A System for Automatic Counting the Number of Collembola Individuals on Petri Disk Images

André R.S. Marçal[1] and Cristina M.R. Caridade[2]

[1] Faculdade de Ciências, Universidade do Porto
DMA, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
[2] Instituto Superior de Engenharia de Coimbra,
Rua Pedro Nunes, Quinta da Nora , 3030-199 Coimbra, Portugal

Abstract. This paper describes an image processing system developed for automatic counting the number of collembola individuals on petri disks images. The system uses image segmentation and mathematical morphology techniques to identify and count the number of collembolans. The main challenges are the specular reflections at the edges of the circular samples and the foam present in a number of samples. The specular reflections are efficiently identified and removed by performing a two-stage segmentation. The foam is considered to be noise, as it is at cases difficult to discriminate between the foam and the collembola individuals. Morphological image processing tools are used both for noise reduction and for the identification of the collembolans. A total of 38 samples (divided in 3 groups according to their noise level) were tested and the results produced from the automatic system compared to the values available from manual counting. The relative error was on average 5.0% (3.4% for good quality samples, 4.6% for medium quality and 7.5% for poor quality samples).

1 Introduction

The ability of organisms to detect contaminated soil can be used to estimate the stress potential of a particular soil [1]. Avoidance behaviour tests as early screening tools in this process have been proposed using collembola species [1]. Collembolan (Folsomia candida) are small invertebrates (1-5 mm), wingless hexapods with antennae always present [2]. These organisms are commonly used in laboratory toxicity tests with polluted soils [1]. Collembola can be kept on moist soil or peat moss, but they thrive best and are most useful when cultured in small jars on a plaster of Paris or petri disk (cylindrical plastic containers typically with 7 cm in diameter and 6 cm in height) [1] [2]. The evaluation of the number of collembola individuals in petri disk samples is thus an important task that is performed on a regular basis. The counting process currently in place at the Instituto de Ambiente e Vida, University of Coimbra, is based on a direct human intervention, and is therefore laborious and time consuming. The purpose of this work was to develop a fully automatic algorithm for performing such task.

The images are acquired using a standard digital camera, coupled to a binocular system, producing RGB colour images (24 bits) with 1024 by 768 pixels. The

illumination is done in a straightforward way using a combination of natural and artificial light. Although an effort is made to minimise reflections, there is always specular reflections at the edges of the circular samples. Figure 1 shows an example of the images acquired (sample C6), here converted to greyscale. There is a bright background and a dark circular area, corresponding to the petri disk sample, which is the area of interest in the image. However, there are very bright areas in the petri disk due to the specular reflection of the incident light, and other grey areas corresponding to foam, which in this case can be seen as noise. The sample presented in Figure 1 has a significant contamination by foam, but there are other samples that are not so heavily contaminated. The presence of foam is considered noise, and poses serious difficulties not only to an automatic counting system but also to the identification of the collembola individuals by a human operator. The histograms of the images are bimodal, as it can be seen in the example of Figure 1. However, there is a variable occupancy of the intermediate levels, which depends both on the amount of foam in the sample disk and on the illumination conditions.
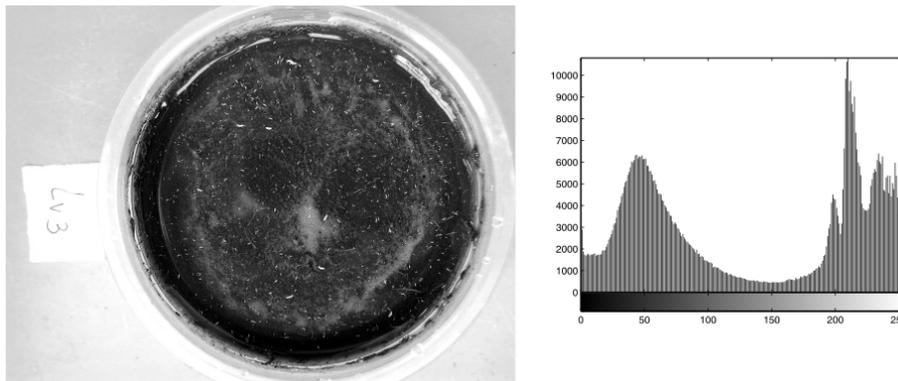

Fig. 1. Example of a petri disk image (C6) and the corresponding histogram

2   Methodology

The algorithm for automatic counting the number of collembola individuals in petri disk images was developed using MATLAB [3]. It can be divided in three parts: (i) the selection of the interest area of the image, (ii) noise removal and segmentation, and (iii) the actual identification and counting of the collembolans.

