DETECTION AND CLASSIFICATION OF PARASITE EGGS FOR USE IN HELMINTHIC THERAPY

Johan Musaeus Bruun[1], Christian M. O. Kapel[1], Jens Michael Carstensen[2]

[1] Faculty of Life Sciences, University of Copenhagen, Frederiksberg, Denmark
[2] DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark

ABSTRACT

Eggs from the small, intestinal pig whipworm *Trichuris suis* constitute the active pharmaceutical ingredient in a novel type of medicine for chronic autoimmune diseases like Crohn's disease. The pharmaceutical potency of such an egg suspension can be assessed by microscopic inspection, as only eggs containing a viable, infective larva provoke the wanted immune reaction. Thus, a precise and objective estimate of the concentration of infective eggs is crucial for dosing the new medicine.

In this paper, a vision-based method for detecting and classifying *T. suis* parasite eggs is described. The detection is based on matched filters and the classification is done using linear and quadratic discriminant analysis on a set of biologically inspired features, including the autocorrelation-based longitudinal anisotropy and the mean scattering intensity under dark field illumination. Despite the presence of impurities and overlapping eggs, the proposed method achieves cross-validated classification rates around 93%.

***Index Terms*—** elliptic filters, longitudinal anisotropy, mean scattering intensity, linear discriminant analysis, *Trichuris suis* eggs

1. INTRODUCTION

Helminthic therapy is a new category of treatments for autoimmune diseases, where the immunoregulatory effect of helminths – intestinal worms – is utilized to downregulate a hypersensitive human immune system [1]. New medicine against chronic autoimmune diseases of the intestines, namely Crohn's disease [2] and ulcerative colitis [3], is currently being developed and tested in humans in several European countries. The larvae inside tiny eggs from the helminth *Trichuris suis* function as the active pharmaceutical ingredient in the new medicine.

When producing the medicine, the concentration (eggs per ml) and potency (infective eggs per ml) of a given egg suspension need to be assessed in order to dose the medicine correctly. This is currently done manually by microscopic inspection of the eggs by a trained technician, but this procedure is subjective and complicated to validate. Instead, an objective, validatable, and automated vision-based procedure for assessing these critical quality parameters is sought developed.

To the authors' knowledge, no automated system has so far been developed for determining the developmental stages and thereby the infectivity of helminth eggs. Yang et al. [4] and others [5, 6] had some success in distinguishing different species of human helminth eggs on size and shape, but this approach is of no use for assessing infectivity. The segmentation part of the problem is somehow similar to a cell counting scenario [7, 8], although impurities are more frequent and vary more in size and shape in this case. Also, the importance of the internal details of the eggs prevents the use of a median filter for noise removal that [4, 5, 6, 7] use.

Trichuris suis eggs are oval, 60×25 µm, with a protruding, operculate plug in each end [9] as seen in Figure 1. This distinct, lemon-like shape makes them easily recognizable. At the time of analysis most of the eggs in the suspension have developed a viable larva and become *infective*, see the right egg in Figure 1. The other two eggs are examples of *non-infective* eggs who ceased their development at the initial stage and at an intermediate stage, respectively.

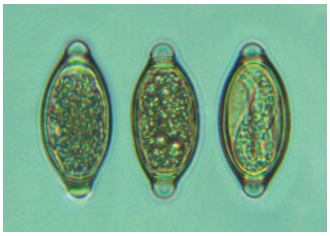


Fig. 1. Three *Trichuris suis* eggs. Two non-infective and one infective with a visible larva inside. Only the latter has the desired medical effect.

A set of 24 images – 12 bright field images and 12 corresponding dark field images – are analyzed. As seen in the example in Figure 2 most of the objects in the images are eggs, but some process-related impurities are present as well. The eggs have an average length of 162 pixels, corresponding to a resolution of 2.7 pixels per µm. Some of the eggs are covered by other eggs and some are standing upright. It is not

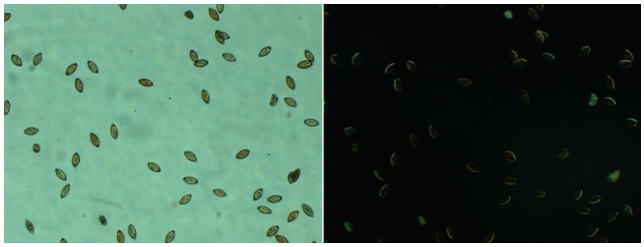


Fig. 2. Microscopic bright field and dark field images of *Trichuris suis* eggs, including a few overlapping eggs and some impurities. The eggs are lemon-shaped and approximately 60 µm long.

possible to determine visually if upright eggs are infective or not, so these are regarded as impurities in the analysis.

