Signal, Image and Video Processing
https://doi.org/10.1007/s11760-018-1239-3

ORIGINAL PAPER

Vegetation segmentation based on variational level set using multi-channel local wavelet texture and color

Tiejun Yang[1] · Yaowen Chen[1] · Zhun Fan[1]

Received: 14 January 2017 / Revised: 25 October 2017 / Accepted: 11 January 2018

Abstract
The existing spectrum index-based methods for detecting vegetation coverage suffer from an over-dependence on spectrum. To address these issues, this paper proposes a graph cut-based variational level set segmentation algorithm that combines multi-channel local wavelet texture (MCLWT) and color. First, the prior color is generated by automatic estimation based on the mathematical morphology with a color histogram. Then, local wavelet texture features are extracted using a multi-scale and orientation Gabor wavelet transformation followed by local median and entropy filtering. Next, in addition to the energy of color, that of MCLWT is integrated into the variational level set model based on kernel density estimation. Consequently, all energies are integrated into the graph cut-based variational level set model. Finally, the proposed energy functional is made convex to obtain a global optimal solution, and a primal-dual algorithm with global relabeling is adopted to accelerate the evolution of the level sets. A comparison of the segmentation results from our proposed algorithm and other state-of-the-art algorithms showed that our algorithm effectively reduces the over-dependence on color and yields more accurate results in detecting vegetation coverage.

Keywords Variational level set · Detecting vegetation coverage · Local wavelet texture · Graph cut

1 Introduction

Detection of vegetation coverage (DVC) is a highly important aspect of machine vision-based agricultural automation applications. DVC methods are generally based on visible spectral indexes (VSIs), including the excess green index (ExG) [1], excess red index (ExR) [2], color index of vegetation extraction (CIVE) [3], excess green minus excess red index (ExGR) [4] and the vegetative index (VEG) [5]. The advantages of VSI-based methods are that they can accurately segment vegetation, soil and sky and that visible spectral sensors are not very expensive. Other DVC methods based on multi-spectral images have advantages of wider coverage and better accuracy than VSIs, although only at considerably higher cost. However, both the VSI-based and multi-spectral-based methods share a common deficiency: they have an over-dependence on spectrum or color while neglecting wider spatial information such as texture.

Therefore, combinations of spectral indexes and other methods have been proposed to improve the accuracy of DVC. Ponti [6] presented a combination of spectral indexes and mean shift (MS) to segment balloon-captured remote sensing images, including MS+CIVE, MS+ExG and MS+VVI. To improve the robustness of DVC in the presence of illumination variations or plant canopy shadows, Bai et al. [7] used particle swarm optimization (PSO) clustering and morphology modeling to segment vegetation from soil in color images acquired by an off-the-shelf digital camera affixed to an image acquisition device. Unsupervised methods such as MS or PSO clustering help in determining an appropriate threshold for spectral indexes from images, although errors always occur when the spectra of the foreground and background overlap.

In addition to spectrum information or color, other cues such as texture and shape have been considered in DVC. Feng et al. [8] used colors in the RGB color space and textures based on the gray-level co-occurrence matrix (GLCM) as the input for a random forest to map urban vegetation from UAV-captured remote sensing images. Aksoy et al. [9]

✉ Zhun Fan
zfan@stu.edu.cn

[1] Department of Electrical Engineering, Shantou University, Shantou, Guangdong, People's Republic of China

Published online: 24 January 2018

exploited both spectral- and Gabor filter-based texture and shape properties to detect hedgerows using decision-making methods. Texture features based on GLCM have a problem determining the scale and orientation of objects, while multi-resolution analysis-based methods such as Gabor filters can extract the textural features of vegetation at different scales and orientations. Machine learning methods are also often used [10], but these methods rely on the availability of numerous good training samples, and they are prone to over-fitting and other issues.

Image segmentation methods based on the variational level set (VLS) model [11] have attracted a substantial amount of attention [12,13]. Unger et al. [14] proposed a variational level set segmentation model based on graph cut and total variation (GCTV) that further exploited the primal-dual and global relabeling (PDGRL) method to minimize the primal-dual energy interval. However, GCTV uses only prior color as an external energy and, consequently, suffers from an over-reliance on it. The VLS segmentation model has a different problem: a non-convex functional leads to a local optimal solution. To eliminate the non-convex factors, Chan et al. [15] proposed convexity transform algorithms for certain non-convex minimization problems. These transform algorithms are called algorithms for finding global minimizers (AFGM).

Hence, we propose an improved method based on GCTV to perform DVC that improves the segmentation accuracy by integrating more features, such as texture and automatically generated prior color, and by convexity transforms of the non-convex energy functionals. In our method, multi-channel textures are extracted by the Gabor wavelet transform combined with local mean and entropy filters in a process called multi-channel local wavelet texture (MCLWT). MCLWT was shown to be effective in vegetation segmentation in our previous work [16]. Prior color is obtained automatically by mathematical morphology and color histogram. Then, these features are integrated into the energy functional of GCTV to evolve the level sets.