2.1   Selection of the Interest Area

Let G be the greyscale image being processed. The image is initially divided in two subgroups – the background and the sample disk (D_0) – by thresholding segmentation. The thresholded image T is obtained from G using equation (1), where $g(x, y)$ is the greylevel in G of a pixel p with coordinates x and y, and

$t(x, y)$ is the 1-bit value for that pixel in T [4]. The new image D_0 is obtained from the original image G by (2).

$$t(x,y) = \begin{cases} 1 & if \;\; g(x,y) \leq h \\ 0 & if \;\; g(x,y) > h \end{cases} \tag{1}$$

$$D_0 = G \cdot T \tag{2}$$

Using a suitable value for h, this process will separate most of the background from the sample disk. However, it will also eliminate other bright parts of the image, including collembola individuals. As an illustration of this, Figure 2 shows the result of the initial segmentation applied to the sample image C6 (Figure 1) using a threshold value of 0.5 (in a range of 0 to 1). Although the image histograms are bimodal, the two areas are not clearly separated. The determination of the threshold value is done by searching for the first greylevel above the first peak maximum with occupancy below L pixels (e.g. $L = 50$). In the segmented image D_0 of Figure 2, the background appears in white and the foreground in black. However, as this two area separation does not match the requirements for extraction of the interest area (the sample disk) further processing is required. After the first segmentation the geometric centre of D_0 is calculated using equations (3) and (4) .

$$x_c = \sum x_i \qquad \forall i \in D_0 \tag{3}$$

$$y_c = \sum y_i \qquad \forall i \in D_0 \tag{4}$$

An iterative process is used to identify the location of the pixels contaminated by specular reflection, as well as the size of the full disk. Initially a circle with a radius R_0 centred in x_c, y_c is considered as a valid area. The radius of this circle (R) is then gradually increased up to the maximum value before the specular reflections (R_1). On a second stage the radius is further increased to the maximum radius before reaching the end of the disk (R_2). The iterative process works as follows:

1. Establish a circular area centred on x_c, y_c with a radius of R_0 pixels (e.g. $R_0 = 100$)
2. Increase the radius R by a step increment d (e.g. $d = 1$ pixel)
3. Divide the arc circle of the additional area in 8 sectors of 45 degrees
4. Compute the fraction of background pixels (f) on each sector
5. If $f > f_1$ for any of the 8 sectors, $R_1 = R$, else go to 2.

A similar process can be used to compute R_2 (instead of R_1), by replacing the value f_1 by a higher value f_2, in the step 5 above. In fact the two radii R_1 and R_2 are computed on a single iterative process, but the algorithm presented here is for a single radius for the sake of simplicity. An example of the application of this process is presented in Figure 3, for the sample C6, using $R_0 = 100$, $d = 1$, $f_1 = 0.015$ and $f_2 = 0.100$. In this figure the centre and circles of radius R_1 and R_2 are presented over the original image. The interest area D_1 is then
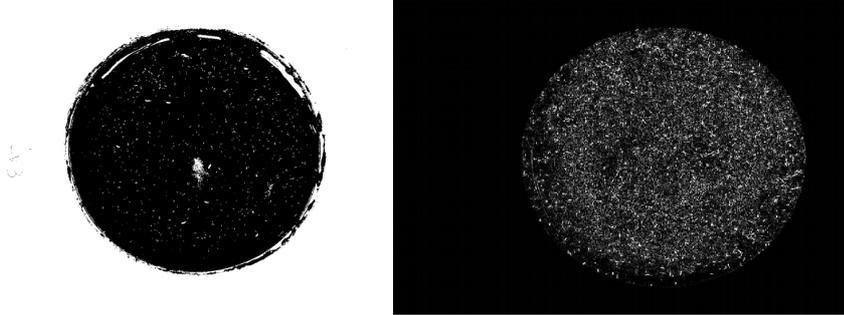
Fig. 2. Example of two intermediate processing stages: binary image D_0 (left) and greyscale image D_1 (right), both from sample C6

selected from the original image G, using (5) and (6), where C is a binary image corresponding to the disk of radius R_1 centred on x_c, y_c.

