

A Real Time Animal Detection And Segmentation Algorithm For IRT Images In Indoor Environments

J. FLEURET^{1,2}, V. OUELLET³, L. MOURA-ROCHA^{3,4}, E. CHARBONNEAU³, L. SAUCIER³, L. FAUCITANO^{3,4},
X. MALDAGUE^{1,2}

¹Faculty of Sciences and Engineering, University Laval, Quebec City, Quebec, Canada,
julien.fleuret.1@univ.laval.ca

²Laboratory of Computer Vision and Numerical Systems, Multipolar Infrared Vision team, Quebec City,
Quebec, Canada

³Faculty of Agriculture and Food Sciences, University Laval, Quebec City, Quebec, Canada

⁴Canada AgriFood, Sherbrooke Research and Development Centre, Sherbrooke, Quebec, Canada

Abstract

The segmentation of objects of interest in an image is one of the first processing steps in numerous applications of image processing. A variety of algorithms and applications of segmentation are described in the literature, but very few have been applied to thermal infrared images, and even less in the context of livestock farming. Most of the literature approaches use either approaches based on optimisation algorithms which cannot be used for real time applications, or linear thresholds which often induce the presence of many artifacts in the resulting image. In this paper we propose a new algorithm to achieve the segmentation of animals in indoor environment images.

1 Introduction

In many applications of computer vision image, segmentation represents the first step in processing. For this reason there are numerous methods in the literature some for general purposes and some for specific tasks. The aim of a segmentation algorithm is to trim the data in order to only keep the information of interest. Recently the trend for a higher automatization of livestock and farming procedures has increased. The identification of the animal within the environment is therefore challenging. The literature proposes mainly three approaches. Wirthgen [9, 10] proposes an approach based on an Active Shape Models (A.S.M.). This approach gives good results but needs priors such as training on manually segmented landmarks. The A.S.M. are based on the optimization of a cost function, that makes them very computational and often not applicable for realtime applications. Later Wirthgen also proposed another approach [8] to refine the results of the A.S.M., which is beyond the scope of this paper. In his approach Christiansen et al. [2] proposed a linear segmentation criterion. This criterion can be used for real time applications, however it is sensitive to thermal noise in the image. Such an approach can be suitable in environments with no or little noise such as constraint environments or some outdoor applications but cannot be used for indoor applications due to the presence of noise sources such as lights, electrical wires and devices. Arruda et al. [1] used a data driven approach based on the statistical distribution of the data in the image. These approaches are often based on algorithms like K-means which require the knowledge of priors such as the number of clusters. If the objects have different statistical representations in the image such approaches can give a precise result and are not very computationally demanding. However in indoor environments most of the time the noisy elements have a statistical distribution close to those of the object of interest which makes them harder to detect even by these methods. Like Christiansen's, Arruda's approaches can be suitable for some outdoor applications or in constraint environments but cannot be used for most indoor applications.

In this paper we propose a method for the segmentation of an image into three regions : background, foreground (the region(s) of interest), and false detection. In this approach the set up of two parameters may need to be changed depending on the application. The experiments showed that this application satisfies the real time constraint. A refinement of the data is also introduced in order to clean the outputs, but this method does not satisfy the real time constraint.

The rest of this paper is organized as follows. In section 2 a review of two automatic thresholding methods is made. Section 3 briefly introduces the four basic morphological operators. The different steps of our method are detailed in section 4. Some experiments are presented, analyzed and discussed in section 5. Then we conclude the paper.

2 Automatic Thresholding

One of the most famous automatic threshold algorithms was introduced by Otsu [6]. The threshold criterion is determined by processing the histogram.

Otsu's Method

For a given grey level image the total number of pixels can be calculated as : $N = \sum_{i=0}^{L-1} n_i$ where L represents the number of range values and n_i the number of pixels with an intensity of i . From the previous equation we can easily deduce : $p_i = \frac{n_i}{N}$ and $\sum_{i=0}^{L-1} p_i = 1$ for $p_i \geq 0$.