The rest of this paper is organized as follows: Sect. 2 discusses the VLS model based on GCTV and the convexity transform method. Then, the proposed algorithm is described in detail in Sect. 3. Section 4 presents the experimental results and a discussion. Finally, Sect. 5 provides conclusions.

2 Background

The graph cut divides all pixels in an image into three categories: source, sink and boundary pixels. Only when a pixel is classified into the category to which it truly belongs is the cut set minimized. The graph cut principle can be used to model the energy functional of a level set [14,17].

Based on GCTV, Bresson et al. [17] also presented a VLS model and defined a boundary weight function for the total variation. In these methods, the source pixels indicate the object, whose cost is denoted by c_f. The sink pixels indicate the background; their cost is denoted by c_b. The boundary pixels separate the object from the background; their cost is denoted by c_e. Using the graph cut principle, an image segmentation problem can be viewed as one of classifying all the pixels into the above three classes. A minimal cut set is obtained only when all pixels have been correctly classified. Thus, the segmentation problem can be viewed as minimizing the following energy functional:

$$E_{\text{gctv}} = \int_{\Omega} c_e \left|\nabla u\right| + \int_{\Omega} c_b u \mathrm{d}x + \int_{\Omega} c_f \left(1-u\right) \mathrm{d}x \quad s.t. u\left(x\right) \in \{0,1\}, \tag{1}$$

where $\Omega \subset \mathbf{R}^2$ is an open set representing the image domain and u is a characteristic function.

To obtain a minimum of the functional in Eq. (1), c_e is set to be inversely proportional to the boundary gradient and c_f and c_b are estimations of the probability distribution of the object color and the background color, respectively. These estimations can be obtained automatically using the method described in Sect. 3.1.

One problem in GCTV is that the functional in Eq. (1) contains a binary function u. In other words, it is a non-convex problem that has only a local optimal solution. Therefore, the problem must be transformed into a convex one. We use a convexity transform method proposed by Chan et al. [15] to solve this problem. We refer readers to [15] for details.

3 Method

Our method includes three major processes: feature extraction, model adaptation and post-processing, as shown in Fig. 1. To improve segmentation accuracy by integrating more features, three kinds of features—automatically estimated prior color, the a^* channel color from the $L^*a^*b^*$ color space and multi-channel local wavelet texture (MCLWT)—are extracted. To utilize the extracted features as supplementary energies in the level set model, the features are integrated as external energies into the graph cut-based level set segmentation model. To achieve global optimization more conveniently, the proposed energy functional is transformed into a convexity functional using Chan's convexity transform method [15] as described above. We also adopt PDGRL to speed up the level set evolution in our method. Finally, in the post-processing stage, mathematical morphology operations are used to erase tiny islands and fill in small holes in the segmentation results.

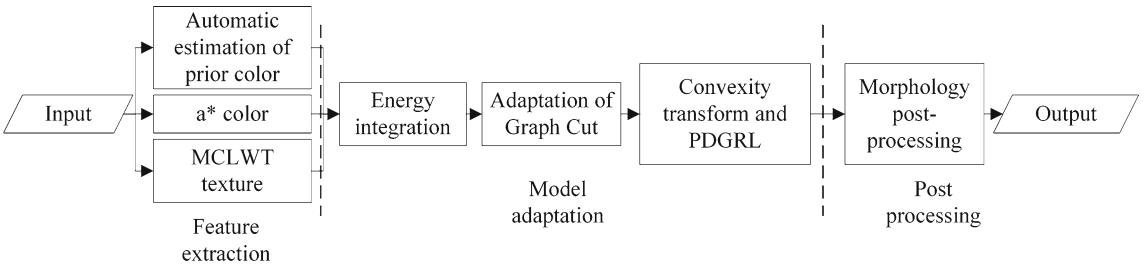

Fig. 1 Flowchart of the proposed method, including pre-processing, segmentation and post-processing

3.1 Automatic estimation of prior color

To obtain prior color, manual labeling is widely used in [14,18]. However, manual labeling requires human interaction and its efficiency is low—especially for large data sets. To address this problem, automatic estimation of prior color (AEPC) is required. Here, we propose an approach based on CIVE and mathematical morphology that can automatically generate approximate probability distributions of vegetation and non-vegetation. CIVE is used here to generate estimated segmentation results. Then, mathematical morphology is applied to obtain the primary components of the objects and background. These processes are described below:

1. A given image $f(x, y)$ is first segmented using CIVE to obtain binary segmentation results. CIVE is computed as follows:

$$f_{\text{CIVE}} = 0.441R - 0.811G + 0.385B + 18.787, \quad (2)$$

where R, G and B are the values of the color channels of $f(x, y)$ in the RGB color space. Then Otsu thresholding is used to transform f_{CIVE} into a binary value, f_{bw}.
2. To obtain the AEPC for vegetation, we use the following steps:
   (a) A mathematical morphology *shrink* operator is applied to the inverse of f_{bw}, shrinking the areas of vegetation to connected lines or points. The results are denoted by f_{obj_shk}.
   (b) Next, a mathematical morphology *thicken* operator is applied, making f_{obj_shk} one pixel thicker.
   (c) Finally, a mathematical morphology *open* operator is imposed to reserve the most likely areas of vegetation using a structuring element (SE), which is a line 10 pixels long at a 2-degree angle. The result is denoted by f_{obj}.
3. To obtain the AEPC for non-vegetation, we use the following steps:
   (a) A *shrink* operator is applied to f_{bw}.