2. DETECTION

The analysis has been divided into three steps; detection, feature extraction, and classification. The first step is to detect the eggs in the image, i.e. determine the position and angle of each egg. Based on the fixed scale, proportions, and distinct shape of the eggs, a matched filter approach is chosen for this task. It has the advantage of being able to find eggs that touch each other without additional processing. An elliptic filter is constructed in order to match the oval shape of the eggs. By measuring the egg dimensions without the protruding plugs it is seen that for most eggs the width is approximately 56% of the length, corresponding to an elliptic eccentricity of 0.83. An elliptic filter with these proportions is therefore used for the detection, see Figure 3(a).

The bright field images are filtered at d angles in order to find eggs at all angles, and a pixelwise maximum image of the d filter response images is computed. It is dilated with a circular structuring element of radius s pixels, whereby local maxima grow to flat, circular plateaus, see Figure 3(d).

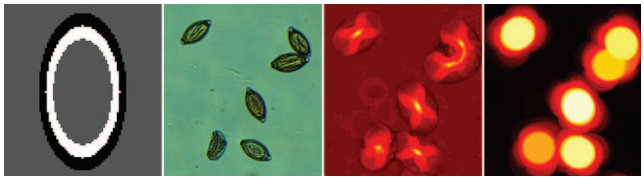


Fig. 3. (a) Elliptic filter. (b) Five eggs and one impurity. (c,d) Original and dilated version of the pixelwise maximum image of the filter responses. Note that the plateaus of the two touching eggs and the impurity are slightly darker, corresponding to a poorer fit of the filter.

The maximum response mage is then compared with the dilated image to identify the local maxima. The values of the local maxima express how well the elliptic filter match that location in the original image. An image threshold τ is applied such that all eggs in the training set are selected as good matches. This parameter seems to be rather robust given standardized and stable imaging conditions.

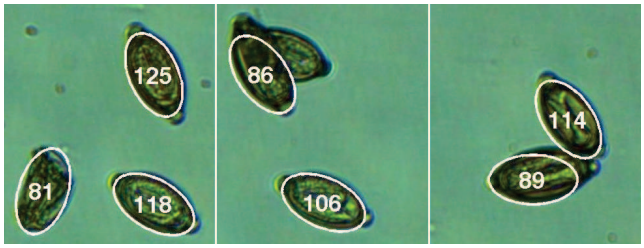


Fig. 4. Detected eggs, including an incorrectly detected impurity (81), an overlapping egg (86), and a suboptimal elliptic fit (89). The numbers on the eggs indicate their 'fit score' which is 1000 times their pixelwise maximum response value. The applied threshold of 0.077 consequently corresponds to a fit score threshold of 77.

2.1. Detection results

Filtering is done at $d = 30$ angles. The radius s depends on the egg size and is set to 85% of an egg width so that two eggs next to each other result in two plateaus while closer peaks will form a single plateau. This reduces the risk of overlapping detections. The maximum response threshold τ is set to 0.077 by histogram inspection, see also Figure 4.

The 12 image pairs in the data set contain 285 eggs in total. Out of these 282 were found in the detection step. The three unfound eggs were all partly overlapping with one or two other eggs. Additionally, 7 impurities were incorrectly detected as being eggs. Overall, the detection is very robust with a low number of false positives, resulting in a high precision rate of $282/289 = 97.6\%$, and a low number of false negatives, resulting in a high statistical power or sensitivity of $282/285 = 98.9\%$.

3. FEATURE EXTRACTION

The second step is to extract features from the detected eggs. Two biologically inspired features were chosen; longitudinal anisotropy and mean scattering intensity.

3.1. Longitudinal anisotropy

The longitudinal anisotropy, denoted λ, is an autocorrelation-based measure of the 'linearity' of the egg contents along the length of the egg versus across the egg. Technically, it is the ratio between the longitudinal and the transverse autocorrelation, ρ, of the egg content intensities.

$$\lambda = \frac{\rho_{long}}{\rho_{trans}} \tag{1}$$

The hypothesis is that eggs which contain fully grown larvae will have high longitudinal anisotropy values because of the way larvae position themselves inside the eggs in contrast to non-infective eggs with homogeneous contents. The autocorrelation is found by computing the correlation between the image itself (egg contents only) and a 1-pixel shifted copy, both longitudinally and transversely.

