Novel Method for Parasite Detection in Microscopic Samples

Patryk Najgebauer, Tomasz Nowak, Jakub Romanowski, Janusz Rygał, Marcin Korytkowski, and Rafał Scherer

Department of Computer Engineering, Częstochowa University of Technology
al. Armii Krajowej 36, 42-200 Częstochowa, Poland
{patryk.najgebauer,tomasz.nowak,jakub.romanowski,janusz.rygal, marcin.korytkowski,rafal.scherer}@kik.pcz.pl
http://kik.pcz.pl

Abstract. This paper describes a novel image retrieval method for parasite detection based on the analysis of digital images captured by the camera from a microscope. In our approach we use several image processing methods to find known parasite shapes. At first, we use an edge detection method with edge representation by vectors. The next step consists in clustering edges fragments by their normal vectors and positions. Then grouped edges fragments are used to perform elliptical or circular shapes fitting as they resemble most parasite forms. This approach is invariant from rotation of parasites eggs or the analyzed sample. It is also invariant to scale of digital images and it is robust to overlapping shapes of parasites eggs thanks to the ability to reconstructing elliptical or other symmetric shapes that represent the eggs of parasites. With this solution we can also reconstruct incomplete shape of parasite egg which can be visible only in some part of the retrieved image.

Keywords: edge detection, shape recognition, objects detection, image processing.

1 Introduction

Content-Based Image Retrieval (CBIR) is a branch of science based on advanced mathematical and computer science methods for analyzing digital images. In truth, the process of computer image analysis is still far from the way in which images are analyzed by the human brain. Nevertheless, existing methods and algorithms allow the use of CBIR-related technologies in many areas of daily life or work. One of the areas in which CBIR gains the immense popularity is analysis of medical imaging (e.g. computed tomography, USG or RTG). The current literature studies do not show, however, real-world solutions in the field of image analysis in parasitological domain. Parasitology is the study of parasites and parasitism, which deals with morphology, anatomy and development of parasites. One of the primary examinations used in parasitology is taking samples and test them under a microscope to identify a parasite on the basis of its physical

L. Rutkowski et al. (Eds.): ICAISC 2012, Part I, LNCS 7267, pp. 551–558, 2012.

characteristics. Each parasite has individual characteristics, which identify its species. The examination itself is time consuming, even for an experienced doctor or lab technician; there are also frequent errors in determining parasite species, resulting in inadequate treatment that can lead to rapid health deterioration and even death of patient.

For many years content-based image retrieval systems (CBIR)[6] have been used in many different areas, such as medicine, biometrics or automotive. Major research areas combining image processing and medical analysis are computed tomography, magnetic resonance and ultrasound scan [8][1][7][2]. Currently, these systems support the work of medical doctors in many different fields of medicine. However, there is no specific application dedicated to parasitology, i.e. to detect and identify parasites in raw laboratory samples. Certain works focused on building knowledge bases, which contained images of selected parasites [10] and were used primarily for academic purposes. In most samples there exists different pollution that can obscure parasites. In many samples there are also structures that can resemble parasites. There are many algorithms for processing images [11][14][4] used in medicine, which specialize in finding specific objects (e.g. bacteria, tumors, fractures, different unwanted changes). An example may be algorithms, which allow searching for strains of malaria in blood [12], or specific algorithms analyzing organs in tomography [13]. General methods used in image processing are not suitable for search of parasites. They can only locate the object (provided that the sample was previously cleaned), but are not able to identify the parasite species. In the case where the sample is contaminated with other structures (such as hair or grass), these structures will be also treated as important objects.

The proposed method can detect edges of objects, which could be parasites, and can supplement the missing parts of the detected edges. In samples obtained directly from the parasite tests there can exist many different objects such as air bubbles, blood, grass, fur, and other non-parasitic objects. Objects can overlap, which may hinder their detection by existing algorithms. A helpful feature here is the shape of the parasite eggs as they are characterized by elliptical construction and have distinctive shell. These features can be a starting point to detect the edges of the object. Often in the samples only a part of the essential object is visible and the next important stage of this method is to develop an algorithm that will be able to predict (reconstruct) the missing elements of the parasites. This is possible because the eggs of parasites are characterized by symmetrical oval shapes, and have similar size [9]. Methods that were presented in the literature are very general and do not provide a solution to this problem. The next problem will be converting the detected edges of objects and their classification.

The remainder of this paper is organized as follows. Section 2 describes the problem of the edge extraction in digital images and the current CBIR methods used in parasitology. Section 3 describes the proposed method of the edge recognition. Section 4 reports the experimental results on a set of real images.