$$c(x,y) = \begin{cases} 1 & if \quad (x - x_c)^2 + (y - y_c)^2 < R_1^2 \\ 0 & if \quad (x - x_c)^2 + (y - y_c)^2 \geq R_1^2 \end{cases} \tag{5}$$

$$D_1 = G \cdot C \tag{6}$$

2.2   Noise Removal and Segmentation

The quality of the images is not uniform, particularly to what concerns the foam present in the sample, which can be seen as noise. The images available were separated in three groups, according to their level of noise: (A) good quality images with low level of foam/noise, (B) medium quality images, (C) poor quality images, which are highly contaminated by foam/noise. Figure 4 shows an example of images from group A (A4) and from group C (C2). As it can be seen in this figure, the presence of foam greatly diminishes the ability to identify the collembola individuals visually, and that is also the case for the automatic system. Another example of an image from group C (test image C6) is presented in Figure 1.

Before attempting to identify and count the collembola individuals it is necessary to reduce the noise. This is done by morphological image processing. The elementary morphological operators dilation ($\oplus$) and erosion ($\ominus$) are used, with a structuring element B [4]. The morphological operator opening ($\circ$) of A by B is defined as an erosion followed by a dilation: $A \circ B = (A \ominus B) \oplus B$. Alternatively, the morphological operator closing ($\bullet$) of A by B is defined as a dilation followed by an erosion: $A \bullet B = (A \oplus B) \ominus B$ [5]. The procedure used to remove the noise is inspired in the example presented in [5] for fingerprint images. Here the purpose is to remove the foam from image D_1. A sequence of opening and closing is applied to the image D_1, using as the structuring element B a 3x3 square, resulting on an image D_2 (7). This process removes all objects bigger than the structuring element from D_1, leaving only small objects that in this case are the collembolans.
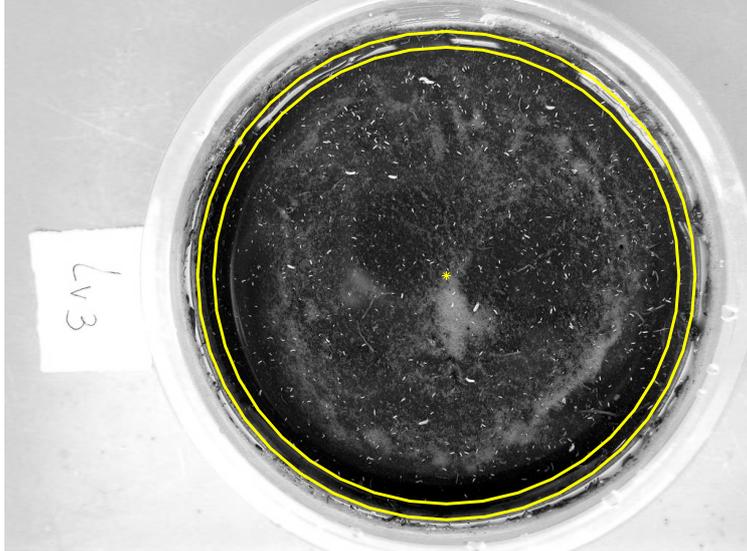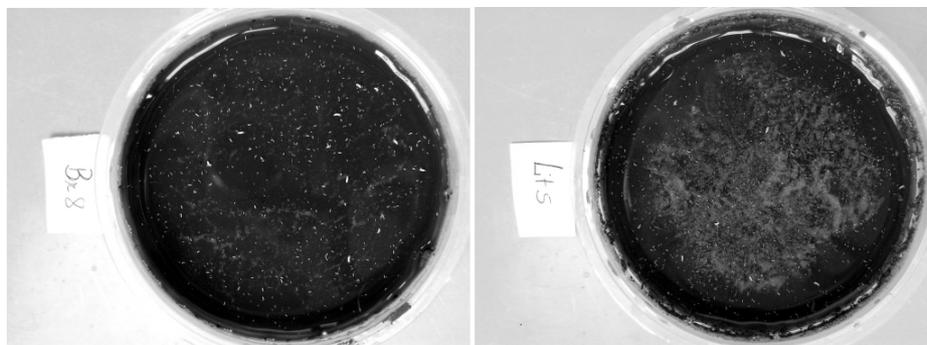
Fig. 3. The original test image C6 with the circular limit before reflection and full disk overlaid

$$D_2 = (D_1 \circ B) \bullet B \tag{7}$$

The image produced by the noise removing process described (D_2) is a greyscale image, with positive values only for the area of interest. The next step is to perform a segmentation of this image in order to have a binary image with non-null values only for the locations of the collembolans. As the greylevel characteristics of the images are not consistent throughout, an adaptive histogram thresholding is used. The images are divided in 48 sections of 128 by 128 pixels each. The basic adaptive thresholding described in [4] is applied to each section of the image, with the threshold value obtained automatically for each section. A large number of the sections are in fact null values, as the initial segmentation removed the background area and the edges of the sample where specular reflection occurs. The resulting binary image D_3 is then used for the subsequent identification of individuals and counting.

