Learning Based Image Segmentation of Pigs in a Pen

Mikael Nilsson, Håkan Ardö, Kalle Åström
Centre of Mathematical Sciences
Lund University
Lund, Sweden
micken@maths.lth.se, ardo@maths.lth.se, kalle@maths.lth.se

Anders Herlin, Christer Bergsten, Oleksiy Guzhva
Department of Agricultural Biosystem and Technology
Swedish University of Agricultural Sciences
Alnarp, Sweden
anders.herlin@slu.se, christer.bergsten@slu.se, oleksiy.guzhva@slu.se

***Abstract*—As farms are getting bigger with more animals, less manual supervision and attention can be given the animals on both group and individual level. In order not to jeopardize animal welfare, automated supervision is in some way already in use. Function and control of ventilation is already in use in modern pig stables, e.g. by the use of sensors for temperature, relative humidity and malfunction connected to alarm. However, by measuring continuously directly on the pigs, more information and more possibilities to adjust production inputs would be possible. In this work, the focus is on a key image processing algorithm aiding such a continuous system - segmentation of pigs in images from video. The proposed solution utilizes extended state-of-the-art features in combination with a structured prediction framework based on a logistic regression solver using elastic net regularization. Objective results on manually segmented images indicate that the proposed solution, based on learning, performs better than approaches suggested in recent publications addressing pig segmentation in video.**

I. Introduction

Health is one pillar of good animal welfare [1]. Thus, prevention and control of diseases and parasites are widely regarded as fundamental to animal welfare [2]. Provision and control of environment is also crucial as animals cannot choose the optimal environment in confined situations. The thermal environment is utterly important for pigs as they lack sweat glands. The pig wants to wet the body during hot weather in order to produce some cooling evaporation and if they cannot do this, their welfare is at stake and they may even die. The control and steering of climate by ventilation is thus of utterly importance. Farmed animals are in the hands of humans and it is our responsibility to provide adequate environment and prevent disease, find it quickly and to treat affected animals.

Precision Livestock Farming (PLF) is defined as a continuous monitoring of farm animals by sensors where the information is processed and compared with predictions and actions [3]. This will help farmers to monitor and reveal deviations from the predicted "normal" behaviour. The use of PLF can thus be a tool to prevent losses in animal production, improve profitability and minimise adverse environmental impact and at the same time promote animal welfare [4]. The use of image analysis technology, in PLF context, to monitor animals, has been used recently with some success [5], [6]. To come further in the practical use of the technology, there is a growing need for computer vision and machine learning approaching these challenges [4]. The focus of this paper is to address image segmentation of pigs in a pen, in order to bring PLF closer to its objectives.

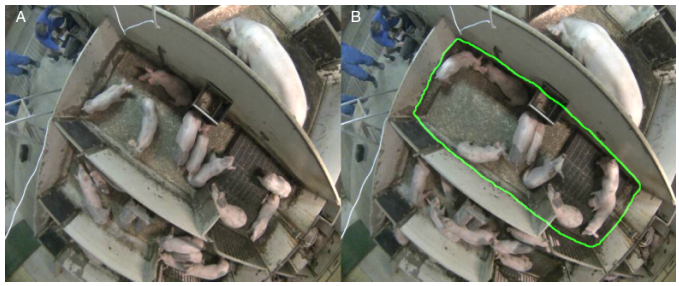

Fig. 1: (A) Top-down view of pigs in a pen. (B) Manually marked region of interest.

The purpose of image segmentation is to partition an image in meaningful information for a specific application. Works on image segmentation in a general context, typically explore various learning methods. These learning methods involve various forms of structured prediction, in order to achieve state of the art image segmentation results [7], [8], [9], [10].

In this work, we approach the specific task of segmenting pigs in a pen by proposing a framework using a structured prediction approach using several image features. This method is further compared and evaluated on manually annotated frames from video.

II. Equipment and Video Data

Pigs in a pen located at the pig husbandry site of Odarslöv in Sweden was the main recording site. Nine pigs in a pen were filmed in a top down view by an Axis M-3006 camera producing a 640×480 color mjpeg video, see Fig 1A. A manually marked Region Of Interest (ROI) capturing the pen was used, see Fig 1B.