For a binary thresholding the pixels are divided into two classes C_1 and C_2 with the grey levels : $[1, \dots, t]$ and $[t + 1, \dots, L]$. The probability distribution for each class is then processed as follows :

$$\begin{aligned} C_1 &: \frac{p_1}{\omega_1(t)}, \dots, \frac{p_t}{\omega_1(t)} \\ C_2 &: \frac{p_{t+1}}{\omega_2(t)}, \dots, \frac{p_L}{\omega_2(t)} \end{aligned} \quad (1)$$

$$\begin{aligned} \omega_1(t) &: \sum_{i=1}^t p_i \\ \omega_2(t) &: \sum_{i=t+1}^L p_i \end{aligned} \quad (2)$$

The means are :

$$\begin{aligned} \mu_1(t) &: \sum_{i=1}^t i \times \frac{p_i}{\omega_1(t)} \\ \mu_2(t) &: \sum_{i=t+1}^L i \times \frac{p_i}{\omega_2(t)} \end{aligned} \quad (3)$$

The threshold value is then determined as the value that maximizes the intra class variance :

$$t^* = \arg \max_{1 \leq t < L} \omega_1(\mu_1 - \mu_T)^2 + \omega_2(\mu_2 - \mu_T)^2 \quad (4)$$

with : $\mu_T = \sum_{i=0}^{L-1} i \times p_i$

A few years after Otsu's method, Zack et al. [12] proposed another method to automatically determine the threshold of an image by processing its histogram. Zack's approach is well known as the triangle method.

The Triangle Method

This method has two main steps :

1. A line is computed between the maximum B_{max} and minimum B_{min} bins of the histogram.
2. The threshold is then determined as the position which maximizes the distance between the line and the number of bins [7, 11]

3 Mathematical Morphology

Mathematical morphology is an approach that is widely used in image processing due to its ability to work with shapes. Among the morphological operators two of the most basic operators include erosion and dilatation. These two operators can be associated to create two other operators named Opening, and Closing. These

operators work with a structuring element that is a shape used to interact with the neighbourhood of a given pixel.

The erosion of a grey level image I by a structuring element S can be defined as follows [5] :

$$I \ominus S(m, n) = \min(F(m + s, n + t) - B(s, t)) \quad (5)$$

The intensity values of the resulting image of the erosion always have values lower or equal to the value of the source image. Due to this property the erosion operator is called a "shrinking" operator.

The dilatation of a grey level image I by a structuring element S can be defined as :

$$I \oplus S(m, n) = \max(F(m + s, n + t) - B(s, t)) \quad (6)$$

The intensity values of the resulting image of the dilation always have values greater or equal to the value of the source image. The dilatation operator is called an "expansion" operator for this reason.

The Opening and Closing operators are defined respectively as :

$$\begin{aligned} I \circ S &= (I \ominus S) \oplus S \\ I \bullet S &= (I \oplus S) \ominus S \end{aligned} \quad (7)$$

The resulting image of an opening operation is smoother and the regions connected with a few pixels linked in the source are separate. The closing operation fills a few pixel holes and connects regions separated only by a few pixels.

4 Method

As pointed out by Wirthgen in [8] automatic diagnostics requires a robust and precise segmentation. The aim of this algorithm is to return an estimation of the foreground i.e. an estimation of the region(s) of interest, the background, and the false detection.

The term false detection refers to some elements that are part of the background but hotter than average background elements such as electrical wire or lights.

Estimation of the Regions of Interest

The first step of this algorithm consists in using Otsu's method to threshold the source image. This step enables only the higher intensities present in the image to be maintained which already provides a reasonably good estimation concerning the background of the image. The thresholded image still contains undesirable elements of the scene such as the lights, electrical wires and others. In order to estimate the foreground we use a coarse to fine approach.

