

Pill Counting Using Watershed-based Segmentation Methods

Luis F. De-la-Rosa-Pineda, Victor M. Leyva-Osorio

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación,
México

{rp226570038, lo226570039, cs226570036, cc226570037}@alm.buap.mx

Abstract. This paper presents a methodology for automatic pill counting using digital image processing techniques. The proposed method starts with the acquisition of a digital image, which is converted to grayscale and smoothed using a Gaussian filter to reduce noise. The image is then binarized using Otsu's method, followed by morphological operations to remove small objects and generate markers. A distance transform and the Watershed algorithm are used to separate pills that are in contact. Finally, the segmented regions are identified and counted using connected components. The methodology was evaluated under controlled conditions using images of groups of pills placed in different positions, including cases where some pills were touching. The method successfully identified and counted 32 pills in the first test image and 17 pills in the second. These results show that the combination of thresholding, morphological operations, and Watershed segmentation can be used for automatic pill counting under the evaluated conditions.

Keywords: Digital image processing, pill counting, Otsu method, image segmentation, Watershed segmentation.

Conteo de pastillas basado en métodos de segmentación de Watershed

Resumen. Este artículo presenta una metodología para el conteo automático de pastillas mediante técnicas de procesamiento digital de imágenes. El método propuesto comienza con la adquisición de una imagen digital, la cual se convierte a escala de grises y se suaviza utilizando un filtro gaussiano para reducir el ruido. A continuación, la imagen se binariza mediante el método de Otsu, seguido de operaciones morfológicas para eliminar objetos pequeños y generar marcadores. Se emplean la transformada de distancia y el algoritmo de Watershed para separar las pastillas que se encuentran en contacto. Finalmente, las regiones segmentadas se identifican y cuentan utilizando componentes conexos. La metodología se evaluó en condiciones controladas utilizando imágenes de grupos de pastillas dispuestas en diferentes posiciones,

incluyendo casos en los que algunas pastillas se tocaban entre sí. El método identificó y contó con éxito 32 pastillas en la primera imagen de prueba y 17 en la segunda. Estos resultados demuestran que la combinación de umbralización, operaciones morfológicas y segmentación Watershed puede utilizarse para el conteo automático de pastillas bajo las condiciones evaluadas.

Palabras clave: Procesamiento digital de imágenes, conteo de pastillas, método de Otsu, segmentación de imágenes, segmentación Watershed.

1 Introduction

Digital image processing is an area focused on the analysis and manipulation of digital images using computational methods. These techniques make it possible to extract useful information from images and have become an important tool in different areas of science and engineering [3].

One of the applications of digital image processing is the automatic detection and counting of objects. In this type of application, an image can be processed to separate the objects of interest from the background and then identify each object individually. However, the counting process becomes more difficult when two or more objects are touching, since after binarization they can appear as a single connected region. This makes it necessary to use additional segmentation techniques before performing the final count [7].

For pill counting, digital image processing provides an alternative to manual counting by allowing pills to be detected from a captured image and counted from the regions obtained after segmentation. Previous works have shown that operations such as grayscale conversion, filtering, thresholding and mathematical morphology can be used for this purpose [5].

The objective of this work is to develop and evaluate a methodology for automatic pill counting using digital image processing. The main problem addressed is the separation of pills that are in contact, since they can be detected as a single object during conventional connected component analysis. The proposed methodology combines grayscale conversion, Gaussian filtering, Otsu thresholding, morphological operations, distance transform and marker based Watershed segmentation. Finally, connected component labeling is used to determine the number of pills present in the processed image.

