

## Research Article

# Adaptive Bacteria Colony Picking in Unstructured Environments Using Intensity Histogram and Unascertained LS-SVM Classifier

Kun Zhang,<sup>1</sup> Minrui Fei,<sup>1</sup> Xin Li,<sup>1</sup> and Huiyu Zhou<sup>2</sup>

<sup>1</sup> School of Mechatronic Engineering & Automation, Shanghai University, M8 Building, 149 Yanchang Road, ZhaBei District, Shanghai 200072, China

<sup>2</sup> The Institute of Electronics, Communications and Information Technology, Queen's University Belfast, Belfast, UK

Correspondence should be addressed to Minrui Fei; [mrfei@staff.shu.edu.cn](mailto:mrfei@staff.shu.edu.cn)

Received 13 March 2014; Accepted 10 April 2014; Published 12 May 2014

Academic Editor: Jinshan Tang

Copyright © 2014 Kun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Features analysis is an important task which can significantly affect the performance of automatic bacteria colony picking. Unstructured environments also affect the automatic colony screening. This paper presents a novel approach for adaptive colony segmentation in unstructured environments by treating the detected peaks of intensity histograms as a morphological feature of images. In order to avoid disturbing peaks, an entropy based mean shift filter is introduced to smooth images as a preprocessing step. The relevance and importance of these features can be determined in an improved support vector machine classifier using unascertained least square estimation. Experimental results show that the proposed unascertained least square support vector machine (ULSSVM) has better recognition accuracy than the other state-of-the-art techniques, and its training process takes less time than most of the traditional approaches presented in this paper.

## 1. Introduction

Bacteria colony isolation [1] is a labor intensive task over the past decades. Manual bacteria colony picking is tedious and experience dependent. Colony screening is in unstructured environments due to different agar mediums and cultivations. Figure 1 is an example of Erythrosin bacteria colony sitting on agar. An automatic colony picking system can be used to make this process consistent and reliable with less time consumption. Researchers worldwide are currently seeking fast and reliable methods for high throughput colony picking. To achieve this, we need to make sure high quality colony illumination and image segmentation are the critical stages of a colony picking system. Currently, there are three major illumination techniques used for image acquisition: (1) drop-in bright-field illumination, (2) back-projective bright-field illumination, and (3) suspended dark-field illumination. Figure 2 shows the different imaging quality based on the three techniques introduced above. Suspended dark-field illumination based approaches can be used to reduce the influence of lights. In this circumstance, colony agar plates are placed in a suspended dark-field environment. Using

reflected and refractive lights, we can obtain volumetric structures of colonies with good image quality. However, the suspended dark-field illumination based approaches could achieve less satisfactory quality of images than the other two approaches, due to a similar color caused by the crowdedness of colonies. Image segmentation approaches such as thresholding [2], region growing [3], watershed [4], and mean shift [5] are commonly used in medical image analysis. Each of these classical methods has its own strengths and weaknesses. For example, a thresholding method is fast but requires the systematic parameters to be changed for different environments. Region growing methods are more robust than the thresholding methods but lack sufficient efficiency. The colony picking systems available on the market have a number of specific requirements in order to achieve good segmentation performance, for example, setting the region of interests, controlling the extent of cluttering, and maintaining appropriate light conditions.

In this paper, we deploy an intensity histogram based morphological features extraction algorithm, which contributes to colony analysis. The proposed method employs a peak-searching method in a standard intensity histogram.



FIGURE 1: Erythrosin bacteria colony on agar.

Afterwards, to achieve correct colony feature classification, we propose an entropy based mean shift algorithm to smooth the image as a preprocess stage. Finally, we introduce an improved approach for feature selection using an unascertained least square support vector machine (ULSSVM) classifier. To our knowledge, this is the first attempt to use the unascertained attributes of the detected features for the purpose of classification. We evaluate the proposed approach on a large dataset of colony images. Based on these experiments, we show that our approach can efficiently deliver correct classification results.

## 2. Proposed Approach

In the smoothing process of image, it is very important for a nonlinear filter to keep the fringe detail of an image during the process. Optimization techniques have been popularly used in image processing. This is driven by the performance need as the target of an application. However, it is very difficult to obtain an optimal solution (stopping criterion) for individual applications. More details can be found in [6–9]. Mean shift has proven to be appliance effective tool for image processing because of its nonparametric property. Smoothing by mean shift algorithm has been reported in the literature. For example, in [10], Zhao and Xi introduced mean shift as a smooth filter for processing YIQ color images and compared it with Wiener filter. In [11], Han and Sohn used mean shift combined with a sigma filter in an illumination and color compensation system. In literature [12], Sahba and Venetsanopoulos applied mean shift to reserve fringe detail and detect breast mass. These results are promising but the computational speed of mean shift is unexpectedly slow.

Entropy is a measure of complexity. We can also use entropy to inspect system uncertainty. Low entropy images have very little contrast and large runs of pixels with the same value. An image that is perfectly flat will have entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images have a great deal of contrast from one pixel to the next and consequently have more details than low entropy images. Entropy has been applied in pattern recognition, object tracking, and image segmentation, for example, [13–15], where entropy has been used as a termination criterion. As mentioned in the first section, the proposed feature based colony classification approach begins with entropy based mean shift filter, and

it is followed by applying intensity histogram analysis to the filtered images. The characteristic peaks' coefficients retrieved from intensity histograms are then applied to colony classification within the framework of unascertained least square support vector machine.

**2.1. Entropy Based Mean Shift Filter.** The following is the idea of a standard mean shift approach [16]. Let  $x_j$  be a numerical sample of  $n$  in a  $d$ -dimensional space. The basic mean shift is defined as

$$M_h(x) = \frac{1}{k} \sum_{x_i \in g_h} (x_i - x), \quad (1)$$

where  $g_h$  is a window based on the center  $x$  and radius  $h$ .  $k$  is the sample set number in  $g_h$ .  $(x_i - x)$  is the relative offset of center  $x$ .