Estimation of the Background and Foreground

Coarse step : In order to obtain a coarse estimation of the foreground we apply several morphological operations. The structuring element used in this method is a cross in order to preserve as much as possible the contours of the regions. Structuring elements such as rectangles can slightly enlarge the surface of the regions. A larger region means more pixels from the background can be estimated as part of the foreground. A morphological opening is applied on the thresholded image followed by several (N) dilations. The variable N can be tuned according to the application.

These operations remove the small elements such as the electrical wires, and significantly reduce the size of the larger one. After these operations a coarse estimation of the background and foreground is made by subtracting the resulting image from the source. These first estimations still contain many artifacts that need additional processing in order to refine them.

Fine step : The refinement information is obtained by processing the background estimation. First a binary image of all the non-zero pixels is created and eroded, then every region lower than the surface S is removed. The variable S can also be fine tuned depending on the application. The erosion of the binary image creates a fusion of closed regions. This is important so as not to erase closed regions with a number of pixels lower than S e.g. the legs of an animal a little bit farther than the animal of interest.

The hole filling operation consists in detecting the contours of the objects present in the image and

then filling them with the values of the background image.

From the result of the hole filling the background and foreground are updated. Any pixel with a non-zero in both the previous estimation of the foreground and the hole filling result is a part of the background. The foreground is also updated as the subtraction of the background from the source. This processing creates an estimation of the background such that anything that is not in the foreground is in the background. Thus the electrical wires are considered as part of the background even if they are warmer than the average elements of the background e.g. electrical wires are always hotter than the walls or the top of the building. Such elements cannot be considered as part of the foreground either, even if their intensity is quite close to those of the foreground elements, but they can also not fit with the background, because their intensity is significantly higher than the average of the background. This illustrates the need to identify these elements and create a false detection estimation map.

False detection estimation

The first step of the false detection estimation consists in normalizing the input image in the range of an 8 bits pixel image. To complete this step we apply the formula :

$$I_8 = \frac{I_{16} - \min_{I_{16}}}{\max_{I_{16}} - \min_{I_{16}}} \times 255 \quad (8)$$

Where $\min_{I_{16}}$ and $\max_{I_{16}}$ are respectively the minimum and maximum values of the input image. This image is then thresholded. The threshold value is determined using the triangle method [12].

The interest in using the triangle method rather than Otsu's method stems from the fact that the threshold value is generally lower than the value obtained with Otsu's method for the same image. That means that the thresholded image will contain more information.

An update of the background estimation and the creation of the foreground estimation is made using the binary image. Any non-zero pixel of the binary image which is also a non-zero pixel in the background is a false detection pixel.

For some applications it might be convenient to know which pixel has been estimated as a part of either the background, the foreground or the false detection. To satisfy this need a colorized map is generated. The color association has been arbitrarily chosen. This map uses a color code where blue refers to foreground, green is the background and red refers to false detection.

The given result obtained by the previous steps may contain some misclassified pixels i.e. some isolated pixel or groups of few pixels considered as a part of the background. For offline applications or post processing, the next section provides a refinement approach of the previous estimations.

Refinement

The refinement is based on the method described by Ding et al. [3]. From the previous steps we only need to update the colorized map. To update the colorized map, the threshold value used in [3] is reused but almost any isochromatic pixel value ($R = G = B$) can be used. When the colorized map has been updated then the estimations of the background, foreground, and false detection are updated using this information.