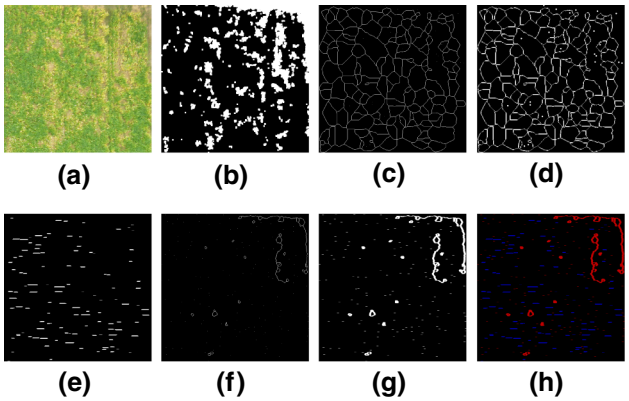

Fig. 2 An AEPC example: **a** the original image from DB1, **b** the segmentation results by CIVE, **c** the shrink operator results on the inverse of **b**, **d** the thicken operator results on **c**, **e** the prior color estimation results of vegetation by the open operator on **d**, **f** the shrink operator results on **b**, **g** the prior color estimation results of non-vegetation by the dilate operator on **f**, and **h** a visualized image of AEPC

   (b) Then, a mathematical morphology *dilate* operator is used to find the likely areas of non-vegetation using an SE line 8 pixels long at a 2-degree angle. The result is denoted by f_{bkg}.
4. Using f_{obj} and f_{bkg} as masks [14], color histograms of the vegetation h_f and non-vegetation h_b are calculated.
5. The histograms are smoothed by a Gaussian filter and then linearly interpolated at each pixel's intensity value from the original image, yielding the final prior color histograms, H_f and H_b, respectively.

Figure 2 shows an example of AEPC. The visualized image of AEPC f_{AEPC} in Fig. 2h is composed as follows:

$$f_{\text{AEPC}} = [R, G, B],$$
$$R = f_{bkg}, \quad G = 0, \quad B = f_{obj}. \quad (3)$$

3.2 Multi-channel local wavelet texture

Because vegetation usually has different textural features than those of surrounding soil, water, etc., we introduce

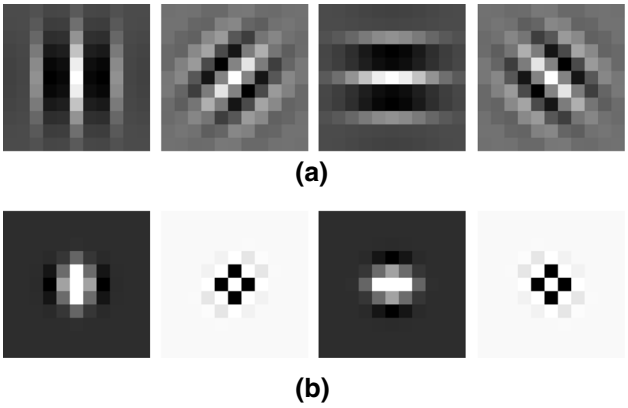

Fig. 3 Gabor filter bank with 2 frequencies and 4 orientations ($\theta = 0, \pi/4, \pi/2, 3\pi/4$): **a** $f_0 = 0.2$, **b** $f_0 = 0.5$

a texture feature called multi-channel local wavelet texture (MCLWT) into the proposed model for vegetation segmentation. A Gabor wavelet transform is insensitive to illumination variation and geometric transformation. Therefore, we use it here to transform the original image into a multi-resolution frequency domain. The two-dimensional Gabor wavelet transform $w(\cdot,\cdot)$ for an image $f(x, y)$ with a frequency of f_0 and an orientation of θ can be written as follows:

$$\begin{aligned} w_{f_0,\theta}(u, v) &= \iint f(x, y)\, h_G^*(u - x, v - y)\, \mathrm{d}x\mathrm{d}y, \\ h_G(x, y) &= \frac{f_0^2}{\pi\sigma_x\sigma_y} \exp\left(-\left(\frac{f_0^2 x_r^2}{\sigma_x^2} + \frac{f_0^2 y_r^2}{\sigma_y^2}\right)\right) \\ &\quad \cdot \exp\left[j2\pi f_0 x_r\right], \\ x_r &= x\cos(\theta) + y\sin(\theta), \\ y_r &= -x\sin(\theta) + y\cos(\theta), \end{aligned} \tag{4}$$

where * denotes the complex conjugate, $h_G(\cdot)$ is the Gabor wavelet function, and σ_x and σ_y are the scaling parameters along the wave and perpendicular to the wave, respectively. Here, $\sigma_x = \sigma_y = 1$, and we use a bank of filters defined in a specific frequency range and orientation. An example of this filter bank with 2 frequencies ($f_0 = 0.2, 0.5$) and 4 orientations ($\theta = 0, \pi/4, \pi/2, 3\pi/4$) is shown in Fig. 3, which yields 8 channels after the transform. Then, the local wavelet texture feature $\mathbf{f}_{\mathrm{MCLWT}}$ is extracted based on the Gabor wavelet transform defined in Eq. (4):