2 Related Work

Different image processing techniques have been used for automatic object counting and for the separation of objects that are in contact. These methods usually combine image preprocessing, segmentation and connected component analysis according to the characteristics of the objects and the conditions in which the images are acquired. Phromlikhit et al. [5] proposed a tablet counting

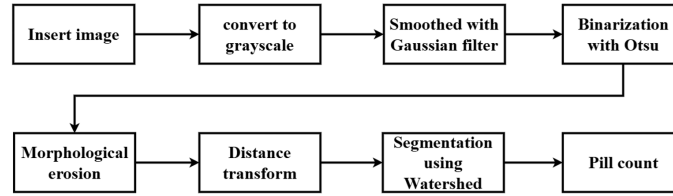


Fig. 1. Block diagram of the proposed methodology.

machine based on image processing. Their method uses images acquired with a digital camera and applies grayscale conversion, noise filtering, thresholding, erosion and dilation. After processing the image, connected component labeling and centroid information are used to determine the number of tablets. This work shows that relatively simple image processing operations can be used for automatic tablet counting. However, the separation of pills that remain connected after thresholding requires additional consideration.

For objects that are touching, Watershed segmentation provides an alternative for separating connected regions. Vincent and Soille [8] presented an efficient Watershed algorithm for digital grayscale images based on an immersion process. In this approach, the image is interpreted as a topographic surface and the resulting watershed lines define the boundaries between regions.

The use of Watershed for object counting has also been studied in applications where several objects can form connected groups. Tan et al. [7] developed a method for the segmentation and counting of touching rice grains. Their approach combines Watershed segmentation with corner detection and a neural network to reduce over segmentation and under segmentation problems. Their results also show that the way in which objects touch and their shape can affect the segmentation process. Based on these works, the methodology presented in this paper uses a combination of thresholding, morphological processing and marker based Watershed segmentation. Unlike approaches based only on connected components, the use of markers and the distance transform is intended to separate pills that remain connected after binarization before performing the final count.

3 Methodology

The methodology used for pill counting is shown in Figure 1. The process consists of a sequence of image processing operations, starting with image acquisition and ending with the identification and counting of the segmented objects. MATLAB was used to implement and evaluate the proposed methodology.

3.1 Image Acquisition

The first step consists of acquiring the image containing the pills to be counted. The images are captured in RGB format, where each pixel is represented by three components corresponding to the red (R), green (G), and blue (B) channels.

After loading the image in MATLAB, it is represented as a numerical matrix that can be processed by the following stages of the methodology. Image acquisition is an important step because factors such as lighting, camera resolution, shadows, and reflections can affect the segmentation process.

3.2 Grayscale Conversion

Before applying the segmentation methods, the RGB image is converted to grayscale. This reduces the image from three color channels to a single intensity channel, since the proposed method uses intensity information rather than color information [3].

The grayscale intensity Y is calculated from the RGB components using the following expression [3]:

$$Y = 0.2989R + 0.5870G + 0.1140B, \quad (1)$$

where R , G , and B represent the intensity values of the red, green, and blue channels, respectively.

3.3 Gaussian Smoothing

After the grayscale conversion, a Gaussian filter is applied to reduce small intensity variations and noise that could affect the subsequent segmentation stages. Gaussian smoothing is performed by convolving the image with a kernel whose values are determined by a Gaussian function [3].

The two-dimensional Gaussian function can be expressed as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (2)$$

where σ controls the amount of smoothing applied to the image. A larger value of σ produces stronger smoothing, while a smaller value preserves more local variations.

For the selected configuration of the proposed method, a value of $\sigma = 2$ was used. This value provided enough smoothing to reduce the small amount of noise present in the test images without excessively blurring the boundaries of the pills.

3.4 Otsu Image Binarization

After smoothing the grayscale image, the next step is to separate the pills from the background. For this purpose, the image is converted into a binary representation, where each pixel is assigned to either the background or the objects of interest.

The binarization process is performed using a threshold T . For an input image $I(x, y)$, the resulting binary image $g(x, y)$ can be defined as:

$$g(x, y) = \begin{cases} 1, & \text{if } I(x, y) \geq T, \\ 0, & \text{if } I(x, y) < T. \end{cases} \quad (3)$$

Instead of using a fixed threshold, the value of T is automatically determined using Otsu's method. This method selects a threshold from the intensity histogram by maximizing the variance between two classes of pixels [4].