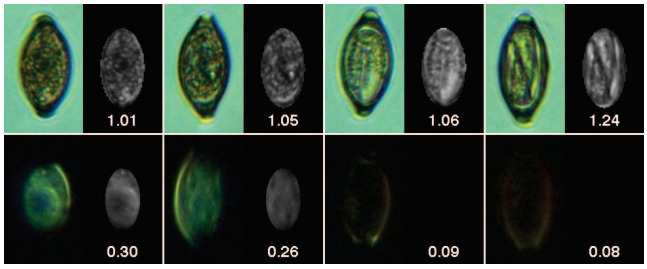


Fig. 5. (Top) Two non-infective and two infective eggs in brightfield and their isolated contents. Note that the third egg only has a longitudinal anisotropy of 1.06 even though it is infective. Its larval structures are not as prominent as the fourth egg's. (Bottom) Two non-infective and two infective eggs in darkfield. The differences in scattering are unfortunately not always this significant.

3.2. Mean scattering intensity

The second feature is the mean scattering intensity under dark field illumination. High amounts of scattering has been observed in several of the non-infective eggs under dark field illumination, resulting in a bright, internal glow in the central region of the eggs which resembles a small white cloud, c.f. Figure 6. The mean scattering intensity, denoted μ_s, is computed as the average intensity of the egg contents, extracted by a mask and converted to grayscale. Four examples are shown in Figure 5.

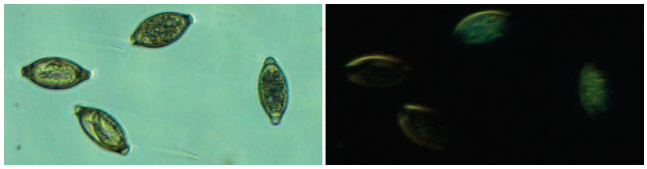


Fig. 6. Light scattering in two infective eggs (lower left) and two non-infective eggs (upper right).

4. CLASSIFICATION

The third and last step is to classify the detected eggs as either infective or non-infective based on the extracted features. This is a binary classification problem with two continuous predictor variables; longitudinal anisotropy, λ, and mean scattering intensity, μ_s. All 12 image pairs in the data set have been manually annotated by marking and categorizing all eggs as either infective or non-infective. These estimates were performed by an experienced technician who is used to assess the egg potency by microscopic inspection.

		Classification outcome	
		Infective	**Non-inf.**
Truth	**Inf.**	248 True pos.	1 False neg.
	Non-inf.	13 False pos.	20 True neg.

Table 1. Confusion matrix for the classification.

4.1. Classification methods

The classification is done using linear and quadratic discriminant analysis. Empirical prior probabilities are used since the images in the data set are assumed to be representative with regards to the amounts of infective and non-infective eggs. Besides the number of actual misclassifications in the data set and the corresponding resubstitution error rate, the leave-one-out and 10-fold cross-validation error rates are computed in order to estimate the classifier's ability to generalize to similar but unseen data. Since the 10-fold cross-validation is depending on a random division into subsets, its performance is computed as an average of 10 repetitions.

4.2. Classification results

The 282 actual eggs among the 285 detected egg candidates are used in the classification. Of these 282 eggs, 249 are infective, and 33 are non-infective so the empirical priors are $[0.883, 0.117]$, which corresponds well with the expected potency of the egg solution.

Figure 7 shows a plot of the longitudinal anisotropy vs. the mean scattering intensity as well as the separating lines of a linear discriminant analysis (LDA) and a quadratic discriminant analysis (QDA) assuming empirical priors. LDA and QDA both misclassifies 14 eggs, corresponding to a resubstitution error rate of 4.96% which turns out also to be their leave-one-out error rates. The cross-validation error rate of LDA is 5.04% and of QDA is 5.11%, but this difference is within the uncertainty. The confusion matrix for the classification using LDA can be seen in Table 1.

5. DISCUSSION

The presented analysis was implemented in MATLAB. The parameters in the analysis, including the thresholds, are em-

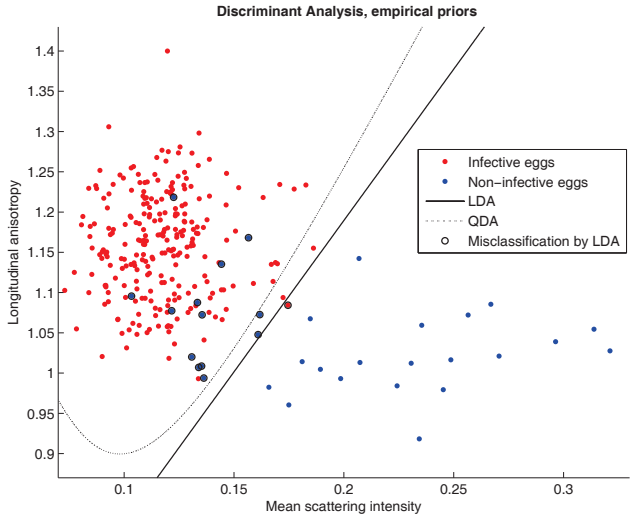
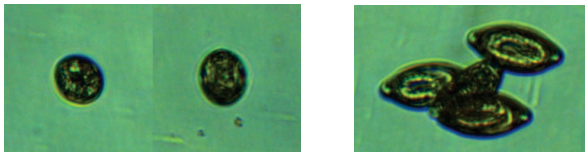