2 Problem Description and Existing Methods

Problems with computer-aided parasites detection start at the edge detection step. Most edges of parasite eggs are blurred. Edges blur is mainly caused by focus loss of the microscope and reflections of the backlight. The second problem is transparent egg bodies, because some pollution are visible in the background. The standard Canny edge algorithm cannot detect blurred edges; this problem can be resolved by using various size of edge detection masks.

After edge detection, we have to deal with a problem with vague egg edges and edges noise from objects texture and pollution. Noise edges from textures and small pollution can be eliminated by curvature test. Vague or invisible edges are caused by close extensive pollutions and other parasite eggs. The extensive pollution can be e.g. crystals, air bubbles or undigested debris. Problems with these pollutions are hard to solve because they exist in different sizes and shapes. Moreover, some of their shapes are similar to parasite eggs. Second problem occurring with the lack of pronounced edges is that edges are sometimes connected with other nearly located objects what increases problem for recognition of objects. To overcome this problem we present a method that transforms edges pixels to vectors to speed up filtering and grouping process.

In the last step of parasites egg detection the main problem is the comparison of their shape to ellipse. The problem outcomes from the fact that the shapes of parasite eggs are not accurately symmetrical. Symmetries and ellipses matched to the shape can be determined only approximately.

There are many algorithms for processing images [11] [14] [4] used in medicine, which specialize in finding specific objects (bacteria, tumors, fractures, different unwanted changes). These specific as well as general methods used in image processing are not suitable for search of parasites. They can only locate objects (provided that the sample was previously cleaned), but are not able to identify the parasite species. In the case when the sample is contaminated with other structures (e.g. hair, grass, crystals, air bubbles), existing algorithms will also classify them as important objects.

The literature shows that previous work focused on building knowledge bases containing images of selected parasites (see e.g. [10]), which are used primarily for academic purposes. None of the previous work was dedicated to an attempt to detect and identify parasites in raw laboratory samples. In most samples there exist various pollutions and in result parasites can be obscured. In many samples there are also structures that can resemble parasites. Existing databases [9] [5] mainly contain parasites isolated from the samples. Moreover they are prepared on various hardware, in different scales and using various methods. These databases are not suitable to for the development of CBIR algorithms for parasite identification.

3 Novel Method for Parasite Egg Detection

Proposed method is designed to detect and to enhance edges of parasite eggs or cyst and finally to estimate the best fit ellipse to their shape. The proposed method detects parasites in following steps.

3.1 Multi-scale Edge Detection

The presented solution is based on the Canny edge detection algorithm [3]. First, the median filter is used to remove noise and small reflections (backlight from microscope), that appear on the test material. The proposed method uses the Sobel filter mask in X and Y axis direction. Frequent blurring on the edge of the examined objects caused by inaccurate focus calibration of microscope forced detection of edges at several scales (Fig. 1). We used 8x8 (Fig. 1b) and larger, 16x16 mask sizes (Fig. 1c). Such action also allowed skipping the part of small impurities and isolating the edges of larger objects from each other. The last

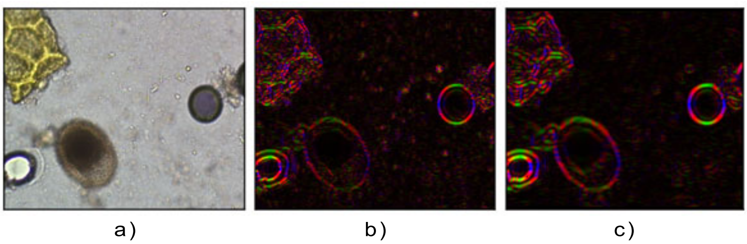

Fig. 1. Multi scale edge detection. (a) Input image. (b) Sobel filter with mask size 8x8 pixels. (c) Sobel filter with mask size 16x16 pixels.

step of the process of edge detection in the presented solution is to search for local extremes among the determined edges. In the result we obtain enhanced edges and steeper gradients of edges. The outcome of this stage is an array of local gradient maximum values in the X and Y direction for each pixel.

3.2 Vector-Based Edge Representation

To speed up the method and to allow for additional filtration of image, edges are converted to a vector form as presented in Fig 2. With this action the amount of data to be analyzed is dramatically reduced. Each part of the edge has information about its location and position of his arms. For this purpose, the image with detected edges is analyzed by means of a grid at a minimum of 10 pixels in x and y axes. This solution speeds up the search for egg edges and ignores the edges of small, unimportant parts of the background. The edges of larger objects will be sliced by the grid at multiple points to provide an outline of the object. For each detected by the grid of edge pixel neighboring pixels are examined.