Equation (1) is a monotonic form and less effective in a practical application. A Kernel based mean shift algorithm is described as follows:

$$M_h(x) = \frac{\sum_{i=1}^n Q((x_i - x)/h) \alpha(x_i - x)}{\sum_{i=1}^n Q((x_i - x)/h) \alpha(x_i)}, \quad (2)$$

where  $\alpha(x)$  is the self-impact factor and  $Q(x)$  is a kernel function.

In a color image with  $n \times n$  pixels, each pixel corresponds to a 5-dimension vector  $R^5 (R, G, B, X, Y)$ . Due to the independence of space and color information, the kernel function is obtained in

$$Q_{g_s, g_r}(x) = \frac{1}{g_s^2 g_r^3} q\left(\left\|\frac{x^s}{g_s}\right\|^2\right) q\left(\left\|\frac{x^r}{g_r}\right\|^2\right), \quad (3)$$

where  $x^s$  is the spatial position of a pixel;  $x^r$  is the color information of a pixel;  $g_s$  is a spatial window based on the center  $x$  and radius  $s$ ;  $g_r$  is a color window based on the center  $x$  and radius  $r$ .

Let  $q(x_t)$  be the gray value probability of the outcome  $x_t$ ,  $k = 1, \dots, n$ .  $A$  is an image with  $\log_2(0) = 0$ ;  $\log_2(1/q(x_t))$  is called the surprisal of the outcome  $x_t$ . Entropy is defined as

$$F(X) \equiv \sum_{t=1}^n q(x_t) \log_2\left(\frac{1}{q(x_t)}\right) = -\sum_{t=1}^n q(x_t) \log_2 q(x_t). \quad (4)$$

Entropy is determined based on the pixels distribution in an image, which is influenced by two factors: foreground and background or called noise. The uncertainty of entropy is dominated by the noise's variance. Entropy can be used to measure the homogeneity of an image area: the more homogeneous image, the less the entropy values. In practice, when we work with images, due to the noise, entropy cannot decrease to zero but it can reach a stable value. Thus, entropy can be applied as a stopping criterion for a mean shift iteration (Algorithm 1).

**2.2. Model of Peak Searching in Intensity Histogram.** Intensity histogram is an important feature of images and can be

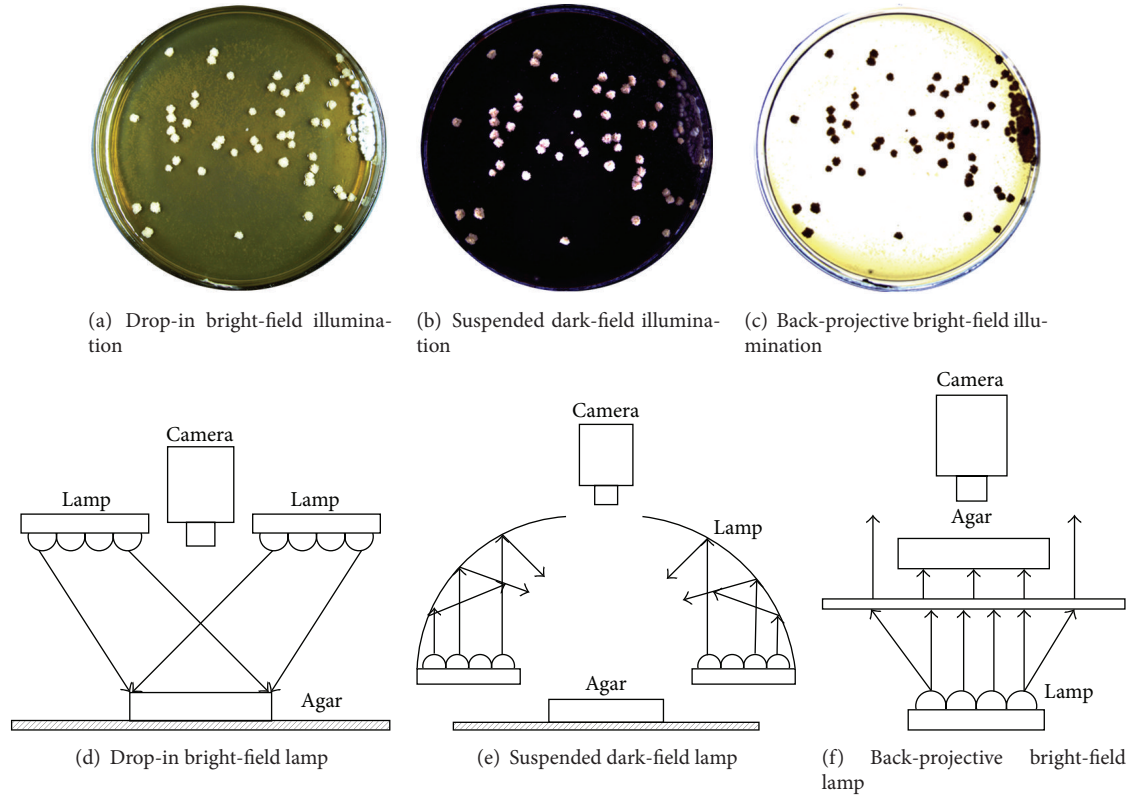


FIGURE 2: Three illumination technologies.

Let  $x_i, i = 1, \dots, n$  be the input image. Let  $O_i, i = 1, \dots, n$  be the filtered image. Pixel  $p \in x_i, p = (R, G, B, x, y) \in R^5$ . Let  $ent.0$  be the entropy initial value,  $ent.1$  be the next iteration of  $ent.0$ , and  $ent.2$  be the next value of  $ent.1$ . Let  $erras$  be the absolute value of the difference between the first two iterations. Let  $edset$  be the thresholding as iterations stopping criteria. Our algorithm comprises the steps listed below:

- (1) Initialize  $h = 1, y_{h,1} = p_h, ent.2 = 1, erras = 1, edset = 0$ .
- (2) While  $erras > edset$ , then
  - (2.1) Filtering image based on mean shift. Store the image in  $O^{[k]}$ .
  - (2.2) Calculating entropy from the  $O^{[k]}$  and store the image in  $ent.1$ .
  - (2.3) Entropy is used to calculate the absolute difference which is obtained in the previous step;  $erras = |ent.1 - ent.2|$ .
  - (2.4) Update the parameters;  $ent.2 = ent.1$  and  $O^{[k+1]} = O^{[k]}$ .
  - (2.5) Calculate the mean shift which is carried out until entropy convergence.
- (3) Store  $Z_i$  which is calculated  $Z_i = (x_i^s, y_{i,c}^r)$ , here  $x_i^s$  is the spatial information and  $y_{i,c}^r$  is the color range information.