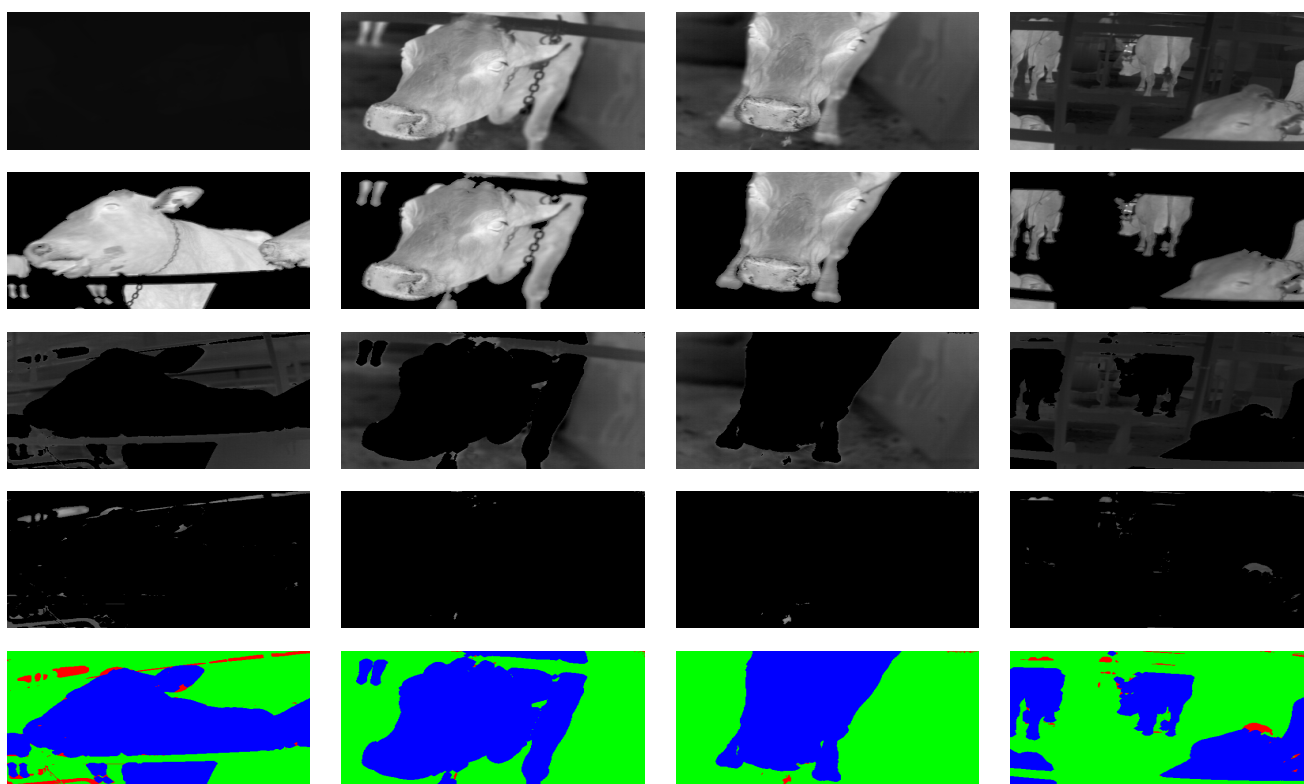


FIGURE 1 – Visualization of the outputs. Row by row from top to bottom. Row 1 : the source image. Row 2 : estimation of the foreground. Row 3 : estimation of the background. Row 4 : estimation of the false detection elements of the scene. Row 5 : map

