37.26	38.05	Zhao [29]	8.66	8.93	7.40	8.33
AVG [36]	63.05	63.47	62.94	63.15	Zhou [20]	12.38	8.95	12.57	11.30
DCT-LP [40]	56.40	56.00	56.70	56.37	Lu [59]	10.81	10.61	11.60	11.01
SHT [20]	162.06	163.48	160.23	161.92	Proposed	2.52	3.45	3.74	3.23
FPDE [34]	113.20	113.91	113.90	113.67					
Proposed	44.65	44.62	44.58	44.62					

6. Conclusions

In this paper, a two-fold BS approach involving multiband fusion followed by polarimetric BS is proposed for MSPI. Combining complementary information from spectral and polarimetric cues in the visible and NIR ranges improves a foreground's contrast and details. This framework and algorithm demonstrate the significance of multiband fusion through analyzing spectral reflectance, correlation and textural features. The proposed hybrid fusion method first decomposes each band image into low and high frequencies using an FFT. Then a PCA is performed, with the first PC of each frequency computed and combined with the others to obtain a final fused image. The proposed fusion algorithm predicts better fusion quality with fewer errors in terms of the MAPE, PSNR, COR and MI than the DWT, PCA, DCT-LP, SHT and FPDE techniques. After fusion, the polarimetric components are calculated through a Stokes vector analysis. Finally, four significant foreground masks are generated and combined with the polarimetric components to segment the complex background of a scene. The proposed BS algorithm is compared with four baseline approaches based on fusion, color mapping and fuzzy C-means clustering using an MSPI dataset. The experimental results illustrate the validity and efficiency of the proposed BS method based on diverse performance evaluation metrics. They also demonstrate that it significantly improves the mean accuracy, G-mean and F1-score to 0.95, 0.93, 0.97, respectively, for three scenes in the MSPI dataset compared with those in the existing literature referenced in this paper.

As an extension of this work, we will investigate an advanced BS method in which the background and foreground of a scene are mixed with polarized and unpolarized intensities. We will also address

the issues related to the specular reflection of the foreground area in a scene that may mistakenly be detected as the background. As it is known that the specular reflection in an area of strong light intensity can destroy the shape of an actual object; developing an algorithm for detecting and reconstructing a secular reflection in the MSPI dataset will be explored.

Author Contributions: Conceptualization, M.N.I. and M.T.; methodology, M.N.I. and M.T.; software, M.N.I.; validation, M.N.I. and M.T.; investigation, M.T. and M.P.; data curation, M.N.I.; writing—original draft preparation, M.N.I.; writing—review and editing, M.N.I., M.T. and M.P.; supervision, M.T. and M.P.; funding acquisition, M.P. All authors have read and agree to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research would like to acknowledge the author, Pierre-Jean Lapray for his technical answers regarding to the creation of the MSPI dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lapray, P.J.; Gendre, L.; Foulonneau, A.; Bigué, L. Database of polarimetric and multispectral images in the visible and NIR regions. In *Unconventional Optical Imaging, Proceedings of the SPIE, Strasbourg, France, 24 May 2018*; SPIE: Bellingham, WA, USA, 2018.
2. Yan, Q.; Shen, X.; Xu, L.; Zhuo, S.; Zhang, X.; Shen, L.; Jia, J. Crossfield joint image restoration via scale map. In *Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013*.
3. Schaul, L.; Fredembach, C.; Susstrunk, S. Color image dehazing using the near-infrared. In *Proceedings of the 16th IEEE International Conference on Image Processing (ICIP), Chiang Mai, Thailand, 7 November 2009*.
4. Salamati, N.; Larlus, D.; Csurka, G.; Süssstrunk, S. Semantic image segmentation using visible and near-infrared channels. In *Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012*.
5. Berns, R.S.; Imai, F.H.; Burns, P.D.; Tzeng, D.Y. Multispectral-based color reproduction research at the Munsell Color Science Laboratory. In *Electronic Imaging: Processing, Printing, and Publishing in Color, Proceedings of the SPIE, Zurich, Switzerland, 7 September 1998*; SPIE: Bellingham, WA, USA, 1998.
6. Thomas, J.B. Illuminant estimation from uncalibrated multispectral images. In *Proceedings of the 2015 Colour and Visual Computing Symposium (CVCS), Gjøvik, Norway, 25–26 August 2015*.
7. Motohka, T.; Nasahara, K.N.; Oguma, H.; Tsuchida, S. Applicability of green-red vegetation index for remote sensing of vegetation phenology. *Remote Sens.* **2010**, *2*, 2369–2387.
8. Dandois, J.P.; Ellis, E.C. Remote sensing of vegetation structure using computer vision. *Remote Sens.* **2010**, *2*, 1157–1176. [[CrossRef](#)]
9. Rfenacht, D.; Fredembach, C.; Süssstrunk, S. Automatic and accurate shadow detection using near-infrared information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1672–1678. [[CrossRef](#)] [[PubMed](#)]
10. Sobral, A.; Javed, S.; Ki Jung, S.; Bouwmans, T.; Zahzah, E.H. Online stochastic tensor decomposition for background subtraction in multispectral video sequences. In *Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 7–13 December 2015*.
11. Tyo, J.S.; Goldstein, D.L.; Chenault, D.B.; Shaw, J.A. Review of passive imaging polarimetry for remote sensing applications. *Appl. Opt.* **2006**, *45*, 5453–5469. [[CrossRef](#)] [[PubMed](#)]
12. Nayar, S.K.; Fang, X.-S.; Boulton, T. Separation of reflection components using color and polarization. *Int. J. Comput. Vis.* **1997**, *21*, 163–186. [[CrossRef](#)]
13. Wolff, L.B. Polarization-based material classification from specular reflection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 1059–1071. [[CrossRef](#)]
14. Atkinson, G.A.; Hancock, E.R. Shape estimation using polarization and shading from two views. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2001–2017.
15. Tan, J.; Zhang, J.; Zhang, Y. Target detection for polarized hyperspectral images based on tensor decomposition. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 674–678. [[CrossRef](#)]
16. Goudail, F.; Terrier, P.; Takakura, Y.; Bigue, L.; Galland, F.; DeVlaminck, V. Target detection with a liquid-crystal-based passive stokes polarimeter. *Appl. Opt.* **2004**, *43*, 274–282. [[CrossRef](#)]