$$\begin{aligned} \mathbf{f}_{\mathrm{MCLWT}} &= \left(\mathbf{f}_{\mathrm{med}}, \mathbf{f}_{\mathrm{etp}}\right), \\ \mathbf{f}_{\mathrm{med}} &= \mathrm{median}\left(\underset{(p,q)\in W}{w_{f_0,\theta}(p, q)}\right), \\ \mathbf{f}_{\mathrm{etp}} &= \mathrm{entropy}\left(\underset{(p,q)\in W}{w_{f_0,\theta}(p, q)}\right), \end{aligned} \tag{5}$$

where median $(\cdot)$ and entropy $(\cdot)$ are functions of the median and entropy filters of a filtered window W with a size of n in $w_{f_0,\theta}$, respectively. Using this feature extraction approach, $\mathbf{f}_{\mathrm{MCLWT}}$ has 16 channels of MCLWT-based textures in total.

3.3 Integration of color and texture energies

First, the energy of prior color E_{p} can be produced by AEPC and integrated directly into the proposed energy functional using Eq. (1) as follows:

$$\begin{aligned} E_{\mathrm{p}} &= \int_\Omega \left(c_b - c_f\right) u\mathrm{d}x\mathrm{d}y, \\ &\quad s.t.\ u(x, y) \in \{0, 1\}, \end{aligned} \tag{6}$$

where $c_b = H_b$ and $c_f = H_f$. The item $\int_\Omega c_f \mathrm{d}x$ at the right of Eq. (1) is omitted because it is independent of u.

Second, the energy of the a^* color channel in the CIE $L^*a^*b^*$ color space, denoted by E_c, is added as a fidelity term based on the idea of the C-V model and [19], which is given by

$$\begin{aligned} E_c &= \int_\Omega \left(\begin{array}{l} H(\phi)\left(c'_1 - f_{a*}(x, y)\right)^2 \\ \quad + (1 - H(\phi))\left(c'_2 - f_{a*}(x, y)\right)^2 \end{array} \right) \mathrm{d}x\mathrm{d}y, \\ c'_1 &= \int_\Sigma f_{a*}(x, y)\mathrm{d}x\mathrm{d}y \Big/ |\Sigma|, \\ c'_2 &= \int_{\Omega\setminus\Sigma} f_{a*}(x, y)\mathrm{d}x\mathrm{d}y \Big/ |\Omega\setminus\Sigma|, \end{aligned} \tag{7}$$

where f_{a*} is the a^* color channel. Because $H(\phi) \in \{0, 1\}$, it can be replaced by u. Then, E_c is transformed as follows:

$$\begin{aligned} E_c &= \int_\Omega \left(\begin{array}{l} (1 - u)\left(c'_1 - f_{a*}(x, y)\right)^2 \\ \quad + u\left(c'_2 - f_{a*}(x, y)\right)^2 \end{array} \right) \mathrm{d}x\mathrm{d}y \\ &= \int_\Omega \left(\begin{array}{l} \left(c'_2 - f_{a*}(x, y)\right)^2 \\ \quad - \left(c'_1 - f_{a*}(x, y)\right)^2 \end{array} \right) u\mathrm{d}x\mathrm{d}y. \end{aligned} \tag{8}$$

The item $\int_\Omega \left(\left(c'_1 - f_{a*}(x, y)\right)^2\right) \mathrm{d}x\mathrm{d}y$ at the right of Eq. (8) is omitted because it is independent of u.

Third, the texture energy based on MCLWT is integrated into our model. There are several approaches for integrating texture into a variational level set model, including mean value-based approaches such as the C-V model [20], the Gaussian mixed model (GMM)-based [21], kernel density estimation (KDE)-based [22], and so on. However, because the mean value-based approaches are not reliable enough for textured objects, and GMM-based approaches are well known for over-fitting, we adopt KDE to build the texture energy based on MCLWT using the following steps:

(a) For each channel of $\mathbf{f}^i_{\mathrm{MCLWT}}$, $i = 1, 2, \ldots, C$, KDE is used to estimate each channel's probability distribution $p_j\left(\mathbf{f}^i_{\mathrm{MCLWT}}\right), j = 1, 2$, where C is the total number of channels of $\mathbf{f}_{\mathrm{MCLWT}}$, and p_1 and p_2 denote the probability distributions of the object and background, respectively.

(b) Calculate the total texture probability distributions of the object and background by

$$t_f = \sum_i p_1\left(\mathbf{f}^i_{\mathrm{MCLWT}}\right), \quad t_b = \sum_i p_2\left(\mathbf{f}^i_{\mathrm{MCLWT}}\right), \tag{9}$$

where t_f is the MCLWT texture energy of the object, and t_b is the texture energy of the background.