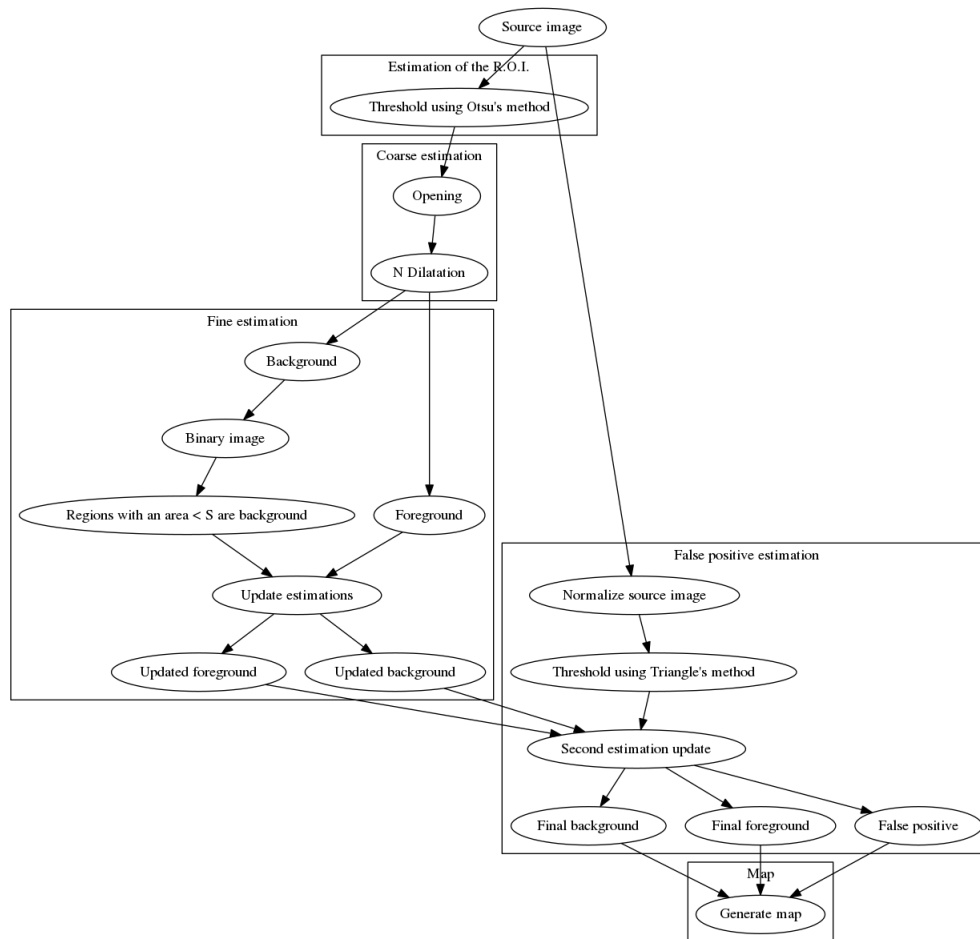


FIGURE 2 – Description of the main steps of the method

Experiments and Analysis

This section shows the performance of our approach.

During these experiments the variable N has been set to 3 while the variable S was automatically determined as : $S = (0.1 \times width) \times (0.1 \times height)$ with $width$ and $height$ the width and height of the source image.

Metrics

To evaluate the quality of results we used several metrics. The Receiver Operating Characteristic (R.O.C.) approach has been used since the second world war to evaluate detection. In order to obtain a quantitative result of the R.O.C. approach we used the F1 score which represents the harmonic mean of the true positive rate and the positive predictive value. The true positive rate represents the ratio between the number of pixels segmented from the result image with the number of pixels segmented from the ground truth (annotated) image. The positive predictive value is the ratio of the number of pixels segmented from the result with the sum of the pixels segmented in both the result image and the ground truth image. To calculate the F1 score we also need to determine the number of true positive, false positive and false negative pixels. The true positive pixels are the pixels segmented in both the result image and the ground truth, the false positive pixels are the pixels segmented in the result image but not in the ground truth and the false negative pixels are the pixels not segmented in the result image but segmented in the ground truth. Figure 3 reuses the illustration of [4]

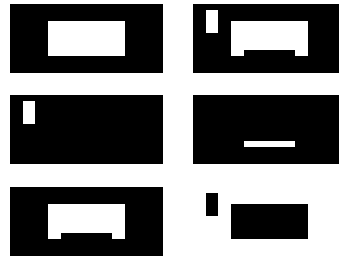


FIGURE 3 – The following description is provided from left to right. In the first row we see the Ground Truth, and a segmented image, in the second row the False Positive region and the False Negative region, in the last row the True Positive and True Negative regions.

Then the F1 is processed as :

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (9)$$

We also used to Jaccard Similarity (J.S.) to describe the union of regions between two images.

$$J.S._{I_1, I_2} = \frac{\sum I_1 \times I_2}{\sum I_1^2 + \sum I_2^2 - \sum I_1 \times I_2} \quad (10)$$

The F1 score is used to evaluate the detection between the *online* and *offline* results with the annotated data. The J.S. was also used to evaluate the similarity between the *online* and *offline* data and the annotated dataset. The J.S. has also been used to evaluate the difference between the results images from *online* and *offline* experiments.