For an image of size $M \times N$ with L intensity levels $[0, 1, \dots, L-1]$, let n_i represent the number of pixels with intensity i . The probability of occurrence of each intensity level is defined as:

$$p_i = \frac{n_i}{MN}. \quad (4)$$

For a possible threshold k , the pixels are divided into two classes: C_1 , containing intensity levels from 0 to k , and C_2 , containing intensity levels from $k+1$ to $L-1$.

The probabilities associated with these two classes are:

$$P_1(k) = \sum_{i=0}^k p_i \quad (5)$$

and

$$P_2(k) = \sum_{i=k+1}^{L-1} p_i = 1 - P_1(k). \quad (6)$$

The mean intensity of each class is then calculated as:

$$\mu_1(k) = \frac{1}{P_1(k)} \sum_{i=0}^k i p_i, \quad (7)$$

$$\mu_2(k) = \frac{1}{P_2(k)} \sum_{i=k+1}^{L-1} i p_i. \quad (8)$$

The global mean intensity of the image is given by:

$$\mu_G = \sum_{i=0}^{L-1} i p_i. \quad (9)$$

Using these values, the between class variance for each possible threshold k is calculated as:

$$\sigma_B^2(k) = P_1(k) [\mu_1(k) - \mu_G]^2 + P_2(k) [\mu_2(k) - \mu_G]^2. \quad (10)$$

Finally, Otsu's method selects the threshold K that maximizes the between class variance:

$$K = \arg \max_{0 \leq k \leq L-1} \{\sigma_B^2(k)\}. \quad (11)$$

In this way, the threshold is obtained directly from the intensity distribution of each image instead of using a fixed value for all test images.

3.5 Small Objects Removal

After binarization, small regions or artifacts that do not correspond to pills may remain in the binary image. To remove these regions, a cleaning process based on the area of the connected components is applied.

First, the connected components in the binary image are identified and their areas are calculated. Components with an area smaller than a predefined threshold of 1000 pixels are removed. This step reduces the presence of small regions that could affect the subsequent segmentation and counting process.

3.6 Marker Generation Through Morphological Erosion

After removing small objects, morphological erosion is applied to the binary image. Erosion is one of the fundamental operations of mathematical morphology and reduces the boundaries of foreground regions according to the shape and size of a structuring element [6].

For a binary image represented by the set A and a structuring element B , erosion can be defined as [6]:

$$A \ominus B = \{z \mid B_z \subseteq A\}, \quad (12)$$

where B_z represents the structuring element B translated to position z . Therefore, a point remains in the resulting set only when the structuring element fits completely inside the object at that position.

In the proposed methodology, a disk shaped structuring element is used to reduce the regions corresponding to the pills. This operation generates smaller regions located approximately at the center of each object. These regions are used as internal markers for the subsequent Watershed segmentation process.

For the selected configuration, a disk with a radius of 72 pixels was used. The effect of changing this parameter is evaluated in the experimental results, since the size of the structuring element influences the separation and preservation of the markers.

3.7 Distance Transform

The distance transform is used to calculate the distance from each foreground pixel to the nearest background pixel in a binary image. As a result, pixels located near the center of an object have larger distance values, while pixels close to its boundary have smaller values [2].

Using the Euclidean distance, the distance transform can be expressed as:

$$D(x, y) = \min_{(x', y') \in B} \sqrt{(x - x')^2 + (y - y')^2}, \quad (13)$$

where $D(x, y)$ represents the distance from the foreground pixel (x, y) to the nearest background pixel, and B is the set of pixels belonging to the image background.

For the proposed methodology, the negative distance transform is used:

$$D_n(x, y) = -D(x, y). \quad (14)$$

By inverting the distance values, the central regions of the pills become local minima. This representation is used in the following Watershed segmentation stage to help separate objects that are in contact.

3.8 Segmentation Using Watershed

The Watershed algorithm is a segmentation method based on the topographic interpretation of an image. In this representation, low intensity values can be interpreted as valleys or basins, while higher values represent ridges. The segmentation process can then be described as a flooding process that starts from the local minima of the image [1, 8].

For the proposed methodology, the negative distance transform obtained in the previous step is used as the input for the Watershed algorithm. Since the central regions of the pills correspond to local minima in this representation, they can be used to guide the segmentation of objects that are in contact.