Subsequently, the MCLWT texture energy is integrated into the GCTV model. The proposed texture energy functional E_{t} is

$$\begin{aligned} E_{\mathrm{t}} &= \int_\Omega t_b u \mathrm{d}x\mathrm{d}y + \int_\Omega t_f\,(1-u)\,\mathrm{d}x\mathrm{d}y \\ &= \int_\Omega (t_b - t_f) u \mathrm{d}x\mathrm{d}y. \end{aligned} \tag{10}$$

The item $\int_\Omega t_f \mathrm{d}x\mathrm{d}y$ at the right of Eq. (10) is omitted because it is independent of u. Therefore, the final functional of our model E_{MCLWT} is given by

$$E_{\mathrm{MCLWT}} = \int_\Omega c_e\,|\nabla u|\,\mathrm{d}x\mathrm{d}y + \lambda_1 E_{\mathrm{p}} + \lambda_2 E_c + \lambda_3 E_{\mathrm{t}}, \tag{11}$$

where λ_1, λ_2 and λ_3 are positive constants. The first term to the right of Eq. (11) is the total variation; the others are the fidelity terms of prior color, a^* color and MCLWT texture energies, respectively. Using Eq. (11), the color and texture energies are integrated into the GCTV framework in a unified manner.

3.4 Global optimization based on convexity transform

To achieve global optimization, we adopt AFGM to eliminate the non-convex factors in the proposed functional defined by Eq. (11). First, $u \in \{0, 1\}$ is relaxed by $u : \mathbf{R}^2 \to [0, 1]$. Then, a globally optimal solution, u_λ, can be achieved. Finally, a binarized solution of the original problem is obtained using the threshold method. The details can be found in [15].

For fast convergence, we use PDGRL [14] to minimize the energy functional in Eq. (11). The PDGRL algorithm reduces the energy interval between the original and its dual problem using iterative and threshold methods, thereby increasing the algorithm's efficiency.

4 Experiments

Our experiments were performed on a PC with a 2.5 GHz Intel Core 2 Q8300 processor and 4 GB of 800 MHz DDR2 RAM under MATLAB 2013b. We used two image databases of vegetation, DB1 and DB2 from [6] and [7], respectively, to perform segmentation accuracy comparisons. DB1 contains 12 images of bean fields, while DB2 contains 200 images of rice fields. The ground truths for all the images were manually labeled by specialists. The default parameter settings were as follows: $n = 11$, $\lambda_1 = \lambda_2 = 1$, and $\lambda_3 = 100$.

4.1 Accuracy evaluation

We used the metric from [6], denoted as Acc_1, to compare the proposed segmentation algorithms on the DB1 images. The Acc_1 metric is defined as follows:

$$\begin{aligned} \mathrm{Acc}_1 &= 1 - \textstyle\sum_{i=1}^{r} \mathrm{Err}_i \Big/ 2r \\ \mathrm{Err}_i &= e_{i,1} + e_{i,2}, \\ e_{i,1} &= \frac{\mathrm{FP}_i}{N - N_i}, \quad e_{i,2} = \frac{\mathrm{FN}_i}{N_i}, \quad i = 1, \ldots, r, \end{aligned} \tag{12}$$

where FP is the number of false positives, and FN is the number of false negatives. FP_i is the number of pixels of region $j \neq i$ misclassified into region i, FN_i represents the pixels of region i misclassified into other regions, N is the number of pixels in a given image, and N_i denotes the pixels that belong to region i. For binary segmentation, here $r = 2$.

We use the metric called the Jaccard index (J) from [7] to evaluate the segmentation performance on the DB2 images. The Jaccard index is calculated as follows:

$$J = \mathrm{TP} \Big/ (\mathrm{TP} + \mathrm{FP} + \mathrm{FN}). \tag{13}$$

where P is the number of true positive pixels.

In addition to the above measures, the metrics of sensitivity (Sens), specificity (Spec), accuracy (Acc_2), precision (Prec) and F-measure (F) were calculated in our experiments for further quantitative comparison. The formulas of these metrics are given in Eqs. (14)–(18), respectively:

$$\mathrm{Sens} = \mathrm{TP} \Big/ \mathrm{TP} + \mathrm{FN}, \tag{14}$$

$$\mathrm{Spec} = \mathrm{TN} \Big/ \mathrm{FP} + \mathrm{TN}, \tag{15}$$

$$\mathrm{Acc}_2 = (\mathrm{TP} + \mathrm{TN}) \Big/ (\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}), \tag{16}$$

$$\mathrm{Prec} = \mathrm{TP} \Big/ \mathrm{TP} + \mathrm{FP}, \tag{17}$$

$$F = 2 * \text{Prec} * \text{Sens} / (\text{Prec} + \text{Sens}). \quad (18)$$

where TN is the number of true negative pixels.