Fig. 7. Linear and quadratic discriminant analysis for classifying the infectivity of *Trichuris suis* eggs based on their longitudinal anisotropy and mean scattering intensity.

pirically assessed by inspecting relevant distribution plots and histograms. An automated, supervised tuning process, possibly combined with additional image data, could therefore lead to better parameter estimates.

It is not possible to assess the infectivity of upright standing eggs visually, see Figure 8(a), so process improvements like sample shaking should be applied to avoid this problem. One source of error is that not all overlapping eggs are found in this analysis, see Figure 8(b). It would be possible to target these directly, for instance by locating local 'fit score' maxima close to each other or using Delaunay triangulation like [7].

As seen in Figure 4, the elliptic fits from the initial detection step are not perfect. This of course influences the computed features, primarily the longitudinal anisotropy. So by improving these fits, e.g. by post-optimizing the ellipse parameters or by fitting an active shape model, more accurate feature scores could be obtained, and some of the falsely detected impurities could potentially be discarded.



(a) (b)

Fig. 8. Examples of upright standing and overlapping eggs.

A natural extension of the presented analysis would be to locate and analyze all impurities in the suspension as well. This could be used to assert that the suspension does not contain any undesirable foreign particles or impurities such as other parasite eggs, which is an essential quality parameter in the final release of the pharmaceutical raw suspension.

6. CONCLUSION

As seen in section 2.1, the detection is very robust. 282 of the 285 actual eggs (98.9%) were detected, while only 7 impurities were incorrectly detected as being eggs. Of these 282 actual eggs, 14 were misclassified wrt. infectivity by the LDA.

So when regarding the 7 impurities as misclassifications as well, the total classification success rate is $268/289 = 92.7\%$. This shows that a fast and non-invasive vision-based assessment of the pharmaceutical potency of *Trichuris suis* eggs is feasible using the presented analysis, although additional infectivity-related features should be investigated in order to increase the classification accuracy.

7. REFERENCES

[1] A. Reddy and B. Fried, "The use of trichuris suis and other helminth therapies to treat crohns disease," *Parasitology Research*, vol. 100, pp. 921–927, 2007, 10.1007/s00436-006-0416-4.

[2] R.W. Summers, D.E. Elliott, J.F. Urban, R.A. Thompson, and J.V. Weinstock, "Trichuris suis therapy in crohns disease," *Gut*, vol. 54, no. 1, pp. 87–90, 2005.

[3] R.W. Summers, D.E. Elliott, J.F. Urban, R.A. Thompson, and J.V. Weinstock, "Trichuris suis therapy for active ulcerative colitis: A randomized controlled trial," *Gastroenterology*, vol. 128, no. 4, pp. 825 – 832, 2005.

[4] Y.S. Yang, D.K. Park, H.C. Kim, M. Choi, and J. Chai, "Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network," *Biomedical Engineering, IEEE Transactions on*, vol. 48, no. 6, pp. 718 –730, june 2001.

[5] D. Avci and A. Varol, "An expert diagnosis system for classification of human parasite eggs based on multi-class svm," *Expert Systems with Applications*, vol. 36, no. 1, pp. 43 – 48, 2009.

[6] E. Dogantekin, M. Yilmaz, A. Dogantekin, E. Avci, and A. Sengur, "A robust technique based on invariant moments anfis for recognition of human parasite eggs in microscopic images," *Expert Systems with Applications*, vol. 35, no. 3, pp. 728 – 738, 2008.

[7] H. Berge, D. Taylor, S. Krishnan, and T.S. Douglas, "Improved red blood cell counting in thin blood smears," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, 30 2011-april 2 2011, pp. 204 –207.

[8] S. Kothari, Q. Chaudry, and M.D. Wang, "Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*, 28 2009-july 1 2009, pp. 795 –798.

[9] J.S. Pittman, G. Shepherd, and B.J. Thacker, "Trichuris suis in finishing pigs: Case report and review," *Journal of Swine Health and Production*, vol. 18, no. 6, 2010.