III. Otsu's Gray-scale Approach

Recent works in the task of pig segmentation in pens for various analysis have utilized Otsu's method on gray-scale images [5], [6], [11]. Hence, as a baseline a comparison to this methods is employed. It should be noted that Otsu's method might work very well in various situations. In fact, it has been successfully applied in some scenarios [5], [6]. These scenarios involves fairly dark background and bright target (i.e. pig) pixels. However, in more uncontrolled scenarios, as the one addressed in this paper, we found it to fall somewhat short for practical use, see Fig. 4C. This observation lead us to pursue and investigate a learning based methodology to overcome

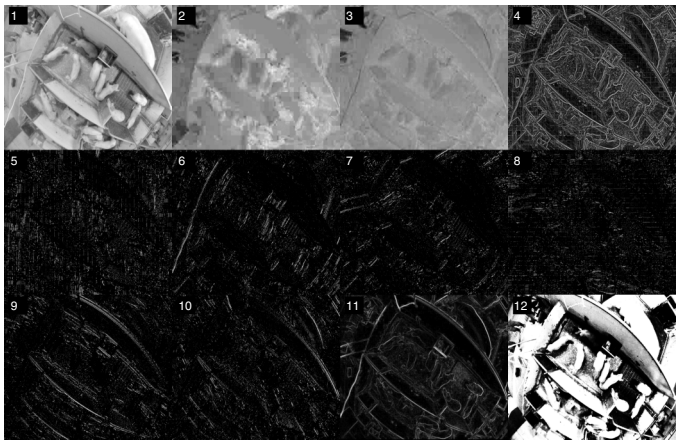

Fig. 2: Channel features used. From left to right and top to bottom are LUV color space (channel 1-3), gradient magnitude (channel 4), six oriented gradients (channel 5-10), max-min filter result (channel 11) and soft Otsu (channel 12).

some the the issues using Otsu's method on gray-scale images for this task.

IV. Channel Features and Simplified Structured Prediction Approach

In this section, the proposed framework for segmentation is presented. In general, the framework follows a typical two step pattern recognition approach; feature extraction and classification. What follows is the description of the proposed solution using extended state-of-the-art features and the framework for a structured prediction classifier based on logistic regression.

A. Channel Features

State-of-the-art channel features from pedestrian detection are used as the base [12], [13]. Ten feature channels corresponding to LUV color space (3 channels), normalized gradient magnitude (1 channel) and oriented gradients (6 channels) are used, see the first ten channels in Fig. 2. The channels, denoted C, can be seen as an $M \times N \times D$ cube with M being rows, N columns and D the number of channels. For example, $C_{1,2,3}$ is the pixel at row one and column two for channel three (V in LUV color space).

These ten channel features are extended with two additional channels. The first channel extension is a max-min filter operating on a 3×3 patch from the gray-scale image $L_{i,j}$ (=$C_{i,j,1}$) by finding the difference between maximum and minimum in the patch

$$C_{i,j,11} = \max_{\substack{k=i-1,i,i+1 \\ l=j-1,j,j+1}} L_{k,l} - \min_{\substack{k=i-1,i,i+1 \\ l=j-1,j,j+1}} L_{k,l}, \forall i,j. \quad (1)$$

This filter captures small local variations, and complements the gradient magnitude with finer edge information, see channel eleven in Fig. 2.

The second channel exploits Otsu's method, this since it has been used previously with some success. Rather than using a hard threshold decision from the method as the channel, a modification is applied. The modification enables a softer decision by finding the two means, μ_{fg} and μ_{bg}, and standard deviations,σ_{fg} and σ_{bg}, for foreground and background pixels given by Otsu's threshold. Then, this soft Otsu channel is found as

$$C_{i,j,12} = \frac{f(L_{i,j},\mu_{fg},\sigma_{fg})}{f(L_{i,j},\mu_{fg},\sigma_{fg}) + f\left(L_{i,j},\mu_{bg},\sigma_{bg}\right)}, \forall i,j \quad (2)$$

where

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

is the normal distribution, see channel twelve in Fig. 2.