Experiments and Results

We have tested the proposed algorithm on three datasets representing more than five thousand images. For each image of these datasets we have a ground truth image of the contents. We conducted two experiments, one *online* without the refinement steps and another *offline* with the refinement steps. For each dataset the processing time of each image has been measured, then an average time has been calculated. The results are presented in table : 1

TABLE 1 – Evaluation of the segmentation results between the *online* and *offline* results

Database	time (m.s.)
1	3.708
2	4.103
3	3.885

However the analysis of the metrics result gave a very poor outcome, as shown in tables 2 and 3.

TABLE 2 – Evaluation of the segmentation results with the ground truth

Database	<i>online</i>		<i>offline</i>	
	J.S.	F1 Score	J.S.	F1 Score
1	0.5815	0.5532	0.5783	0.5540
2	0.3794	0.7066	0.3809	0.6952
3	0.6492	0.4903	0.6378	0.4966

TABLE 3 – Evaluation of the segmentation results between the *online* and *offline* results

Database	J.S.
1	0.5607
2	0.2603
3	0.6383

With the first database almost 45% of the pixels segmented were misclassified while 42% of the segmented pixels structures mismatch those of the ground truth. The results are little bit better with the second database but worse on the third. The comparison between the *online* and *offline* images only confirm the results. After several investigations on the performances of the algorithm, the ground truth images were examined more closely. The ground truth has been labelled by only one non expert observer. The lack of precision he introduced is probably the main reason for the poor results obtained by both the F1 score and the J.S. Score.

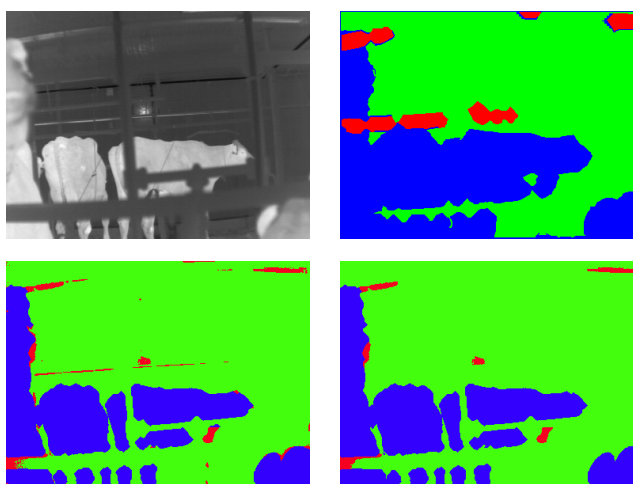


FIGURE 4 – Top left : the source image. Top right : the ground truth map (green = background, blue = foreground, red = false detection). Bottom left : online algorithm result map. Bottom right : offline algorithm result map

However the observer was not an expert nor a professional so he is not to blame. From a statistical point of view if several observers could have carried out the same work of labelling, the ground truth would be more accurate for the input images. This issue illustrates the need to have a well labelled database.

However this does not explain the poor quality of the J.S. between the *online* and *offline* images. The setting provided to the algorithm of [3] as a minimum number of elements to create an existing cluster can partly explain the results.

Qualitatively the majority of the results produced either by the *online* or by the *offline* approach closely resemble the source image, but without an accurate ground truth to quantitatively evaluate these results, it is not possible to reach a conclusion concerning the robustness of the proposed method.

Conclusion

The segmentation of an object of interest is one very basic and long-time studied application of image processing. In animal welfare the literature in this field provides many segmentation approaches such as linear thresholding that is not adapted for indoor applications due to the temperature of lights, electrical wires and other electrical devices that could be present in the scene. Other thresholding approaches such as Active Shape Model are based on the optimization of a cost function that makes them very computationally demanding and not suitable for real time applications. A last category of methods use data driven approaches. These approaches are not very computationally demanding and give good results but they require that some priors be set and the quality of the result is influenced by the statistical distributions of the data. On indoor applications the noisy scene elements such as light often have a statistical distribution very close to that of the object of

interest that makes such approaches inefficient for such applications.