17. Denes, L.J.; Gottlieb, M.S.; Kaminsky, B.; Huber, D.F. Spectropolarimetric imaging for object recognition. In Proceedings of the 26th AIPR Workshop: Exploiting New Image Sources and Sensors, Washington, DC, USA, 1 March 1998.
18. Romano, J.M.; Rosario, D.; McCarthy, J. Day/night polarimetric anomaly detection using SPICE imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 5014–5023. [[CrossRef](#)]
19. Islam, M.N.; Tahtali, M.; Pickering, M. Man-made object separation using polarimetric imagery. In Proceedings of the SPIE Future Sensing Technologies, Tokyo, Japan, 12–14 November 2019.
20. Zhou, P.C.; Liu, C.C. Camouflaged target separation by spectral-polarimetric imagery fusion with shearlet transform and clustering segmentation. In Proceedings of the International Symposium on Photoelectronic Detection and Imaging 2013: Imaging Sensors and Applications, Beijing, China, 21 August 2013.
21. Domadiya, P.; Shah, P.; Mitra, S.K. Fast and Accurate Foreground Background Separation for Video Surveillance. In Proceedings of the 6th International Conference on Pattern Recognition and Machine Intelligence (PReMI), Warsaw, Poland, 30 June–3 July 2015; Springer: Cham, Switzerland, 2015.
22. Bouwmans, T. Traditional and recent approaches in background modeling for foreground detection: An overview. *Comput. Sci. Rev.* **2014**, *11*, 31–66. [[CrossRef](#)]
23. Perazzi, F.; Pont-Tuset, J.; McWilliams, B.; Van Gool, L.; Gross, M.; Sorkine-Hornung, A. A benchmark dataset and evaluation methodology for video object segmentation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
24. Zitova, B.; Flusser, J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977–1000. [[CrossRef](#)]
25. Benezeth, Y.; Sidibé, D.; Thomas, J.B. Background subtraction with multispectral video sequences. In Proceedings of the IEEE International Conference on Robotics and Automation workshop on Nonclassical Cameras, Camera Networks and Omnidirectional Vision (OMNIVIS), Hong Kong, China, 11 June 2014; HAL-00986168f.
26. Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* **2019**, *45*, 153–178. [[CrossRef](#)]
27. Zhan, L.; Zhuang, Y.; Huang, L. Infrared and visible images fusion method based on discrete wavelet transform. *J. Comput.* **2017**, *28*, 57–71. [[CrossRef](#)]
28. Li, H.; Liu, L.; Huang, W.; Yue, C. An improved fusion algorithm for infrared and visible images based on multi-scale transform. *Infrared Phys. Technol.* **2016**, *74*, 28–37. [[CrossRef](#)]
29. Zhao, Y.Q.; Zhang, L.; Zhang, D.; Pan, Q. Object separation by polarimetric and spectral imagery fusion. *Comput. Vis. Image Underst.* **2009**, *113*, 855–866. [[CrossRef](#)]
30. Weinberger, M.J.; Seroussi, G.; Sapiro, G. The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS. *IEEE Trans. Image Process.* **2000**, *12*, 1309–1324. [[CrossRef](#)]
31. Rizzo, F.; Carpentieri, B.; Motta, G.; Storer, J.A. Low-Complexity Lossless Compression of Hyperspectral Imagery via Linear Prediction. *IEEE Signal Process. Lett.* **2005**, *12*, 138–141. [[CrossRef](#)]
32. Seki, M.; Wada, T.; Fujiwara, H.; Sumi, K. Background subtraction based on cooccurrence of image variations. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003.
33. Naidu, V.P.S.; Raol, J.R. Pixel-level image fusion using wavelets and principal component analysis. *Def. Sci. J.* **2008**, *58*, 338. [[CrossRef](#)]
34. Bavirisetti, D.P.; Xiao, G.; Liu, G. Multi-sensor image fusion based on fourth order partial differential equations. In Proceedings of the 2017 20th International Conference on Information Fusion (Fusion), Xi'an, China, 10–13 July 2017.
35. Lapray, P.J.; Thomas, J.B.; Gouton, P.; Ruichek, Y. Energy balance in Spectral Filter Array camera design. *J. Eur. Opt. Soc.-Rapid Publ* **2017**, *13*, 1–13. [[CrossRef](#)]
36. Malviya, A.; Bhirud, S.G. Image fusion of digital images. *Int. J. Recent Trends Eng.* **2009**, *2*, 146.
37. Jian, M.; Dong, J.; Zhang, Y. Image fusion based on wavelet transform. In Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD), Qingdao, China, 30 July–1 August 2017.
38. Raju, V.B.; Sankar, K.J.; Naidu, C.D.; Bachu, S. Multispectral image compression for various band images with high resolution improved DWT SPIHT. *Int. J. Signal Process. Image Process. Pattern Recognit.* **2016**, *9*, 271–286. [[CrossRef](#)]