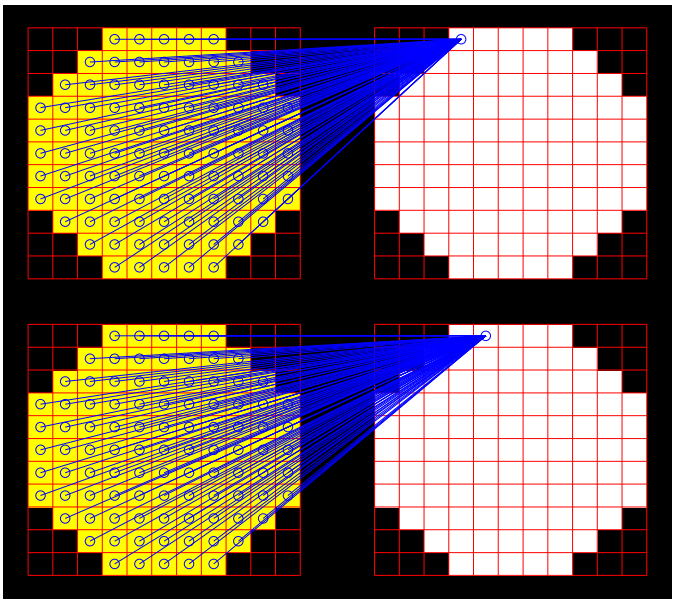

Fig. 3: Examples of two, of a total A corresponding to the area, learning problems building up the structured output. Left is input and right is output. Yellow position indicate a vector of size D (the channels) and white indicate a single output value which will be a probability from logistic regression.

B. Structured Prediction

The approach utilized here involves learning using structured prediction. That is, given $i = 1,2,\ldots,N$ training examples $\mathbf{x}_i \in \mathbb{R}^{D_x}$ and the associated structured output $\mathbf{y}_i \in \mathbb{R}^{D_y}$ a model, given an input $\mathbf{x}$, should be able to predict the structured output $\mathbf{y}$. Note that the training input can be seen as a matrix of size $N \times D_x$ and, similarly, the output as a matrix of size $N \times D_y$. The input here is a circular area, of A positions, containing the channel features at each position (i.e. $D_x = A \cdot D$) and the output is the same circular area with probabilities at each position (i.e. $D_y = A$), see Fig. 3.

The structured prediction performed here is that individual components in y are treated independently. That is, the problem is simplified by treating it by D_y individual learning problems (learning from input matrix $\mathbf{x}$ to each row of output matrix $\mathbf{y}$). Each of these learning problems will use an elastic net regularized logistic regression as mapping, se next subsection. Thus, for a given pixel in the input image, the local circular area of the feature channels will be calculated, see Fig. 3 left, and D_y classifiers will results in a structured output, see Fig. 3

right. Hence, this procedure is performed for every pixel (pixel by pixel) in the ROI and will result in D_y probabilities from the logistic regression at each pixel. The final result, at each pixel, will then be found as the mean of these probabilities.

1) Elastic Net Regularized Logistic Regression: Given $i = 1,2,\ldots,N$ training samples $\mathbf{x}_i \in \mathbb{R}^D$, in form of features for an area, and the associated class labels $y_i \in \{-1,+1\}$, where $+1$ indicate target and -1 background, the goal is to design a classifier by optimization of an objective function. Note that y_i here is taken as a single row from the main output matrix $\mathbf{y}$. The main objective function used here is the logistic loss function, $L(\mathbf{w},b)$, defined as the *negative average log-likelihood*

$$L(\mathbf{w},b) = -\frac{1}{N}\log \mathcal{L}(\mathbf{w},b), \tag{4}$$

where $\mathcal{L}(\mathbf{w},b) = \prod_{i=1}^{N} \frac{1}{1+e^{-y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)}}$, $\mathbf{w}$ is a weight vector and b a bias [14]. This can be reformulated as

$$L(\mathbf{w},b) = \frac{1}{N}\sum_{i=1}^{N}\log\left(1+e^{-y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)}\right). \tag{5}$$

Adding an L_1-norm (lasso) regularization term

$$R_1(\mathbf{w}) = \|\mathbf{w}\|_1 = \sum_{j=1}^{D}|w_j| \tag{6}$$

and a squared L_2-norm (ridge) regularization term

$$R_2(\mathbf{w}) = \|\mathbf{w}\|_2^2 = \sum_{j=1}^{D} w_j^2 \tag{7}$$

together with regularization parameters λ_1 and λ_2, the objective function to minimize using elastic net (combined lasso and ridge) regularization is

$$J(\mathbf{w},b) = L(\mathbf{w},b) + \lambda_1 R_1(\mathbf{w}) + \lambda_2 R_2(\mathbf{w}). \tag{8}$$

Further details regarding solving this optimization can be found in the work by Nilsson [14].