2.3   Identification and Counting

The binary image resulting from the segmentation and noise reduction stages (D_3) has a number of background ($= 0$) and foreground ($= 1$) pixels. An initial estimation of the number of collembola individuals (n_i) is done by identifying the number of objects in D_3. In this context, an object is defined as a connected component of the binary image, for a given definition of adjacency [5]. The 8-neighbourhood adjacency is used, where all 8 pixels of the 3x3 window centred on a pixel are considered to be neighbours. The number of connected components of

Fig. 4. A sample with low level of noise (A4, left) and one highly contaminated by foam/noise (C2, right)
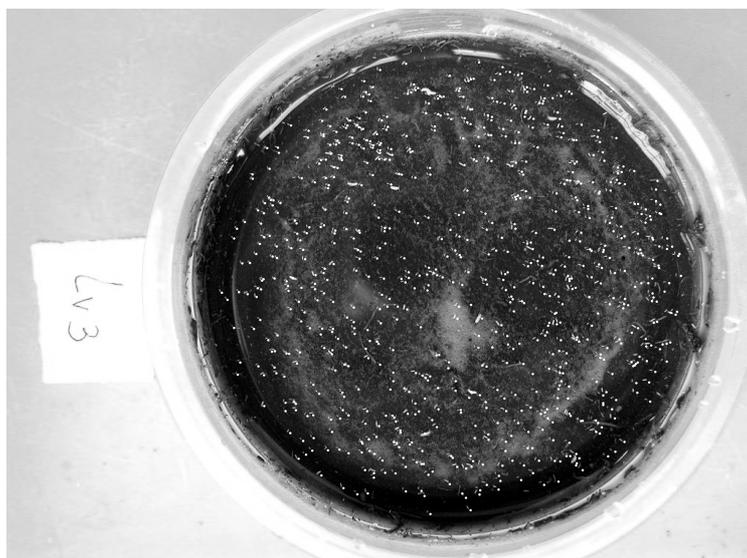
Fig. 5. The original test image C6 with the identified collumbelans overlaid

the binary image D_3 are computed using a MATLAB Image Processing Toolbox function [3] [5]. The result for the test image C6 is presented in Figure 5, where the objects identified in the binary image are overlaid in white over the original image.

The number of objects identified by the process described (n_i) corresponds to the number of collembolans in the circular area of radius R_1. A final step is required to compensate for the arc circle eliminated due to the specular reflection. An estimate of the total number of collembolans in the whole sample (n_f) is

obtained using (8). This assumes a homogeneous distribution of collembolans in the sample disk, which might not always be the case.

$$n_f = n_i \times R_2^2/R_1^2 \tag{8}$$

3   Results

A total of 38 images were available for testing: 9 with low level of foam/noise (group A), 20 with intermediate level of foam/noise (group B), and 9 highly contaminated by foam/noise (group C). The automatic counting system described in section 2 was applied to these images, using the following parameters: $L = 50$, $R_0 = 100$, $d = 1$, $f_1 = 0.015$ and $f_2 = 0.100$. The results were compared to the values available from the manual counting. Table 3 shows the results from both methods, as well as the difference and relative error of the automatic counting, for the images of group A. The maximum relative error is 7.8% but the average relative error is only 3.4%.