ALGORITHM 1: Entropy based mean shift filtering algorithm.

regarded as the approximate expression of a density function of image intensities. It shows the frequency of an intensity appearing in an image. An intensity histogram is described in

$$G_i = \sum_{m=1}^M \sum_{n=1}^N P(i, m, n), \quad (5)$$

where  $M$  and  $N$  represent the total numbers of rows and columns, respectively, and  $G_i$  represents the appearance times of intensity  $i$ , and  $P(i, m, n)$  is described as follows:

$$P(i, m, n) = \begin{cases} 1 & P(m, n) = i \\ 0 & P(m, n) \neq i, \end{cases} \quad (6)$$

where  $P(m, n)$  is the intensity value of point  $(m, n)$ . As the sizes of different images may be different, to avoid the impact of image size, we normalize each image according to the following equation:

$$G'_i = \frac{G_i}{T}, \quad (7)$$

where  $G'_i$  is a normalization value and  $T$  represents the total amount of the pixels of an image. Figure 1 shows an example of bacteria colony and its intensity histogram using (5), (6), and (7).

In Figure 1, the entire image can be divided into two main zones: culture medium zone and colony zone, according

- (1) Obtain  $\Delta G'_i$  from the image transformed based on Meanshift algorithm, and using a second order differentiation described in (8).
- (2) Find the highest peak from the entire curve.
- (3) Adaptively select a defined threshold: The threshold used to obtain the global peak is determined by both the Maximum and the minimum peaks (threshold = (max (peak) – min (peak))/8).  
If two peaks are detected, continue. If two peaks are found, jump to Step 6.
- Repeat
- (4) Increase the threshold, from ((max (peak) – min (peak))/8) to ((max (peak) – min (peak))/4).
- Until
- (5) Stop if finding two characteristic peaks.
- (6) End

ALGORITHM 2: The proposed adaptive quickly peaks detection.

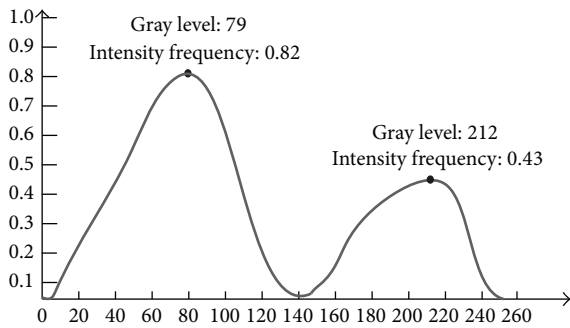


FIGURE 3: Normalised intensity histogram.

to the contrast and density of the image. In Figure 3, there are two peaks, which represent high frequency values of the corresponding intensities. The difference between the gray levels of the two peaks is 133, and the difference between the two intensity frequencies of the two peaks is 0.39. The intensity histogram mathematical model will be introduced in the following sections.

The peaks and valleys shown in Figure 3 can be obtained using the second-order derivative. The method to find a peak or a valley is described in

$$\Delta G'_i = \begin{cases} 1 & G'_i - G'_{i+1} < -H \\ 0 & |G'_i - G'_{i+1}| < H \\ -1 & G'_i - G'_{i+1} > H, \end{cases} \quad (8)$$

where  $H$  refers to a positive threshold which is set according to a specific image to reduce inaccuracy because of infinitesimal disturbance.  $\Delta G'_i$  represents the tendency of the histogram curve at the point where the intensity equals  $i$ .  $\Delta G'_i = 1$  means that the curve ascends,  $\Delta G'_i = -1$  means that the curve descends, and  $\Delta G'_i = 0$  means that the curve is flat. Therefore,  $1(0 \cdots 0) - 1$  indicates the peak of the curve, and  $-1(0 \cdots 0)1$  refers to the valley of the curve. In Figure 3, the peaks have been extracted according to the method described above.

Through the experiments, it is found that most of the intensity histogram curves change from double-peak to multipeak due to different bacteria colony in different illumination conditions. It is possible to obtain multiple local peaks

or valleys in case the boundary of the intensity histogram is not smooth. The proposed peaks searching algorithm is shown in Algorithm 2.

### 3. Unascertained Least Square Support Vector Machine

Support vector machines (SVM) have been well studied in the machine learning field, which was proposed by Vapnik [17]. The performance of SVM has been verified in many applications, such as handwriting recognition [18], face recognition [19], and medical pattern matching [20]. But the training speed of SVM is too slow and this hinders its applications. Different from the classical support vector machine methods, the least squares support vector machines (LSSVM) proposed by Suykens and Vandewalle [21] were to change the form of the original convex quadratic optimization problem into a linear optimization problem and they effectively enhance the training speed. But it is hard to classify some uncertain information. Based on LSSVM and unascertained mathematical models, we propose the ULSSVM algorithm. For unascertained information, we can use unascertained number [22] and unascertained programming [23] to describe our algorithm. Please see below for a summary of these theories.

**Theorem 1.**  $[\alpha, \beta], \alpha = x_1 < x_2 < \cdots, x_n = \beta$ , if function  $f(x)$  satisfies

$$f(x) = \begin{cases} \phi_i & x = x_i, i = 1, \dots, l \\ 0 & x \neq x_i, x \in [\alpha, \beta], \end{cases} \quad (9)$$

$$s.t. \sum_{i=1}^l \phi_i = \phi, \quad 0 < \phi_i \leq 1 (i = 1, \dots, l),$$

where  $[\alpha, \beta]$  and  $f(x)$  form  $l$  order unascertained numbers and are set as  $[[\alpha, \beta], f(x)]$ .  $\phi$  is the main reliability.  $[\alpha, \beta]$  is the value range.  $f(x)$  is the main reliability distribution density function.  $\{x_i\}$  is a possible value sequence of the unascertained numbers.  $\{\phi_i\}$  is a confidence value sequence of the unascertained numbers.