V. Experiments and Evaluation

In order to get objective results, ten frames, taken spread out over the time of recording, were manually segmented. To speed up this manual segmentation, an interactive segmentation solution proposed by Gulshan et al. was used [15]. The manual segmentation results in a mask to be used as the desired output, see Fig. 4B.

Training and testing of the proposed system, using $A = 97$ (area size) and $D = 12$ (channels) implying $D_x = 1164$ and $D_y = 97$, was performed using a five-fold cross validation over the ten images. Each training was performed by using 25% of the image pixels from the ROI by random sampling to avoid exceeding RAM memory.

Otsu, in its basic form, is a single threshold and thereby a single operation point in Receive Operation Characteristic (ROC) space [16]. In order to compare ROC curves, and associated Area Under Curve (AUC), the soft Otsu, see Eq. (2), is employed. Results using Otsu and the proposed method using five-fold cross validation can be found in Fig. 5 and Table I.

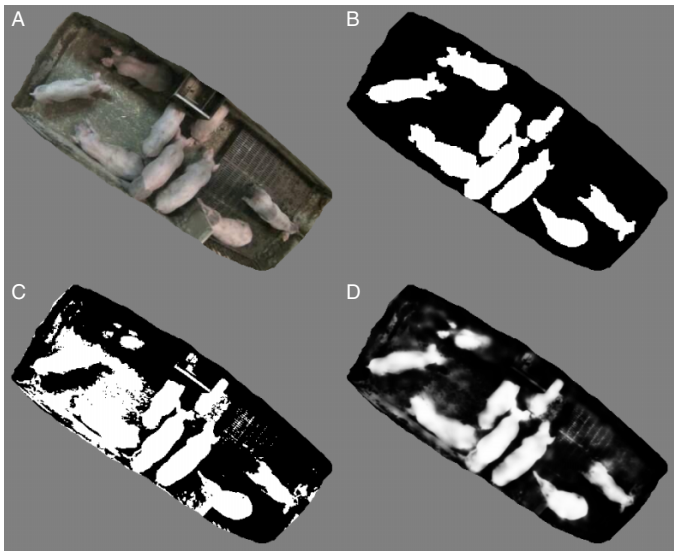

Fig. 4: (A) ROI of image. (B) Manual segmentation. (C) Otsu segmentation. (D) Proposed segmentation with probability result.

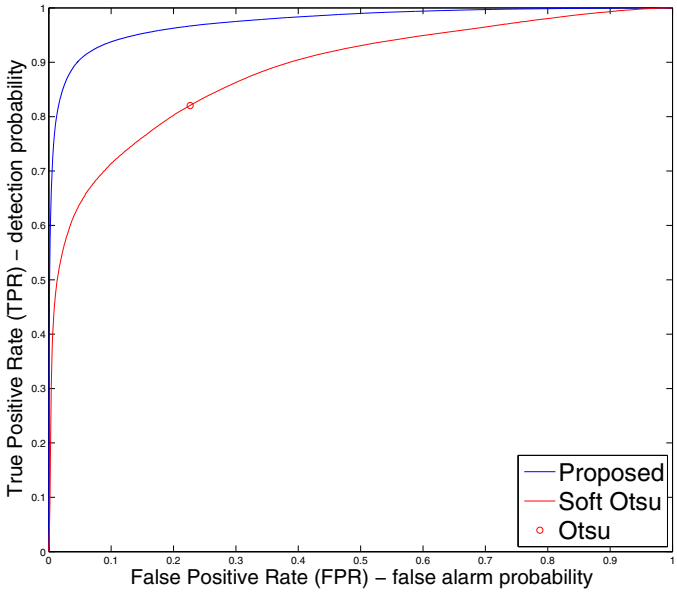

Fig. 5: Receive Operation Characteristic (ROC) curve for the proposed method, soft Otsu and the operation point for Otsu.