39. Desale, R.P.; Verma, S.V. Study and analysis of PCA, DCT & DWT based image fusion techniques. In Proceedings of the 2013 International Conference on Signal Processing, Image Processing & Pattern Recognition, Coimbatore, India, 7–8 February 2013.
40. Naidu, V.P.S.; Elias, B. A novel image fusion technique using DCT based Laplacian pyramid. *Int. J. Inventive Eng. Sci. (IJIES)* **2013**, *1*, 1–9.
41. Liu, R.; Ruichek, Y.; El Bagdouri, M. Extended Codebook with Multispectral Sequences for Background Subtraction. *Sensors* **2019**, *19*, 703. [[CrossRef](#)] [[PubMed](#)]
42. Zhao, J.; Sen-ching, S.C. Human segmentation by geometrically fusing visible-light and thermal imageries. *Multimed. Tools Appl.* **2012**, *73*, 61–89. [[CrossRef](#)]
43. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 898–916. [[CrossRef](#)] [[PubMed](#)]
44. Rother, C.; Kolmogorov, V.; Blake, A. “GrabCut” interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph. (TOG)* **2004**, *23*, 309–314. [[CrossRef](#)]
45. Tron, R.; Vidal, R. A benchmark for the comparison of 3-d motion segmentation algorithms. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007.
46. Cheng, J.; Tsai, Y.H.; Wang, S.; Yang, M.H. SegFlow: Joint learning for video object segmentation and optical flow. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
47. Jain, S.D.; Xiong, B.; Grauman, K. Fusionseg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hawaii, HI, USA, 21–26 July 2017.
48. Rother, C.; Minka, T.; Blake, A.; Kolmogorov, V. Cosegmentation of image pairs by histogram matching—Incorporating a global constraint into MRFs. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06), New York, NY, USA, 17–22 June 2006.
49. Zhu, H.; Meng, F.; Cai, J.; Lu, S. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *J. Vis. Commun. Image Represent.* **2016**, *34*, 12–27. [[CrossRef](#)]
50. St-Charles, P.L.; Bilodeau, G.A.; Bergevin, R. Online mutual foreground segmentation for multispectral stereo videos. *Int. J. Comput. Vis.* **2019**, *127*, 1044–1062. [[CrossRef](#)]
51. Jeong, S.; Lee, J.; Kim, B.; Kim, Y.; Noh, J. Object segmentation ensuring consistency across multi-viewpoint images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 2455–2468. [[CrossRef](#)] [[PubMed](#)]
52. Djelouah, A.; Franco, J.S.; Boyer, E.; Le Clerc, F.; Pérez, P. Sparse multi-view consistency for object segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1890–1903. [[CrossRef](#)]
53. Riklin-Raviv, T.; Sochen, N.; Kiryati, N. Shape-based mutual segmentation. *Int. J. Comput. Vis.* **2008**, *79*, 231–245. [[CrossRef](#)]
54. Bleyer, M.; Rother, C.; Kohli, P.; Scharstein, D.; Sinha, S. Object stereo-joint stereo matching and object segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 20–25 June 2011.
55. Del-Blanco, C.R.; Mantecón, T.; Camplani, M.; Jaureguizar, F.; Salgado, L.; García, N. Foreground segmentation in depth imagery using depth and spatial dynamic models for video surveillance applications. *Sensors* **2014**, *14*, 1961–1987. [[CrossRef](#)] [[PubMed](#)]
56. Fernandez-Sanchez, E.J.; Diaz, J.; Ros, E. Background subtraction based on color and depth using active sensors. *Sensors* **2013**, *13*, 8895–8915. [[CrossRef](#)]
57. Zhou, X.; Liu, X.; Jiang, A.; Yan, B.; Yang, C. Improving video segmentation by fusing depth cues and the visual background extractor (ViBe) algorithm. *Sensors* **2017**, *17*, 1177. [[CrossRef](#)]
58. Zhang, C.; Li, Z.; Cai, R.; Chao, H.; Rui, Y. Joint Multiview segmentation and localization of RGB-D images using depth-induced silhouette consistency. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
59. Lu, X.; Peng, F.; Li, G.; Xiao, Z.; Hu, T. Object Segmentation for Linearly Polarimetric Passive Millimeter Wave Images Based on Principle Component Analysis. *Prog. Electromagn. Res.* **2017**, *61*, 169–176. [[CrossRef](#)]
60. Lapray, P.J.; Gendre, L.; Foulonneau, A.; Bigué, L. A Database of Polarimetric and Multispectral Images in the Visible and NIR Regions. In Proceedings of the SPIE Photonics Europe, Strasbourg, France, 22–26 April 2018.