**Theorem 2.** Setting the unascertained number  $a = [[\alpha_1, \alpha_n], f(y)], b = [[\beta_1, \beta_m], k(x)],$

$$f(y) = \begin{cases} \phi_i, & y = y_i, \quad i = 1, \dots, l \\ 0, & y \neq y_i, \quad y \in [y_1, y_l], \end{cases} \quad (10)$$

$$\sum_{i=1}^l \phi_i = \phi, \quad 0 < \phi_i \leq 1 \quad (i = 1, \dots, l), \quad (11)$$

$$k(x) = \begin{cases} \gamma_j, & x = x_j, \quad j = 1, \dots, p \\ 0, & x \neq x_j, \quad x \in [x_1, x_p], \end{cases} \quad (12)$$

$$\sum_{j=1}^p \gamma_j = \gamma, \quad 0 < \gamma_j \leq 1 \quad (j = 1, \dots, p). \quad (13)$$

One calls inequalities  $a \leq b, b \leq a$  unascertained events.

**Theorem 3.** One calls the following programing as an unascertained constraint programing:

$$\begin{aligned} \max \quad & \bar{f} \\ \text{s.t.} \quad & Cr \{f(X, a) \geq \bar{f}\} \geq \gamma \\ & Cr \{g_j(X, b_j) \leq 0, j = 1, \dots, p\} \geq \phi, \end{aligned} \quad (14)$$

where  $X$  is a decision vector and  $a, b_j$  ( $j = 1, \dots, p$ ) are unascertained parameter vectors.  $f(X, a)$  is the target function.  $g_j(X, b_j)$  is a constraint function.  $\phi, \gamma$  ( $\phi, \gamma \in (0, 1]$ ) are confidence levels of the constraint and target function.  $Cr\{\cdot\}$  is a credible degree of the unascertained events.

Based on the preliminary knowledge mentioned above, if the SVM training data obtains unascertained information, we can transform the unascertained information into unascertained number

$$a = [[x_1, x_n], f(x)], \quad (15)$$

$$f(x) = \begin{cases} \phi_j & x = x_j, \quad j = 1, \dots, l \\ 0 & x \neq x_j, \quad x \in [x_1, x_l]. \end{cases} \quad (15)$$

The training set is defined in

$$K = \{(y_1, a_1), (y_2, a_2), \dots, (y_l, a_l)\}, \quad (16)$$

where  $y_i \in R^l, a_i$  is an unascertained number,  $(y_i, a_i)$  ( $i = 1, \dots, l$ ) is an unascertained training point, and  $K$  is an unascertained training set.

The objective function can be minimized as follows:

$$\min_{\omega, b, \xi} J(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2, \quad (17)$$

$$\text{s.t.} \quad A_i [w^T \phi(x_i) + b] = 1 - \xi_i, \quad i = 1, \dots, n.$$

We then define a Lagrange function as

$$L(\omega, b, \xi, \alpha) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i \{A_i [w^T \phi(x_i) + b] - 1 + \xi_i\}. \quad (18)$$

According to the KKT condition,

$$\begin{aligned} \frac{\partial L}{\partial \omega} = 0 & \implies \omega = \sum_{i=1}^n \alpha_i A_i \phi(x_i), \\ \frac{\partial L}{\partial b} = 0 & \implies \sum_{i=1}^n \alpha_i A_i = 0, \end{aligned} \quad (19)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \implies \alpha_i = \gamma \xi_i,$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \implies A_i [w^T \phi(x_i) + b] - 1 + \xi_i = 0.$$

Equation (19) can turn into the following matrix problem:

$$\begin{bmatrix} L & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma^I & -L \\ Z & Y & L & 0 \end{bmatrix} \begin{bmatrix} \bar{\omega} \\ b \\ \xi \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ L \end{bmatrix}, \quad (20)$$

where  $Z = [\varphi(x_1)^T y_1, \varphi(x_2)^T y_2, \dots, \varphi(x_n)^T y_n]^T, Y = [y_1, y_2, \dots, y_n], L = [1, 1, \dots, 1]^T, \xi = [\xi_1, \xi_2, \dots, \xi_n],$  and  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n].$

We eliminate  $w$  and  $\xi$  and then get the following equations:

$$\begin{bmatrix} 0 & -Y^T \\ Y & \Omega + \frac{1}{\gamma} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ L \end{bmatrix}, \quad (21)$$

where  $\Omega = ZZ^T,$

$$\Omega_{kl} = y_k y_l \phi(x_k)^T \phi(x_l) = y_k y_l \psi(x_k, x_l), \quad (22)$$

where  $\Omega(\cdot)$  is a kernel function and satisfies the Mercer theorem.

A set of linear equations will be solved instead of a QP problem. Finally, we can obtain the following optimal classification function:

$$f(x) = \text{sign} \left( \sum_{i=1}^l A_i \alpha_i^* (x \cdot x_i) + b^* \right), \quad (23)$$

where  $\alpha_i^*$  is the optimal solutions and corresponding bias  $b^*.$   $A_i$  is an unascertained set.

#### 4. Experimental Results

The proposed algorithm is evaluated on colony image databases which are captured using a Basler CCD sensor. The