4.2 Numerical tests

First, the proposed algorithm was compared to 8 state-of-the-art vegetation segmentation algorithms on the DB1 images. These algorithms cover the major types of effective segmentation techniques. Among these algorithms, the one proposed by Cheng et al. [23] is clustering-based; the ones presented by Kataoka et al. [3], Gée et al. [24] and VVI [6] are VSI-based; and the mean shift combined with VVI (MS+VVI), mean shift combined with ExG (MS+ExG) and mean shift combined with CIVE (MS+CIVE) algorithms proposed in [6] are hybrid methods that combine the preceding two types of methods. The GCTV algorithm [14] is the most similar to ours. Table 1 lists the accuracy scores of the comparison results measured by Acc_1 as defined in Eq. (12) for the DB1 images, where μ and σ denote the mean and standard deviation of Acc_1, respectively. As shown in Table 1, CIVE has the lowest accuracy (66.1%). This result occurs because CIVE simply uses each pixel's color to perform segmentation. The algorithms that combine clustering and VSI such as MS+ExG and MS+CIVE use the mean values of regional features to merge or split areas based on the segmentation results of the VSI algorithms. These hybrid algorithms achieve higher accuracies than do the VSI-based ones, 85 and 86.4%, respectively. Using AEPC, GCTV achieves a high accuracy of 88.9%. Our method integrates both color and textures in the proposed level set segmentation model; consequently, it achieves the highest accuracy (89.1%) and the lowest standard deviation (5.7%).

A sample image from DB1 and the segmentation results of ExG, CIVE, MS+ExG, MS+CIVE and our method are shown in Fig. 4. These compared algorithms were claimed to be superior to other VSI-based algorithms in [6]. Figure 4a, b shows the original image and the ground truth, respectively. From a visual comparison, the index methods (ExG and CIVE) obtain the worst results because they contain numerous FPs or FNs, as shown in Fig. 4c, d. The hybrid methods (MS+ExG and MS+CIVE) perform better than the indexed methods, as depicted in Fig. 4e, f, respectively, because they consider regional color features during segmentation. Our results are quite similar to those of the hybrid methods, although some details are different, as demonstrated in Fig. 5.

Table 1 Comparison data of the state-of-the-art algorithms using DB1

Algorithm	μ (%)	σ (%)
CIVE [3]	66.1	11.9
ExG [24]	78.5	8.6
MS [23]	76.5	10.7
VVI [6]	70.4	10.5
MS+VVI [6]	72.6	13.3
MS+ExG [6]	85	8.4
MS+CIVE [6]	86.4	7.2
GCTV [14]	88.9	6.1
Proposed	**89.1**	**5.7**

Bold values indicate the best indices

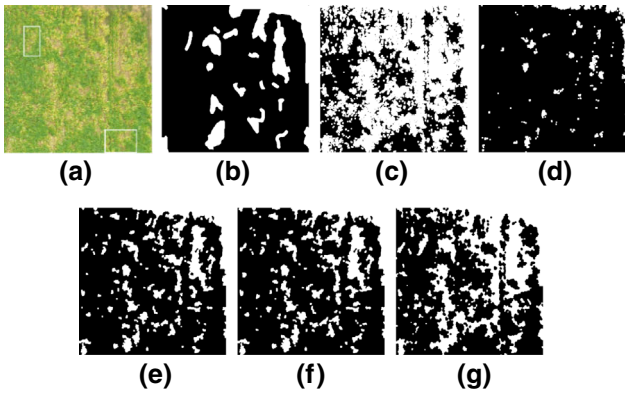

Fig. 4 Comparison of our algorithm with 4 other state-of-the-art algorithms reported in [6]. The black pixels denote vegetation, and the white pixels represent the background. **a** An original image from DB1, **b** ground truth, **c**–**g** the segmentation results by ExG, CIVE, MS+ExG, MS+CIVE and our algorithm, respectively

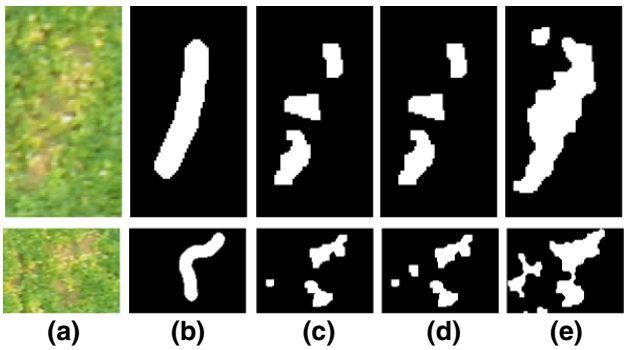

Fig. 5 Comparison of our algorithm and the hybrid algorithms on an image of DB1 [6]: **a** two areas cropped from the original image are marked by the two labeled white boxes in Fig. 4a. The first row refers to box No. 1, and the second row refers to box No. 2. **b** Ground truth. **c**–**e** the segmentation results of the areas by MS+ExG, MS+CIVE and our method, respectively

Figure 5 shows two areas cropped from the original image marked by the two labeled white boxes in Fig. 4a. The 1st row of Fig. 5 refers to box No. 1 and the 2nd row to box No. 2. As shown, the hybrid methods segment more background pixels as vegetation. Apparently, the background areas segmented by our method are more precise compared to those obtained the other two methods. Note that the ground truth obtained by manual labeling tends to lose some details of the

Table 2 Segmentation results of state-of-the-art algorithms on the DB2 images

Algorithm	μ (%)	σ (%)
ExG&Otsu [25]	76.2	7.7
ExGExR [25]	62.3	18.1
EASA [26]	80.2	7.8
GMM [27]	86.9	6.9
ColourHist [28]	82.1	6.4
ClusterMorph [7]	88.1	4.7
GCTV [14]	82.9	9.7
Proposed	**92.7**	**4.3**

Bold values indicate the best indices

vegetation and background, while our method catches these subtle differences surprisingly well.