VI. Conclusion

A learning based framework using several features and a structured prediction approach has been proposed and presented for image segmentation of pigs in a pen. The system showed improved segmentation results compared to Otsu's method which has previously been proposed, and applied in practice, for pig segmentation in video. The method showed a 0.08 improvement in AUC (0.97 vs 0.89) when evaluating it on ten manually segmented images taken from video. Future work involves adding temporal information, for example by adding channel(s) with information from previous frames, in the framework to get better temporal coherence. Furthermore, it is desirable to further collect more data to investigate and com-

Method	AUC	TPR at Otsu FPR	FPR at Otsu TPR
Proposed	0.97	0.97	0.02
Soft Otsu	0.89	0.82	0.23

TABLE I: Area Under Curve (AUC) for proposed method, soft Otsu and the operating point for Otsu.

pare how methods, the two explained herein as well as adopting other techniques from recent computer vision research in segmentation, to see how variations such as other sites, longer videos, lightning conditions, etc effects performance. The segmentation framework presented is fairly generic. Creating new training data and investigating its applicability on other animals in the agriculture area is a natural next step.

REFERENCES

[1] M. S. Dawkins, "A user's guide to animal welfare science," *Trends in Ecology and Evolution*, vol. 21, no. 2, pp. 77 – 82, 2006.

[2] D. Fraser, I. J. Duncan, S. A. Edwards, T. Grandin, N. G. Gregory, V. Guyonnet, P. H. Hemsworth, S. M. Huertas, J. M. Huzzey, D. J. Mellor, J. A. Mench, M. Spinka, and H. R. Whay, "General principles for the welfare of animals in production systems: The underlying science and its application," *The Veterinary Journal*, vol. 198, no. 1, pp. 19 – 27, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1090023313003122

[3] D. Berckmans, "Automatic on-line monitoring of animals by precision livestock farming," in *Proceedings of the ISAH Conference on Animal Production in Europe: The Way Forward in a Changing World*, vol. 1, 2004, pp. 27–31.

[4] C. Wathes, H. Kristensen, J.-M. Aerts, and D. Berckmans, "Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall?" *Computers and Electronics in Agriculture*, vol. 64, no. 1, pp. 2 – 10, 2008, smart Sensors in precision livestock farming.

[5] S. Ott, C. Moons, M. Kashiha, C. Bahr, F. Tuyttens, D. Berckmans, and T. Niewold, "Automated video analysis of pig activity at pen level highly correlates to human observations of behavioural activities," *Livestock Science*, vol. 160, no. 0, pp. 132 – 137, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1871141313005568

[6] M. A. Kashiha, C. Bahr, S. Ott, C. P. Moons, T. A. Niewold, F. Tuyttens, and D. Berckmans, "Automatic monitoring of pig locomotion using image analysis," *Livestock Science*, vol. 159, no. 0, pp. 141 – 148, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1871141313005003

[7] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ser. ICCV '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 10–.

[8] J. Carreira and C. Sminchisescu, "CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts," *IEEE TPAMI*, vol. 34, no. 7, pp. 1312–1328, 2012.

[9] S. Nowozin, P. Gehler, and C. Lampert, "On parameter learning in crf-based approaches to object class image segmentation," in *ECCV*, September 2010.

[10] J. Domke, "Structured learning via logistic regression," in *Advances in Neural Information Processing Systems 26*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 647–655. [Online]. Available: http://papers.nips.cc/paper/4870-structured-learning-via-logistic-regression.pdf

[11] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Systems, Man and Cybernetics*, vol. 9, pp. 62–66, Mar. 1979, minimize inter class variance.

[12] P. Dollár, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," in *BMVC*, 2010.

[13] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *PAMI*, 2014.

[14] M. Nilsson, "Elastic net regularized logistic regression using cubic majorization," in *ICPR*, 2014.

[15] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[16] T. Fawcett, "Roc graphs: Notes and practical considerations for researchers," *ReCALL*, vol. 31, no. HPL-2003-4, pp. 1–38, 2004.