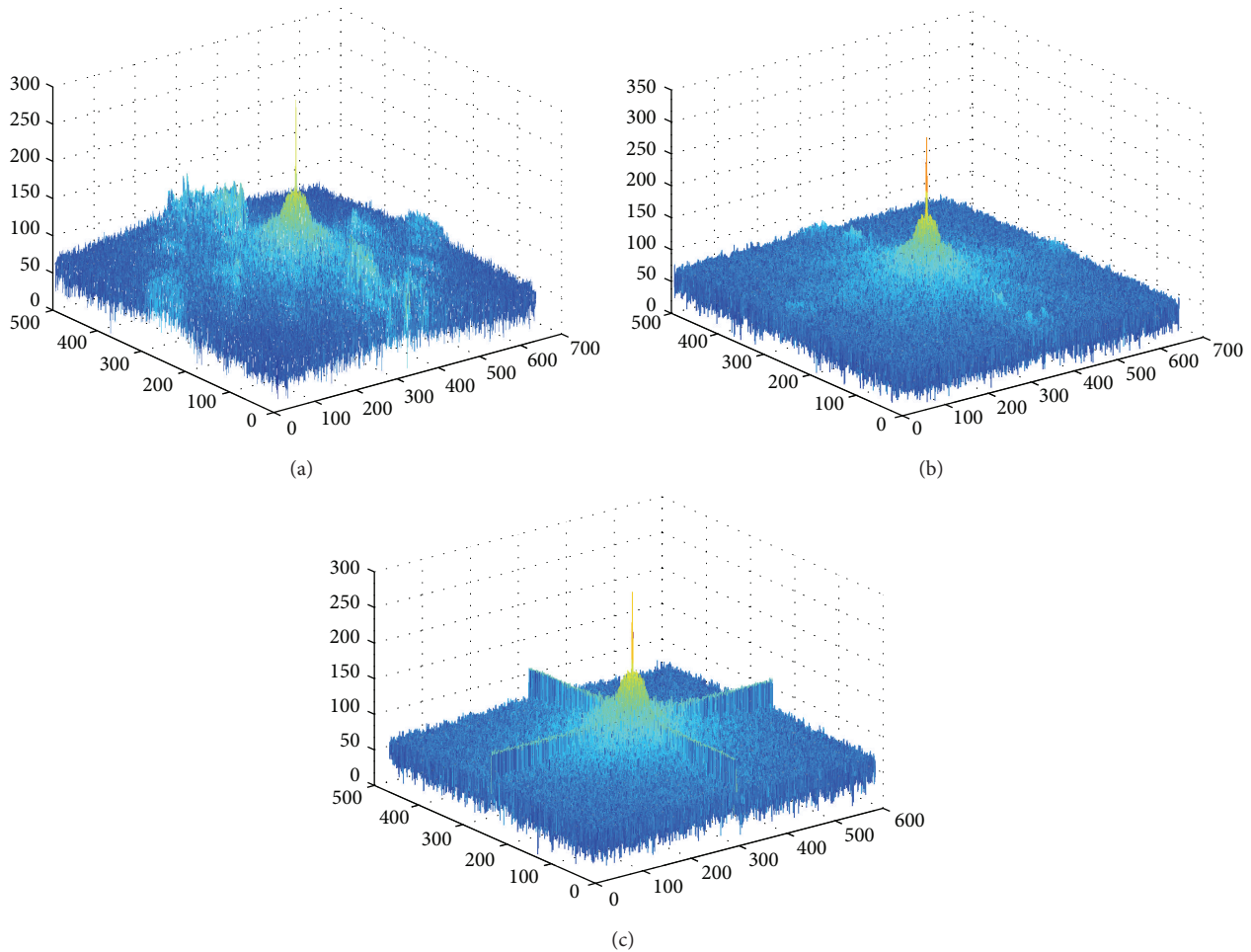


FIGURE 4: Energy spectrum at different stages.

TABLE 1: The filtering performance of three filtering algorithms.

Performance comparison	MSE		
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$
Median filter	38.541	47.259	85.3254
Fuzzy filter	36.573	49.694	82.3651
Mean shift filter	33.258	40.258	79.6938

images are resized to be  $640 \times 480$ . The used computer is of a 3.2 G CPU running Windows 7 with a 4 G memory. The first three experiments have been carried out using the colony images to analyze the feasibility and efficiency of the proposed algorithm with MATLAB 7.2. The last experiment is carried out to demonstrate the segmentation effect and performed with Visual Studio 2010.

In image denoising, classical low-passing filters can suppress high frequency noise [24]. However, it is hard for them to preserve the edges of images due to the mixture in some frequency bands. Here, we use energy density spectrums to illustrate the outcomes of different filters [25, 26]. In Figure 4(a), this is the energy spectrum of the original image,

where yellow-orange indicates the major energy of symbol “+” and this area is contaminated by the background noise (i.e., blue and green areas). Figure 4(b) shows the outcome of a low-pass filter, where only the central area of the symbol is kept but the edges of symbol “+” are mixed with the background. In Figure 4(c), based on mean shift, it is clear that the central area of the symbol is outstanding and the edges are also kept. Traditional low-pass filters have good performance on image smoothing but also affect the edge details. In Figure 5 we observe that, after 5 iterations, entropy can reach a stable value, and meanwhile the mean shift iteration automatically stops. Listeria colony entropy is different from the other two. Second row in Figure 6 is Listeria colony. This is because of the cluster colony and agar color is approximated with colony. Figure 6 shows different kinds of colony. The first column shows three different cultures: *Microsporium audouinii*, *Listeria monocytogenes*, and *Cephalosporin*. During the development of cultures, their biochemical reactions appear to be significantly different. As a result, the histograms of cultures in drop-in bright-filed illumination may accompany a number of noisy peaks, illustrated on the second column. Using mean shift based filtering algorithms, we can remove the irregular backgrounds

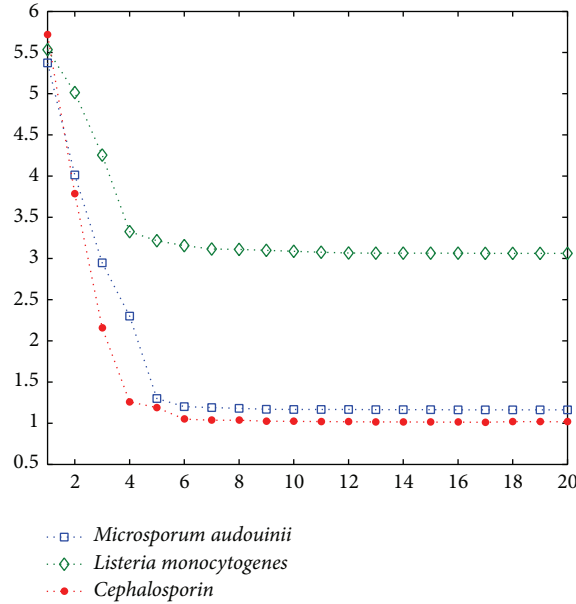


FIGURE 5: Entropy calculation using Mic, Listeria, and Ceph.