Let $M_1, M_2, \dots, M_R$ represent the set of regional minima present in an image $g(x, y)$. For a flooding level n , the set of points below this level can be expressed as [8]:

$$T_n = (x, y) \mid g(x, y) < n. \quad (15)$$

For each regional minimum M_i , the portion of its catchment basin that has been flooded at level n is defined as:

$$C_n(M_i) = C(M_i) \cap T_n, \quad (16)$$

where $C(M_i)$ represents the catchment basin associated with the minimum M_i .

As the flooding level increases, different catchment basins begin to approach each other. When regions associated with different minima would merge, a separation line is created between them. The collection of these lines forms the Watershed boundaries that divide the image into different regions [8].

In this work, a marker based Watershed approach is used. The regions previously obtained through morphological erosion are used as internal markers for the pills. These markers guide the segmentation process and help reduce the generation of unnecessary regions caused by small local variations in the image.

3.9 Final Cleaning and Object Counting

After the Watershed segmentation, a final cleaning process is applied to remove small regions that do not correspond to valid pills. These regions may appear due to small segmentation errors or fragments generated during the separation of objects.

The cleaning process is based on the area of the segmented regions. Connected components with an area smaller than 15000 pixels are removed from the binary

image. This threshold was selected according to the approximate size of the pills in the test images.

Once the final binary image is obtained, the remaining connected components are identified using 8-connectivity. For a pixel $p = (x_p, y_p)$, its 8-neighborhood can be defined as [3]:

$$N_8(p) = (x, y) \mid \max(|x - x_p|, |y - y_p|) = 1. \quad (17)$$

Using connected component labeling, each independent region in the binary image is assigned a different label. The total number of detected pills is then obtained from the number of labeled connected components. If k represents the total number of valid labeled regions, the final count is given by $N = k$.

Therefore, after the cleaning and segmentation stages, each remaining connected component is considered as one detected pill.

4 Results and Experimentation

This section presents the results obtained with the proposed methodology and evaluates the effect of some of the parameters used during image processing. In addition to the final pill count, the intermediate results are analyzed to identify the conditions in which the segmentation performs correctly and the cases where errors may occur.

4.1 Experimental Setup

The proposed methodology was implemented in MATLAB and evaluated using images containing pills with different shapes and arrangements. The test images include circular pills and elongated capsules, as well as cases where multiple objects are in contact.

The selected configuration used a Gaussian filter with $\sigma = 2$. After binarization using Otsu's method, connected components smaller than 1000 pixels were removed. A disk-shaped structuring element with a radius of 72 pixels was then used to generate the internal markers for the Watershed segmentation. After segmentation, regions with an area smaller than 15000 pixels were removed before performing the final connected-component labeling. The parameters of the selected configuration are summarized in Table 1.

These values were selected from the evaluated parameter ranges and were used for the final pill counting experiments. Some of the parameters were subsequently varied to evaluate their influence on the segmentation and final counting results.

4.2 Image Processing Results

Two images were initially used to evaluate the proposed methodology. The first image contains circular pills, while the second contains elongated capsules. In both cases, some objects are in contact, allowing the segmentation process to be evaluated with different pill shapes and arrangements.

Table 1. Selected parameters used in the proposed methodology.

Parameter	Selected value
Gaussian smoothing (σ)	2
Initial area threshold	1000 pixels
Structuring element	Disk
Structuring element radius	72 pixels
Final area threshold	15000 pixels
Connectivity	8

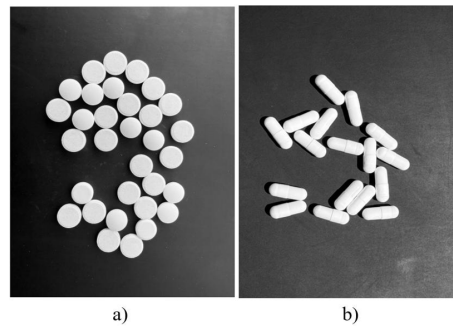


Fig. 2. Preprocessed images after grayscale conversion and Gaussian smoothing: (a) circular pills and (b) elongated capsules.