In this paper we have proposed a new algorithm which not only detect the object of interest but also evaluates the object in the scene that could be considered as a false detection due to the fact that they share some properties with the object of interest.

The proposed algorithm provides an intermediary solution, with a complexity higher than linear approaches but it does not need to optimize any cost functions in order to give an accurate result.

In addition to the accuracy of the detection, by making use of theory mainly based on mathematical morphology and blob processing, this approach can satisfy the realtime requirement.

However due to an imprecise ground truth, it was impossible to quantitatively evaluate the results of this algorithm. Further investigation will be made in order to quantitatively evaluate the accuracy of the method. «However, one can say that «qualitatively» the obtained results are «sound and satisfactory» as confirmed in the presented results»

Acknowledgement

This research was supported by the Fond Quebecois de Recherche - Nature et Technologie (FQRNT). The author would like to thank Ahmed Trabelsi for his efforts, Thierry Moszkowicz for his advice during the development of the algorithm and Annette Schwerdtfeger for the time and energy she spent reviewing this article.

Références

- [1] Mauro dos Santos de Arruda, Bruno Brandoli Machado, Wesley Nunes Gonçalves, João Henrique Pinheiro Dias, Laury Cullen, Cristina Cunha Garcia, José Fernando Rodrigues Junior, et al. Thermal image segmentation in studies of wildlife animals. In *Workshop de Visão Computacional, XI*. Universidade de São Paulo–USP, 2015.
- [2] Peter Christiansen, Kim Arild Steen, Rasmus Nyholm Jørgensen, and Henrik Karstoft. Automated detection and recognition of wildlife using thermal cameras. *Sensors*, 14(8) :13778–13793, 2014.
- [3] Jian-Jiun Ding, CJ Kuo, and WC Hong. An efficient image segmentation technique by fast scanning and adaptive merging. *CVGIP, Aug*, 2009.
- [4] Yuncong Feng, Xuanjing Shen, Haipeng Chen, and Xiaoli Zhang. A weighted-roc graph based metric for image segmentation evaluation. *Signal Processing*, 119 :43–55, 2016.
- [5] Ying-Tung Hsiao, Cheng-Long Chuang, Joe-Air Jiang, and Cheng-Chih Chien. A contour based image segmentation algorithm using morphological edge detection. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 3, pages 2962–2967. IEEE, 2005.
- [6] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296) :23–27, 1975.
- [7] Sylwia Sikorska. Segmentation with image thresholding algorithms. *Challenges of Modern Technology*, 2(2) :19–22, 2011.
- [8] Tom Wirthgen, Georg Lempe, Stephan Zipser, and Ulrich Grunhaupt. Level-set based infrared image segmentation for automatic veterinary health monitoring. In *Computer Vision and Graphics*, pages 685–693. Springer, 2012.
- [9] Tom Wirthgen, Stephan Zipser, Ulrike Franze, Steffi Geidel, Franz Dietel, and Theophile Alary. Automatic segmentation of veterinary infrared images with the active shape approach. In *Image Analysis*, pages 435–446. Springer, 2011.
- [10] Tom Wirthgen, Stephan Zipser, Ulrike Franze, Steffi Geidel, and Georg Lempe. Automatic infrared based temperature measuring system for health monitoring in veterinary applications. In *Sensors, 2011 IEEE*, pages 1800–1803. IEEE, 2011.

10.21611/qirt.2016.034

- [11] Ian T Young, Jan J Gerbrands, and Lucas J Van Vliet. *Fundamentals of image processing*. Delft University of Technology Delft, The Netherlands, 1998.
- [12] GW Zack, WE Rogers, and SA Latt. Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry & Cytochemistry*, 25(7) :741–753, 1977.