TABLE 2: The value corresponding to the unascertained information.

$f_1(y) = \begin{cases} 0.4 & y = 1 \\ 0.3 & y = 2 \\ 0.2 & y = 3 \\ 0.1 & y = 4 \end{cases}$	$f_2(y) = \begin{cases} 0.3 & y = 1 \\ 0.5 & y = 2 \\ 0.2 & y = 3 \\ 0 & y = 4 \end{cases}$	$f_3(y) = \begin{cases} 0 & y = 1 \\ 0.5 & y = 2 \\ 0.1 & y = 3 \\ 0.4 & y = 4 \end{cases}$	$f_4(y) = \begin{cases} 0.7 & y = 1 \\ 0.3 & y = 2 \\ 0 & y = 3 \\ 0 & y = 4 \end{cases}$	$f_5(y) = \begin{cases} 0.4 & y = 1 \\ 0 & y = 2 \\ 0.1 & y = 3 \\ 0.5 & y = 4 \end{cases}$
$f_6(y) = \begin{cases} 0.5 & y = 1 \\ 0.1 & y = 2 \\ 0.4 & y = 3 \\ 0 & y = 4 \end{cases}$	$f_7(y) = \begin{cases} 0.1 & y = 1 \\ 0.4 & y = 2 \\ 0.3 & y = 3 \\ 0.2 & y = 4 \end{cases}$	$f_8(y) = \begin{cases} 0.7 & y = 1 \\ 0.2 & y = 2 \\ 0 & y = 3 \\ 0.1 & y = 4 \end{cases}$	$f_9(y) = \begin{cases} 0.5 & y = 1 \\ 0.1 & y = 2 \\ 0.3 & y = 3 \\ 0 & y = 4 \end{cases}$	$f_{10}(y) = \begin{cases} 0.2 & y = 1 \\ 0.3 & y = 2 \\ 0.1 & y = 3 \\ 0.4 & y = 4 \end{cases}$

and hence reduce noisy peaks. The result of using mean shift is shown on the 3rd column. Furthermore, we apply an adaptive thresholding based peak searching approach in order to detect two peaks, which indicate the features of cultures. This results in the 4th column.

The performance of mean shift filtering can be measured with mean square error (MSE):

$$MSE = \frac{1}{BG} \sum_{x=0}^{B-1} \sum_{y=0}^{G-1} [(I'(x, y) - I(x, y))^2], \quad (24)$$

where  $I'(x, y)$  is denoted by the filtered output image and  $I(x, y)$  is denoted by the original input image; Table 1 shows the performance of the three kinds of filtering approach.

After extracting the characteristic peaks, we apply the ULSSVM classifier to the data for classification. We now evaluate the performance of our ULSSVM classifier against the classical SVM [27], LSSVM [28], and fuzzy SVM [29] using 400 colony images, where 150 samples belong to class 1, 50 samples belong to class 2, 100 belong to class 3, and the remainder belong to class 4. We randomly select 300

TABLE 3: Classifiers' experimental results.

Classifier	SVM	FSVM	LSSVM	ULSSVM
Accuracy	75.9%	84.6%	73.2%	<b>92.7%</b>
Training speed/s	12.1 s	13.7 s	5.1 s	<b>2.7 s</b>

samples as the training set and the remaining 100 samples are considered as the testing set. There are 130 samples labeled as the unascertained numbers, and half of them are set as training samples. Table 2 shows the values corresponding to the unascertained information. Table 3 shows the classification results. In Table 3, the experimental results demonstrate that the ULSSVM effectively improves the performance of classification.

We carry on screening colony using the ULSSVM classifier. Figure 7 shows the interface of colony screening. Figure 8 is adaptive colony segmentation. The segmentation outcomes of two different colony picking methods are as follows: the first row shows the colony with homogeneous medium and its segmentation results using region growing, and the next two rows show the colony with inhomogeneous medium and