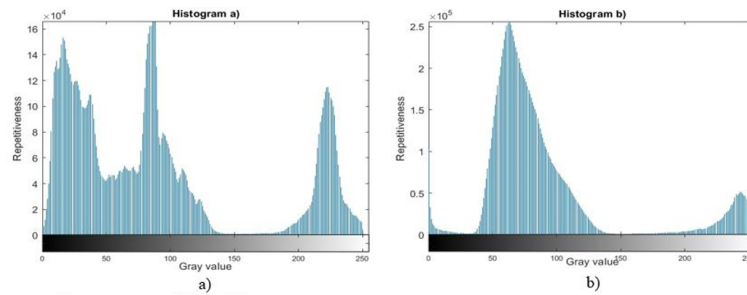


Fig. 3. Intensity histograms of the test images.

The images were first converted to grayscale, removing the color information and keeping the intensity information used in the following processing stages. A Gaussian filter with $\sigma = 2$ was then applied to reduce small intensity variations and noise. The resulting preprocessed images are shown in Figure 2.

The intensity histograms of the test images are shown in Figure 3. Differences can be observed between the gray-level distributions of both images. These variations can be produced by factors such as lighting, reflections, and the conditions in which each image was captured.

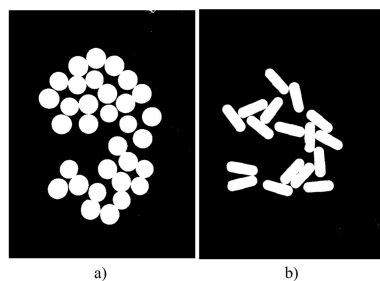


Fig. 4. Binarized images using Otsu's method.

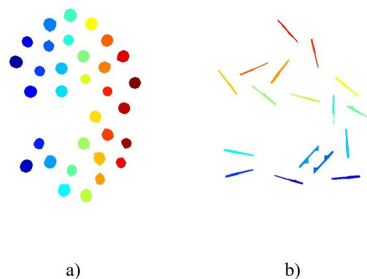


Fig. 5. Internal markers obtained through morphological erosion.

Because the intensity distribution is not the same for both images, Otsu's method was used to calculate the binarization threshold independently for each case. The resulting binary images are shown in Figure 4, where the pills are separated from the background.

After removing small regions, morphological erosion was applied using a disk-shaped structuring element with a radius of 72 pixels. The resulting internal markers are shown in Figure 5. These markers represent the regions used to identify the individual pills during the Watershed segmentation.

The distance transform was then calculated from the binary images. The distance values increase toward the central regions of the objects, and the negative of this transformation was used for the subsequent Watershed segmentation.

Figure 6 shows the result obtained after applying the Watershed algorithm. The generated boundaries allow several regions that were previously in contact to be separated.

Finally, the segmented images were cleaned by removing regions below the final area threshold. The resulting images are shown in Figure 7. After connected-component labeling, 32 regions were detected in the first image and 17 regions in the second image.

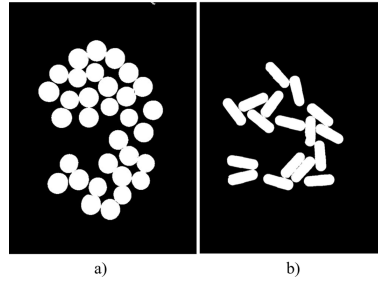


Fig. 6. Watershed segmentation of the test images.

Table 2. Effect of the structuring element radius on the final pill count.

Radius [px]	Image A	Image B
36	22	16
48	27	16
60	32	16
72	32	17
84	32	20

Table 3. Effect of Gaussian smoothing on the final pill count.

σ	Image A	Image B
1	32	17
2	32	17
3	32	16
4	32	16

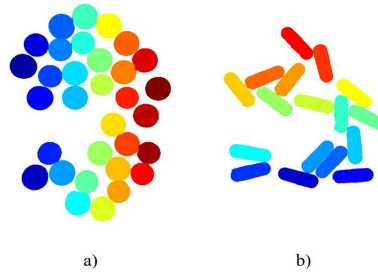


Fig. 7. Final segmented images used for pill counting.