Tables 2 and 3 show the results from the automatic counting method for the images of groups B and C. The maximum relative error for the images of group B is 14.5% and the average relative error is 4.6%. Although the maximum error is high, in only 2 out of the 20 images tested the relative error is above 10%. For the heavily contaminated images of group C the results are, as expected, considerably worst. The average error is 7.5% and the maximum error reaches 24.8%. A third of the images of group C have a relative error above 10%. These 3 cases are all overestimates. However, apart from these 3 cases, all remaining samples in group C were counted by the automatic system with very good accuracy (relative error below 4%). That is the case for example for the test image C6 used for the illustration of the method throughout the paper. In that case, although the sample was highly contaminated by noise (foam), the automatic system counting and the manual result only differ by 1.1% (813 instead of 822 collembolans). The reason for the high variance in the results of the group C images is not fully understood. The foam in these images is highly widespread

Table 1. Results of the automatic counting system applied to the 9 images of type A (low level of noise)

SAMPLE	MANUAL	AUTOMATIC	DIFFERENCE	ERROR
A1	696	663	-33	4.7%
A2	1139	1050	-89	7.8%
A3	853	837	-16	1.9%
A4	649	654	5	0.8%
A5	1050	1064	14	1.3%
A6	740	758	18	2.4%
A7	897	898	1	0.1%
A8	891	826	-65	7.3%
A9	1011	1050	39	3.9%

Table 2. Results of the automatic counting system applied to the 9 images of type B (low level of noise)

SAMPLE	MANUAL	AUTOMATIC	DIFFERENCE	ERROR
B1	656	751	95	14.5%
B2	847	815	-32	3.8%
B3	730	720	-10	1.4%
B4	676	705	29	4.3%
B5	636	657	21	3.3%
B6	758	670	-88	11.6%
B7	706	735	29	4.1%
B8	765	728	-37	4.8%
B9	902	908	6	0.7%
B10	891	829	-62	7.0%
B11	967	943	-24	2.5%
B12	939	895	-44	4.7%
B13	880	865	-15	1.7%
B14	936	925	-11	1.2%
B15	832	773	-59	7.1%
B16	973	901	-72	7.4%
B17	826	846	20	2.4%
B18	786	798	12	1.5%
B19	1345	1414	69	5.1%
B20	939	974	35	3.7%

Table 3. Results of the automatic counting system applied to the 9 images of type C (low level of noise)

SAMPLE	MANUAL	AUTOMATIC	DIFFERENCE	ERROR
C1	843	998	155	18.4%
C2	797	767	-30	3.8%
C3	607	695	88	14.5%
C4	897	921	24	2.7%
C5	669	835	166	24.8%
C6	822	813	-9	1.1%
C7	986	997	11	1.1%
C8	466	468	2	0.4%
C9	731	739	8	1.1%

and irregular. In some cases the filtering process seems to be effective in removing the foam component on the image whilst in other cases it does not perform adequately, causing a large number of false positive collembola identifications.

4 Conclusions

A system for automatic counting the number of collembola individuals on petri disks images was developed and presented here. The main challenges facing the

implementation of such a system were the different characteristics of the samples acquired, both in terms of illumination and in the presence of noise in the sample disk. The main problem with the illumination is the specular reflections that occur at the edge of the disk. These reflections are efficiently identified and removed by the proposed method by performing a two-stage segmentation. The presence of noise (foam in the sample) is harder to handle, as for extreme cases it can be difficult to discriminate between the foam and the collembolans themselves, even by a human operator. The method developed makes use of morphological image processing operators to reduce the amount of noise in the petri disk images.

The automatic counting system was tested with 38 images divided in three groups, according to their noise level. The average accuracy of the automatic counting system was 3.4% for good quality samples, 4.6% for medium quality and 7.5% for poor quality samples. The average relative error for the 38 images was 5.0%.

Plans for future work include the development of a better approach to deal with the samples highly contaminated by foam, or at least an automatic identification of the group that each sample belongs to. This could be used as a confidence value label to be paired with the actual counting figure produced. A more restrictive selection of the areas affected by the specular reflections could also be implemented. This could help reducing the counting error for those samples where the spatial distribution of the collembolans is not uniform, particularly at the edges of the sample.

References

1. Luz T.N., Ribeiro R., and Sousa, J.P.: Avoidance tests with collembola and earthworms as early screening tools for site-specific assessment of polluted soils. Environmental Toxicology and Chemistry. **23** (2004) 2188–2193
2. Bellinger, P.F., Christiansen, K.A., Janssens, F.: Checklist of the Collembola of the World. (1996-2006). http://www.collembola.org
3. Using Matlab, Version 6.5. The MathWorks, Inc. Natick. MA. (2002)
4. Gonzalez, R. C., Woods, R. E.:Digital Image Processing. Prentice Hall (2002)
5. Gonzalez, R. C., Woods, R. E., Eddins, S. L.:Digital Image Processing using MATLAB. Prentice Hall (2004)