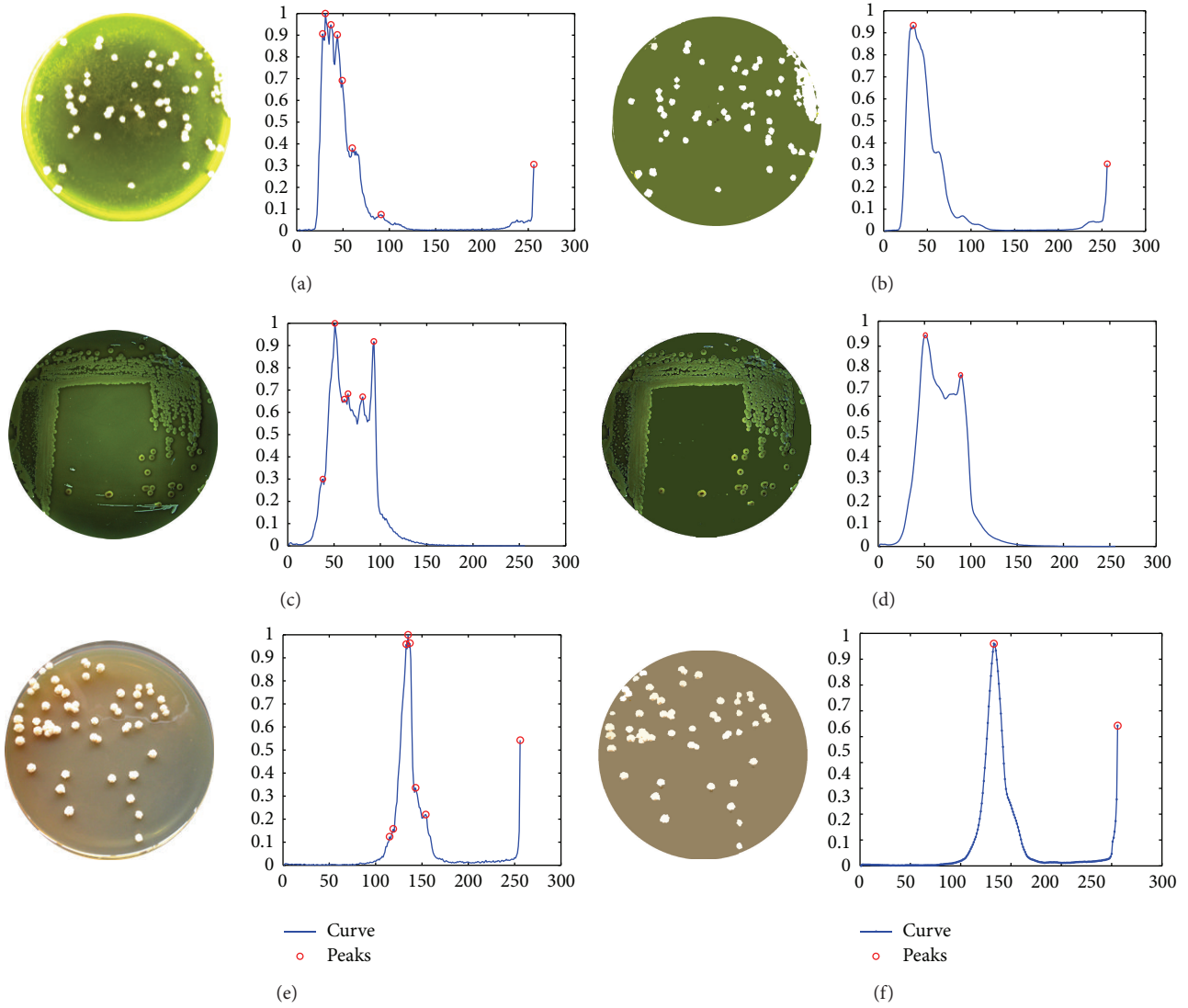


FIGURE 6: Colony imaging and morphological feature extract.

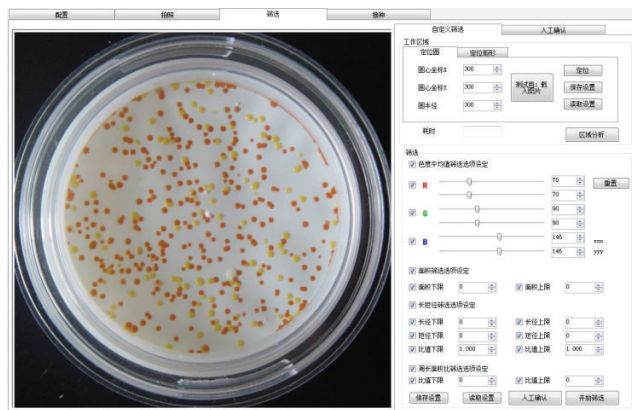


FIGURE 7: Interface of colony screening.



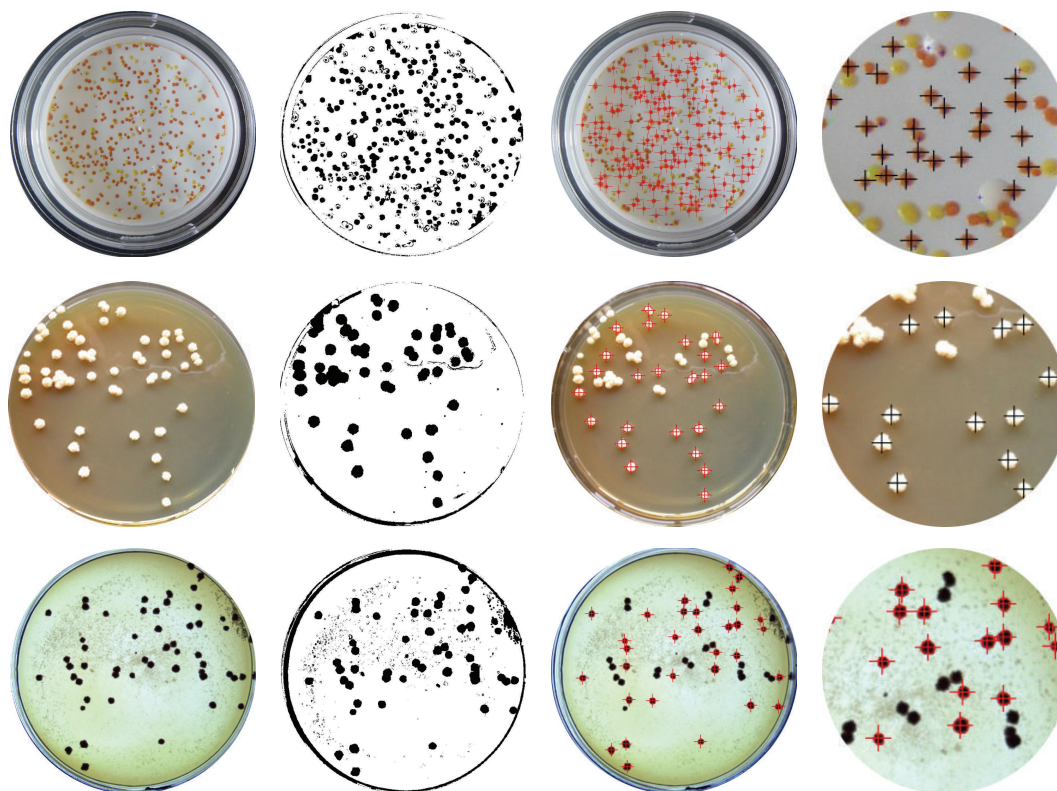


FIGURE 8: Adaptive colony segmentation based on morphological features.

the corresponding segmentation results using thresholding. The first column is original images, the second column is image process based on our approach, the third column is the identification results, and the fourth column is the local zooming of the screening. Meanwhile, we calculate the time consumption of the colony screen. The process of using the thresholding method took 2.57 s and the process of using the region growing method took 8.61 seconds.

## 5. Conclusions

In this paper, we have deployed an approach to perform adaptive colony segmentation in unstructured environments using feature extraction and selection in an intelligence classifier. We used the intensity histogram peaks as features. To properly determine the importance of the extracted features for colony classification, we used an unascertained theory based LSSVM classification algorithm. Experimental results show that this new approach had better performance than other state-of-the-art techniques in terms of accuracy and speed. This approach works well for adaptive colony segmentation, whilst optimizing the time consumption of colony picking.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Key Scientific Instrument and Equipment Development Projects (2012YQ15008703 and 2012YQ15008702).