4.3 Parameter Evaluation

To evaluate the influence of the parameters used in the proposed methodology, three parameters were varied independently: the radius of the structuring element used for marker generation, the value of σ used for Gaussian smoothing, and the final area threshold. During each experiment, only one parameter was modified while the remaining parameters were kept at their initial values.

Table 4. Effect of the final area threshold on the pill count.

Area threshold [px]	Image A	Image B
1000	33	17
5000	32	17
10000	32	17
15000	32	17
20000	32	17
25000	32	17

The first experiment evaluated the radius of the disk-shaped structuring element. The results are shown in Table 2.

The results show that the structuring element radius has an important effect on the final segmentation. Small radius values produced an incorrect number of detected regions, especially in Image A. Increasing the radius improved the result until a radius of 72 pixels, where 32 and 17 pills were detected in Images A and B, respectively. However, increasing the radius to 84 pixels caused Image B to produce 20 regions. This shows that increasing the radius does not always improve the segmentation and that its value must be selected according to the size and shape of the objects.

The second experiment evaluated the effect of Gaussian smoothing. The results are presented in Table 3.

Image A maintained a count of 32 pills for all the evaluated values of σ . Image B produced the expected count of 17 pills with $\sigma = 1$ and $\sigma = 2$, while increasing the value to 3 or 4 resulted in 16 detected regions. This indicates that, for the evaluated images, the elongated capsules in Image B were more sensitive to stronger smoothing.

Finally, the final area threshold was varied between 1000 and 25000 pixels. The results are shown in Table 4.

Compared with the previous parameters, the final area threshold showed less influence on the counting results within the evaluated range. A threshold of 1000 pixels produced one additional region in Image A.

From 5000 to 25000 pixels, both images maintained counts of 32 and 17 pills. The selected value of 15000 pixels was therefore kept for the final configuration.

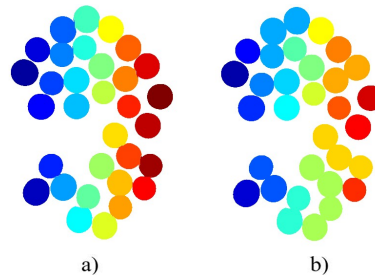
4.4 Pill Counting Results

After evaluating the different parameter values, the selected configuration was used for the final counting process. This configuration uses $\sigma = 2$, a structuring element radius of 72 pixels, an initial area threshold of 1000 pixels, and a final area threshold of 15000 pixels. The final counting results are summarized in Table 5.

The selected configuration produced 32 detected pills in Image A and 17 in Image B. These results are consistent with the final segmented images shown previously. The parameter evaluation also showed that the correct separation of

Table 5. Final pill counting results using the selected configuration.

Test image	Detected pills
Image A	32
Image B	17

**Fig. 8.** Effect of the structuring element radius on Image A: (a) $r = 36$, resulting in under segmentation and 22 detected regions; (b) $r = 72$, resulting in 32 detected regions.

the objects depends mainly on the generation of appropriate internal markers, particularly on the radius used during morphological erosion.

4.5 Analysis and Validity of Results

The experiments show that the proposed methodology can correctly separate and count the pills in the two test images when appropriate parameter values are used. With the selected configuration, 32 pills were detected in Image A and 17 capsules in Image B.

The parameter evaluation showed that the radius of the structuring element had the greatest influence on the final count among the parameters evaluated. Figure 8 shows the effect of using different radius values in Image A. With $r = 36$, some pills are not separated correctly, resulting in 22 detected regions. With the selected radius of $r = 72$, the pills are separated individually and 32 regions are obtained.

A different effect can be observed in Image B, as shown in Figure 9. With the selected radius of $r = 72$, 17 regions are obtained.

When the radius is increased to 84 pixels, some capsules are divided into more than one region, increasing the final count to 20. In this case, the larger radius produces over segmentation.