We also compared the segmentation qualities of the proposed algorithm with GCTV [14] and those of six other methods reported in [7] on the DB2 images. Among these algorithms, ExG&Otsu and ExGExR are VSI-based methods; EASA, GMM and ColourHist are statistical methods; and ClusterMorph is a hybrid method that uses clustering and mathematical morphology. The performances of these algorithms are compared using the means μ and standard deviations σ of the metric defined by Eq. (13) in Table 2.

According to the experiments, the VSI-based methods (ExG&Otsu and ExGExR) performed the worst, and their means on the J metric were only 76.2 and 62.3%, respectively. EASA, GMM and ColourHist use the probability density function of color instead of the thresholding used by the VSI-based methods, and they yielded better results: $\mu > 80\%$. ClusterMorph obtained a high μ of 88.1 through its use of particle swarm optimization clustering and morphology modeling in the CIE $L^*a^*b^*$ color space, although it also required an extra offline learning stage. Our algorithm performs the best ($\mu = 92.7\%$ and $\sigma = 4.3\%$), which is much better than GCTV.

Figure 6 depicts some segmentation examples from DB2 obtained by ClusterMorph and our algorithm. ClusterMorph was chosen for comparison here because its performance is superior to the other methods based on the results in Table 2. The 3 rows demonstrate 3 examples under different illumination: the black pixels represent rice areas, and the white pixels represent the background. As shown, ClusterMorph produced more FPs than did our algorithm under these different illumination conditions—especially in shaded areas.

In addition to the preceding evaluation, Tables 3 and 4 list the means and standard deviations of the other performance metrics defined by Eqs. (14)–(18) obtained by our algorithm on the DB1 and DB2 images, respectively. Among these metrics, the mean value of *Spec* is approximately 83%, while those of the other metrics are all above 90%.

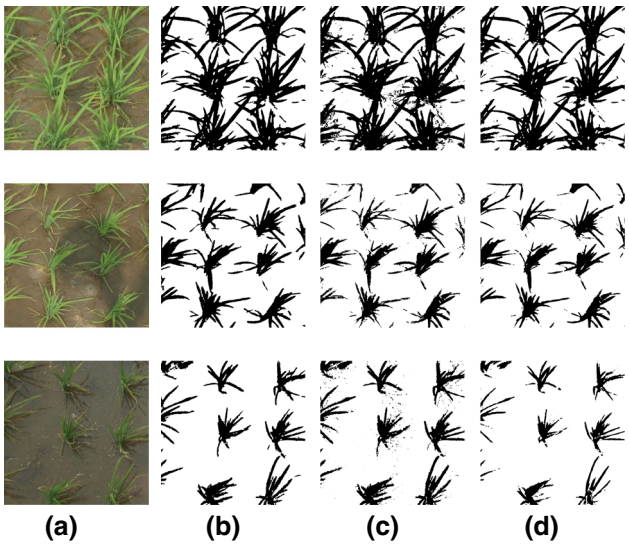

Fig. 6 Examples of segmentation results by our algorithm and ClusterMorph from [7]. The black pixels represent rice, and the white pixels represent the background: **a** The original images from DB2; **b** the ground truth; **c**–**d** the results by ClusterMorph and our algorithm, respectively

Table 3 Other performance metrics of our algorithm on the DB1 images (μ and σ denote the mean values and standard deviations, respectively)

	Sens	Spec	Acc_2	Prec	J	F
μ (%)	95.04	83.17	93.90	98.07	93.31	96.47
σ (%)	4.98	10.55	4.67	1.72	5.27	2.91

Table 4 Other performance metrics of our algorithm on the DB2 images (μ and σ denote the mean values and standard deviations, respectively)

	Sens	Spec	Acc_2	Prec	F	Acc_1
μ (%)	99.08	82.95	94.65	93.49	96.14	91.15
σ (%)	2.52	9.24	3.08	4.04	2.50	3.84

5 Conclusion

The proposed method, which integrates both multi-channel local wavelet texture and color, obtains a higher vegetation segmentation accuracy compared to several other state-of-the-art methods. By combining the prior color of the estimated object and the background with local wavelet texture features, the proposed method uses both the colors of pixels and regional texture features. In addition, it avoids the strong dependence on prior color exhibited by GCTV-based methods and substantially improves the accuracy of segmentation results. Due to the regional texture fidelity term, the segmentation results for areas with similar colors but different textures are more accurate.

Acknowledgements This research was partly supported by the Guangdong Provincial Key Laboratory of Digital Signal and Image Processing Techniques (2013GDDSIPL-03) and by the Guangxi Basic Ability Promotion Project for Young and Middle-aged Teachers (2017KY0247).