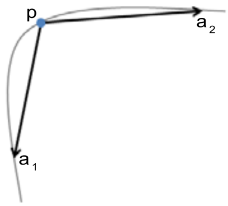

Fig. 2. Edge vector based representation. p - position of edges. a_1, a_2 - vectors of edge arm

The algorithm follows the edge of the pixel tracking in two opposite directions relative to the axis in which the component value of the gradient was the highest. After the passing of the distance in pixels equal to the number of pixels between grid lines are designated as vectors for the edges of each arm from the starting point. An example of vector representation of image edges is shown in Fig 3a.

3.3 Edge Filtration

Filtering of the detected edges is possible thanks to the determined arm vectors from the previous stage. At the beginning we remove the edges where the angle

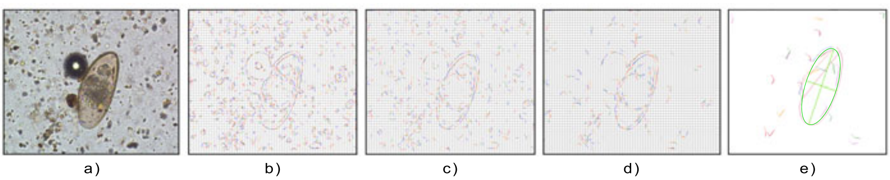

Fig. 3. Vector-based edge representation with filtering and grouping. (a) Input image. (b) Image of vector-based representation of detected edges . (c) Edges filtered by angle between arm vectors. (d) Edge filtered by relation between neighbor edges.

between their arms is less than 140 °. For that purpose, we determine cosine of the angle (dot product) between their normal vectors

$$A = \overrightarrow{an} \cdot \overrightarrow{bn} \tag{1}$$

where $\overrightarrow{an}$, $\overrightarrow{bn}$ are normal vectors of edge arms. In this way, we remove part of the detected edges which represent object texture or small pollution (see Fig. 3b). The second step of filtering is removing isolated edges, i.e. without near placed edges of the same direction, forming continuous line (Fig. 3c). The degree of linkage between the lines is determined by the following formula

$$B = |\overrightarrow{n} \cdot \overrightarrow{abn_1}| \cdot |\overrightarrow{n} \cdot \overrightarrow{abn_2}| \tag{2}$$

where: $\overrightarrow{n}$ specifies the normal vector between positions of the edge fragments, and $\overrightarrow{abn_1}$ normal vector between the positions of the first edge arms and $\overrightarrow{abn_2}$ normal vector between the positions of second edge arms. Edges of the value of the coefficient B above values of 0.7 are treated as related edges. All the edges that are not related with at least one edge are removed from the list of edges.

3.4 Edge Grouping and Ellipse Approximation

In this step of the proposed method we group related edges in similar way to filtering by the coefficient B from Section 3.3. However, one edge can be joined only with two related edges, one per arm. Additionally, the method rejects groups composed of less than three edges. Now, only a few groups of edges remain from the entire image. At the final step, the algorithm finds ellipses which are best fit to edges groups by last square estimation.

4 Experiments

In the experiments we use images directly from the clinical parasitological laboratory at a resolution of 720x576 pixels. Images were captured directly from the camera which is built into the microscope. Presented method was implemented in C# language. The first step of analysis was to use median filter with mask

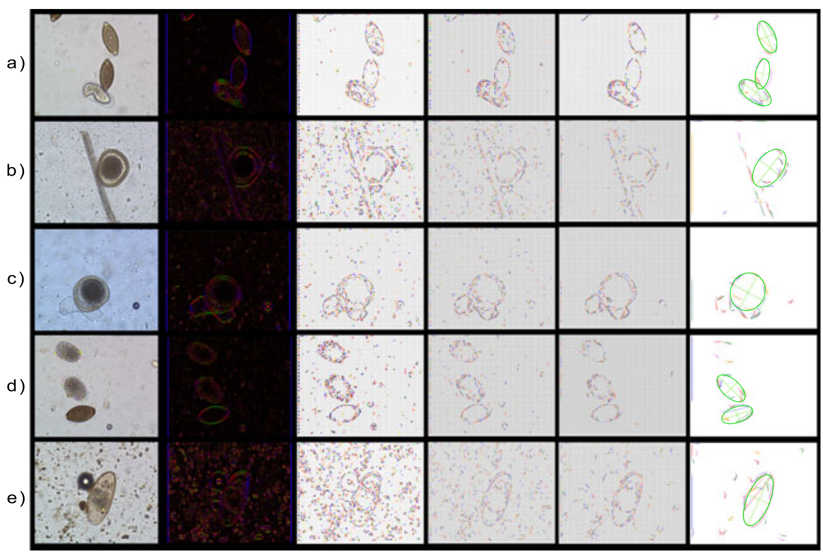

Fig. 4. Experimental results for parasite edge detection on 5 real images. First column, original images; second column, Canny edge detection with Sobel 8x8 filter; third column, created vectors from edges pixels; fourth column, the rejection of sharp edges; fifth column, rejection of the edges that do not form a continuity with the neighbor; sixth column, edge grouping by the proposed method.