These results indicate that the generation of internal markers is an important stage of the methodology. If the selected radius is not appropriate for the objects in the image, the number of regions obtained after Watershed segmentation can change and produce an incorrect final count. A small radius produced under segmentation in Image A, while a larger radius produced over segmentation in Image B.

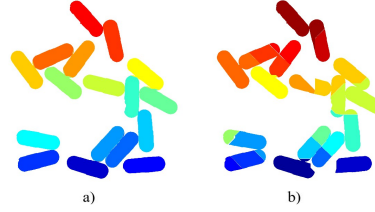


Fig. 9. Effect of the structuring element radius on Image B: (a) $r = 72$, resulting in 17 detected regions; (b) $r = 84$, producing over segmentation and 20 detected regions.

The experiments with Gaussian smoothing showed a smaller effect on the final count. Image A maintained 32 detected regions for all the evaluated values of σ . Image B maintained 17 regions with $\sigma = 1$ and $\sigma = 2$, but the count decreased to 16 when σ was increased to 3 and 4. This indicates that the result for Image B was more sensitive to stronger smoothing within the evaluated range.

The final area threshold showed the smallest influence among the three evaluated parameters. A threshold of 1000 pixels resulted in 33 detected regions in Image A, while Image B maintained a count of 17. For threshold values between 5000 and 25000 pixels, both images maintained counts of 32 and 17 regions, respectively.

Although the selected configuration produced the expected results for both test images, the experiments also show some limitations of the methodology. The same parameter values may not provide the same segmentation result for objects with different shapes and sizes. In particular, the generation of internal markers depends on the radius of the structuring element, which can lead to under-segmentation or over-segmentation when an inappropriate value is used.

Therefore, the results obtained in this work validate the methodology for the evaluated images and parameter ranges. Additional images should be captured under different conditions, considering variations in pill arrangement, size, and shape, as well as changes in lighting and background, to evaluate the performance of the method under different scenarios.

5 Conclusions

In this work, a methodology for automatic pill counting using digital image processing was presented. The method combines grayscale conversion, Gaussian smoothing, Otsu thresholding, morphological operations, distance transform, and marker based Watershed segmentation. With the selected configuration, the method detected 32 pills in Image A and 17 capsules in Image B.

Although the methodology is based on classical image processing tools and does not use artificial intelligence, their combination made it possible to successfully count the pills in the images evaluated in this work. In particular, the use of morphological operations, distance transform, and Watershed segmentation allowed pills that were in contact to be separated before the final counting stage.

The experiments also showed that the selection of the processing parameters can affect the segmentation result.

One limitation of the proposed methodology occurs when pills overlap. Since the method works with two dimensional images, an overlapping pill can partially hide another object, making their separation more difficult with the techniques used in this work. Therefore, the current methodology is mainly intended for cases where the pills are visible but may be in contact with each other.

Future work could include images captured under different conditions, with variations in pill arrangement, size, shape, lighting, and background. The evaluation of methods for handling partially overlapping pills and automatically adjusting the processing parameters could also be considered.

References

1. Beucher, S., Lantuéjoul, C.: Use of watersheds in contour detection. In: Proceedings of the International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation. Rennes, France (1979)
2. Borgefors, G.: Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing* 34(3), 344–371 (1986)
3. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Pearson, 4 edn. (2018)
4. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9(1), 62–66 (1979)
5. Phromlikhit, C., Cheevasuvit, F., Yimman, S.: Tablet counting machine base on image processing. In: 2012 Biomedical Engineering International Conference (BMEiCON). pp. 400–403 (2012)
6. Soille, P.: *Morphological Image Analysis: Principles and Applications*. Springer, Berlin, Heidelberg, 2 edn. (2003)
7. Tan, S., Ma, X., Mai, Z., Qi, L., Wang, Y.: Segmentation and counting algorithm for touching hybrid rice grains. *Computers and Electronics in Agriculture* 162, 493–504 (2019)
8. Vincent, L., Soille, P.: Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(6), 583–598 (1991)