References

1. Von Bargen, K., Woebbecke, D.M., Meyer, G.E., Mortensen, D.A.: Color indices for weed identification under various soil, residue, and lighting conditions. Trans. ASAE **38**(1), 259–269 (1995)
2. Meyer, G.E., Hindman, T.W., Laksmi, K.: Machine vision detection parameters for plant species identification. In: Photonics East (1999)
3. Kataoka, T., Kaneko, T., Okamoto, H., Hata, S.: Crop growth estimation system using machine vision. In: 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2003. AIM 2003. Proceedings, pp. b1079–b1083
4. Meyer, G.E., Neto, J.C.: Verification of color vegetation indices for automated crop imaging applications. Comput. Electron. Agric. **63**(2), 282–293 (2008)
5. Hague, T., Tillett, N., Wheeler, H.: Automated crop and weed monitoring in widely spaced cereals. Precis. Agric. **7**(1), 21–32 (2006)
6. Ponti, M.P.: Segmentation of low-cost remote sensing images combining vegetation indices and mean shift. IEEE Geosci. Remote Sens. Lett. **10**(1), 67–70 (2013)
7. Bai, X., Zhiguo Cao, Y., Wang, Z.Y., Zhu, H., Zhang, X., Li, C.: Vegetation segmentation robust to illumination variations based on clustering and morphology modelling. Biosyst. Eng. **125**(3), 80–97 (2014)
8. Feng, Q., Liu, J., Gong, J.: Uav remote sensing for urban vegetation mapping using random forest and texture analysis. Remote Sens. **7**(1), 1074–1094 (2015)
9. Aksoy, S., Akcay, H.G., Wassenaar, T.: Automatic mapping of linear woody vegetation features in agricultural landscapes using very high resolution imagery. IEEE Trans. Geosci. Remote Sens. **48**(1), 511–522 (2010)
10. Nia, M.S., Wang, Z., Gader, P., Bohlman, S.A., Graves, S.J., Petrovic, M.: Impact of atmospheric correction and image filtering on hyperspectral classification of tree species using support vector machine. J. Appl. Remote Sens. **9**(1), 095990 (2015)
11. Osher, S., Sethian, J.: Fronts propagating with curvature dependent speed: algorithms based on Hamilton–Jacobi formulations. J. Comput. Phys. **79**(1), 12–49 (1988)
12. Liu, K., Wan, J., Han, Z.: Abnormal event detection and localization using level set based on hybrid features. Signal Image Video Process. **11**(6), 1–7 (2017)
13. Shahvaran, Z., Kamran, K., Helfroush, M.S.: Simultaneous vector-valued image segmentation and intensity nonuniformity correction using variational level set combined with Markov random field modeling. Signal Image Video Process. **10**(5), 887–893 (2016)
14. Unger, M., Pock, T., Bischof, H.: Global relabeling for continuous optimization in binary image segmentation. In: Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science , vol. 6819, pp. 104–117 (2011)
15. Nikolova, M., Chan, T.F., Esedoglu, S.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J. Appl. Math. **66**(5), 1632–1648 (2006)
16. Yang, T., Huang, L., Hu, P.: Vegetation segmentation based on level set using local wavelet texture. In: 2016 IEEE 13th International Conference on Signal Processing, pp. 222–227
17. Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.-P., Osher, S.: Fast global minimization of the active contour/snake model. J. Math. Imaging Vis. **28**(2), 151–167 (2007)
18. Rother, C., Kolmogorov, V., Blake, A.: Grabcut: interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**(3), 309–314 (2004)
19. Yang, T., Huang, L., Jiang, C.: Wavelet and total variation based level set segmentation method for portraits. Int. J. Digit. Content Technol. Appl. **7**(4), 973–982 (2013)
20. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)
21. Xie, Z., Wang, S., Dewen, H.: New insight at level set & Gaussian mixture model for natural image segmentation. Signal Image Video Process. **7**(3), 521–536 (2013)
22. Chang, J., Fisher, J.W.: Efficient MCMC sampling with implicit shape representations. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2081–2088 (2011)
23. Cheng, Y.: Mean shift, mode seeking clustering. IEEE Trans. Pattern Anal. Mach. Intell. **17**(8), 790–799 (1995)
24. Gée, C., Bossu, J., Jones, G., Truchetet, F.: Crop/weed discrimination in perspective agronomic images. Comput. Electron. Agric. **60**(1), 49–59 (2008)
25. Neto, J.C.: A combined statistical-soft computing approach for classification and mappingweed species in minimum-tillage systems. Thesis (2006)
26. Tian, L.F., Slaughter, D.C.: Environmentally adaptive segmentation algorithm for outdoor image segmentation. Comput. Electron. Agric. **21**(3), 153–168 (1998)
27. Bergasa, L.M., Mazo, M., Gardel, A., Sotelo, M.A., Boquete, L.: Unsupervised and adaptive Gaussian skin-color model. Image Vis. Comput. **18**(12), 987–1003 (2000)
28. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. Int. J. Comput. Vis. **46**(1), 81–96 (2002)