61. Richards John, A.; Xiuping, J. *Remote Sensing Digital Image Analysis: An Introduction*, 4th ed.; Springer: Berlin/Heidelberg, Germany, 1999.
62. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern* **1973**, SMC-3, 610–621. [[CrossRef](#)]
63. Zhang, X.; Cui, J.; Wang, W.; Lin, C. A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. *Sensors* **2017**, *17*, 1474. [[CrossRef](#)]
64. Stokes, G.G. On the composition and resolution of streams of polarized light from different sources. *Trans. Camb. Philos. Soc.* **1851**, *9*, 399.
65. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
66. Somvanshi, S.S.; Kunwar, P.; Tomar, S.; Singh, M. Comparative statistical analysis of the quality of image enhancement techniques. *Int. J. Image Data Fusion* **2017**, *9*, 131–151. [[CrossRef](#)]
67. Haghighat, M.B.A.; Aghagolzadeh, A.; Seyedarabi, H. A non-reference image fusion metric based on mutual information of image features. *Comput. Electr. Eng.* **2011**, *37*, 744–756. [[CrossRef](#)]
68. Rani, K.; Sharma, R. Study of different image fusion algorithm. *Int. J. Emerg. Technol. Adv. Eng.* **2013**, *3*, 288–291.
69. Chiu, S.Y.; Chiu, C.C.; Xu, S.S.D. A Background Subtraction Algorithm in Complex Environments Based on Category Entropy Analysis. *Appl. Sci.* **2018**, *8*, 885. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).