of 4x4 pixels to denoise image. The next stage is the process of edge detection which uses Canny edge algorithm with Sobel mask of size 4x4 and 8x8 pixels. Figure 3 shows the step-by-step algorithm. The next step was to replace the detected edges by vector based representation - this operation is described in detail in section 3.2. The resulting vectors are filtered and grouped by their normal vectors (Section 3.3). The last step in the proposed solution is the detection of ellipses based on the created groups of the edges, using the least squares method (chapter 3.4). Fig. 4 shows 5 real images taken in different focus which were used during the experiments. As shown in the presented experimental examples the goal of the method is to find the edges of parasites and to remove all edges that belong to the impurities.

5 Conclusions

Existing algorithms used in image analysis of are not suitable for the analysis of images from parasitological examinations in a way that would enable the detection and identification of the parasite. After consultation with the parasitological clinical laboratory and preliminary research, it can be assumed that the creation of new algorithms targeted for parasitological examination will enable more profound research of such images. The result of the algorithms should be to locate and identify parasites. Characteristics that allow for proper classification of objects similar to parasite eggs and cysts are elliptical shapes with smooth edges.

Acknowledgements. The authors would like to thank Dr Andrzej Połozowski from Wrocław University of Environmental and Life Sciences for the opportunity to capture microscopic parasite video and inspiring talks.

The project was funded by the National Center for Science under decision number DEC-2011/01/D/ST6/06957.

References

1. Akgül, C.B., Rubin, D.L., Napel, S., Beaulieu, C.F., Greenspan, H., Acar, B.: Content-based image retrieval in radiology: current status and future directions. J. Digit. Imaging 2, 208–222 (2011)
2. Antonie, M.L., Zaiane, O.R., Coman, A.: Application of Data Mining Techniques for Medical Image Classification. In: Proceedings of the Second International Workshop on Multimedia Data Mining (2001)
3. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 8(6), 679–698 (1986)
4. Chuctaya, H., Portugal, C., Beltran, C., Gutierrez, J., Lopez, C., Tupac, Y.: M-CBIR: A medical content-based image retrieval system using metric datastructures. JCC (2011)
5. Gardner, S.L.: The parasite collection search page in the Manter Laboratory of Prasitology, `http://manter.unl.edu/hwml`

6. Gudivada, V.N., Raghavan, V.V.: Content based image retrieval systems, vol. 28(9), pp. 18–22. IEEE Computer Society (1995)
7. Gueld, M.O., Keysers, D., Deselaers, T., Leisten, M., Schubert, H., Ney, H., Lehmann, T.M.: Comparison of global features for categorization of medical images. Medical Imaging (2004)
8. Hsua, W., Antani, S., Long, L.R., Neve, L., Thoma, G.R.: SPIRS: a Web-based image retrieval system for large biomedical databases. Int. J. Med. Inform. 78, 13–24 (2009)
9. Mallik, J., Samal, A.K., Gardner, S.L.: A Content Based Pattern Analysis System for a Biological Specimen Collection. In: 7th IEEE International Conference on Data Mining (2007)
10. Meduri, R., Samal, A., Gardner, S.L.: Worm-Web Search: A Content-Based Image Retrieval (CBIR) System for the Parasite Image Collection in the Harold W. Manter Laboratory of Parasitology, University of Nebraska State Mueum (2008)
11. Muller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A Review of Content - Based Image Retrieval Systems in Medical Applications - Clinical Benefits and Future Directions. Int. J. Med. Inform. (2009)
12. Tek, F.B., Dempster, A.G., Kale, I.: Malaria Parasite Detection in Peripheral Blood Images. In: British Machine Vision Conference (2006)
13. Urschler, M., Mayer, H., Bolter, R., Leberl, F.: The LiveWire Approach for the Segmentation of Let Ventricle Electron-Beam CT Images. OCG 160, 319–326 (2002)
14. Wanjale, K., Borawake, T., Chaudhari, S.: Content Based Image Retrieval for Medical Images Techniques and Storage Methods-Review Paper. IJCA Journal 1(19) (2010)