## References

- [1] T. Hartley, C. Stewart, R. Stewart, and D. J. Munroe, "Cost-effective addition of high-throughput colony picking capability to a standard liquid-handling platform," *Journal of the Association for Laboratory Automation*, vol. 14, no. 1, pp. 22–26, 2009.
- [2] J. Romeo, G. Pajares, M. Montalvo, J. M. Guerrero, M. Guijarro, and A. Ribeiro, "Crop row detection in maize fields inspired on the human visual perception," *The Scientific World Journal*, vol. 2012, Article ID 484390, 10 pages, 2012.
- [3] R. K. Justice, E. M. Stokely, J. S. Strobel, R. E. Ideker, and W. M. Smith, "Medical image segmentation using 3-D seeded region growing," in *Medical Imaging 1997: Image Processing*, vol. 3034 of *Proceedings of SPIE*, pp. 900–910, February 1997.
- [4] J. A. Solves Llorens, M. J. Rupérez, C. Monserrat, E. Feliu, M. García, and M. Lloret, "Segmentation of the breast skin and its influence in the simulation of the breast compression during an x-ray mammography," *The Scientific World Journal*, vol. 2012, Article ID 876489, 8 pages, 2012.
- [5] H. Y. Zhou, G. Schaefer, A. H. Sadka, and M. E. Celebi, "Anisotropic mean shift based fuzzy C-means segmentation of dermoscopy images," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 26–34, 2009.

- [6] J. G. Kuk and N. I. Cho, "Weighted gradient domain image processing problems and their iterative solutions," *Eurasip Journal on Image and Video Processing*, vol. 2013, 2013.
- [7] H. Hrizi and N. Sboui, "Reducing the numerical calculation in the wave iterative method by image processing techniques," *Applied Computational Electromagnetics Society Journal*, vol. 27, no. 6, pp. 524–531, 2012.
- [8] Y. Awatsuji and T. Kubota, "Two-dimensional H-tree parallel optical interconnect for two-dimensional image by using optical iterative processing," *IEEE Photonics Technology Letters*, vol. 13, no. 1, pp. 79–81, 2001.
- [9] J. S. Jrgensen, E. Y. Sidky, and X. C. Pan, "Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-Ray CT," *IEEE Transactions on Medical Imaging*, vol. 32, no. 2, pp. 460–473, 2013.
- [10] L. F. Zhao and Z. H. Xi, "Halo inhibition algorithm based on mean shift filter," *Systems Engineering and Electronics*, vol. 35, no. 7, pp. 1558–1563, 2013.
- [11] H. Han and K. Sohn, "Automatic illumination and color compensation using mean shift and sigma filter," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, pp. 978–986, 2009.
- [12] F. Sahba and A. Venetsanopoulos, "Breast mass detection using bilateral filter and mean shift based clustering," in *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pp. 88–94, July 2010.
- [13] H. Zhang, J. E. Fritts, and S. A. Goldman, "An entropy-based objective evaluation method for image segmentation," in *Storage and Retrieval Methods and Applications for Multimedia*, M. M. Yeung, R. W. Lienhart, and C. S. Li, Eds., vol. 5307 of *Proceedings of SPIE*, pp. 38–49, January 2004.
- [14] B. Chen, Y. Zhu, J. Hu, and M. Zhang, "On optimal estimations with minimum error entropy criterion," *Journal of the Franklin Institute*, vol. 347, no. 2, pp. 545–558, 2010.
- [15] T. Zeng, R. Wang, and F. Li, "SAR image autofocus utilizing minimum-entropy criterion," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1552–1556, 2013.
- [16] Y. A. Ghassabeh, T. Linder, and G. Takahara, "On some convergence properties of the subspace constrained mean shift," *Pattern Recognition*, vol. 46, no. 11, pp. 3140–3147, 2013.
- [17] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 2000.
- [18] Y. Wen, Y. Lu, and P. Shi, "Handwritten Bangla numeral recognition system and its application to postal automation," *Pattern Recognition*, vol. 40, no. 1, pp. 99–107, 2007.
- [19] N. M. Khan, R. Ksantini, I. S. Ahmad, and B. Boufama, "A novel SVM+NDA model for classification with an application to face recognition," *Pattern Recognition*, vol. 45, no. 1, pp. 66–79, 2012.
- [20] H. Zhou, G. Schaefer, M. E. Celebi, F. Lin, and T. Liu, "Gradient vector flow with mean shift for skin lesion segmentation," *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 121–127, 2011.
- [21] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [22] T. E. J. Behrens, M. W. Woolrich, M. E. Walton, and M. F. S. Rushworth, "Learning the value of information in an uncertain world," *Nature Neuroscience*, vol. 10, no. 9, pp. 1214–1221, 2007.
- [23] W. Li and G. X. Wang, "A purchase model with unascertained information," in *Proceedings of the 5th International Conference on Machine Learning and Cybernetics*, pp. 1840–1844, August 2006.
- [24] P. K. Parlewar and K. M. Bhurchandi, "A 4-quadrant curvelet transform for denoising digital images," *International Journal of Automation and Computing*, vol. 10, no. 3, pp. 217–226, 2013.
- [25] X. H. Wang, H. Wang, and Z. Y. Qui, "Vehicles detection based an adaptive median filter," *International Journal of Advances in Computing Technology*, vol. 4, no. 6, pp. 219–225, 2012.
- [26] Z. Zhang and G. Zhao, "Butterworth filter and Sobel edge detection to image," in *Proceedings of the 2nd International Conference on Multimedia Technology*, pp. 254–256, July 2011.
- [27] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Transactions on Neural Networks*, vol. 11, no. 5, pp. 1188–1193, 2000.
- [28] W. J. Liao and Z. Balzen, "LSSVM network flow prediction based on the self-adaptive genetic algorithm optimization," *Journal of Networks*, vol. 8, no. 2, pp. 507–512, 2013.
- [29] H. Ma, J. Chan, T. K. Saha, and C. Ekanayake, "Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 2, pp. 468–478, 2013.