University of Groningen

A Computer Vision Pipeline that Uses Thermal and RGB Images for the Recognition of Holstein Cattle

Bhole, Amey; Falzon, Owen; Biehl, Michael; Azzopardi, George

Published in:
Computer Analysis of Images and Patterns

DOI:
[10.1007/978-3-030-29891-3_10](https://doi.org/10.1007/978-3-030-29891-3_10)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Bhole, A., Falzon, O., Biehl, M., & Azzopardi, G. (2019). A Computer Vision Pipeline that Uses Thermal and RGB Images for the Recognition of Holstein Cattle. In M. Vento, & G. Percannella (Eds.), *Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, Salerno, Italy, September 3–5, 2019, Proceedings, Part II* (pp. 108–119). (Lecture Notes in Computer Science; Vol. 11679). Springer. https://doi.org/10.1007/978-3-030-29891-3_10

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

A Computer Vision Pipeline that Uses Thermal and RGB Images for the Recognition of Holstein Cattle

Amey Bhole¹(✉), Owen Falzon², Michael Biehl¹,
and George Azzopardi¹(✉)

¹ University of Groningen, Groningen, The Netherlands
{a.bhole,m.biehl,g.azzopardi}@rug.nl

² University of Malta, Msida, Malta
owen.falzon@um.edu.mt

Abstract. The monitoring of farm animals is important as it allows farmers keeping track of the performance indicators and any signs of health issues, which is useful to improve the production of milk, meat, eggs and others. In Europe, bovine identification is mostly dependent upon the electronic ID/RFID ear tags, as opposed to branding and tattooing. The RFID based ear-tagging approach has been called into question because of implementation and management costs, physical damage and animal welfare concerns. In this paper, we conduct a case study for individual identification of Holstein cattle, characterized by black, brown and white patterns, in collaboration with the Dairy campus in Leeuwarden. We use a FLIR E6 thermal camera to collect an infrared and RGB image of the side view of each cow just after leaving the milking station. We apply a fully automatic pipeline, which consists of image processing, computer vision and machine learning techniques on a data set containing 1237 images and 136 classes (i.e. individual animals). In particular, we use the thermal images to segment the cattle from the background and remove horizontal and vertical pipes that occlude the cattle in the station, followed by filling the blank areas with an inpainting algorithm. We use the segmented image and apply transfer learning to a pre-trained AlexNet convolutional neural network. We apply five-fold cross-validation and achieve an average accuracy rate of 0.9754 ± 0.0097 . The results obtained suggest that the proposed non-invasive approach is highly effective in the automatic recognition of Holstein cattle from the side view. In principle, this approach is applicable to any farm animals that are characterized by distinctive coat patterns.

Keywords: Transfer learning · Individual identification · Computer vision · Holstein cattle

1 Introduction

Modern agriculture is developing rapidly in different sectors, such as organic farming, agribusiness, animal husbandry and precision agriculture, due to

(a) Thermal image

(b) RGB image

Fig. 1. Examples of thermal and RGB images acquired at the same time by the FLIR E6 thermal camera from a distance of five meters.

technological advancements. In animal husbandry, it is important for the farmer to keep track of the animal's performance and health in order to improve the production of commodities such as milk, cheese, meat and other products. An example where traceability of animals is important is when the cattle are approaching or leaving the milking station. Identification methods based on collar id and ear-tagging are used for individual recognition of cattle. These methods do not provide effective solutions for the identification of cattle due to loss of ear-tags, fading of labels on ear-tags and physical damage [6, 9, 21]. In Europe, bovine identification is mostly dependant on electronic identification devices, such as radio frequency identification (RFIDs) [26]. It is a technology that uses electromagnetic waves for the identification and tracking of the RFID tags placed on the ears of Holstein cattle. The RFID tags are also used with an ear-tagging approach for cattle identification. That approach, however, has issues related to implementation, cost and management of large scale monitoring of livestock animals [8]. There is, therefore, a need for non-invasive methods for the identification of cattle in farms.

In this paper, we explore the possibility of integrating animal biometric with modern agriculture. We perform a case study to identify individual Holstein cattle based on the unique coat pattern. We propose a solution consisting of a pipeline of image acquisition using a thermal camera, image processing and transfer learning [19] using a pre-trained convolution neural network. We also release a benchmark data set consisting of side-view images of Holstein cattle, as shown in Fig. 1, collected at the Dairy Campus in Leeuwarden, the Netherlands. The proposed solution provides an effective system for the individual identification of Holstein cattle using a non-invasive computer vision approach, which can be set up in a practically relevant environment.

2 Related Work

In this section, we evaluate three different methodologies proposed in the literature for individual identification of cattle. Specifically, for each approach, we

describe the classification performance, the ability to automate the process, and the adequacy of using the method in a relevant practical setting.

Almost a century ago it was found that the muzzle patterns of cows exhibit individually unique traits [20]. Such patterns have been investigated and turned out to be, indeed, promising biometric features. Handcrafted feature-description algorithms, such as Scale-Invariant Feature Transform (SIFT) [17], have been used for muzzle prints for individual identification of cattle with classification accuracy of 93% [2]. In more recent work a deep learning framework [15] was used for individual identification of cattle and identification accuracies in the region of 95.98%–98.9% were reported for tests carried out on 500 subjects. That approach, however, can be burdensome to be completely automated and executed in an application where the cattle has to be identified during the milking process. Due to the non-cooperative behaviour of cattle, it is difficult to set up a system for acquiring good quality images of muzzles.

Identification of animals, such as horses or cattle, can also be performed based on iris patterns including arching ligaments, furrows, ridges, crypts, rings and zigzag collarette [4]. A new cattle identification system was proposed based on iris analysis and recognition in [18], and it achieved identification accuracy of 96% for 6 cows. While those results look promising more experiments need to be carried out on a large database to better assess the robustness of that approach. Moreover, the eye image captured by a video camera may lack uniformity making this method difficult to automate.

In recent work, the coat patterns of the Holstein cattle, which exhibit black, brown and white coat patterns, patches, and markings over their bodies were used for individual identification. In particular, an R-CNN [22] adaptation of the VGG CNN M 1024 network was used for individual identification of cattle [1]. That approach achieved a mean average precision of 86.07% over two-fold cross validation for the identification of 89 cows using top-view still RGB images. The promising results demonstrated that individual identification of Holstein cattle is possible based on the coat patterns.

In the method that we propose we also use the coat patterns for individual identification but unlike the work in [1], we leverage the side-view instead of the top-view images to capture maximum information related to the coat patterns of cattle. We also use thermal images to capture the temperature distribution in the scene and use it to segment the cow from the background. Moreover, we use an inpainting technique to replace occluding components with more meaningful values related to the coat pattern of the cattle. Finally, we use the AlexNet convolutional neural network [16] with transfer learning to demonstrate the effectiveness of the conceptually simple pipeline that we propose. AlexNet [14] made a breakthrough in computer vision by achieving outstanding results in the image classification tasks of the ImageNet Large Scale Visual Recognition Challenge [23]. In recent applications, transfer learning has been employed with pre-trained CNNs to overcome the problem of the restricted availability of data in certain applications, and it has been shown how effective it is in different applications, including health care [7, 27], natural language processing [13] and self-driving cars [3].

Fig. 2. Schematic overview of the proposed solution

3 Method

In Fig. 2 we illustrate a schematic overview of the proposed solution. It consists of two main components, namely configuration and application. The configuration part includes data acquisition, data pre-processing and determination of the classification model by transfer learning. The data acquisition is used to capture the side-view of the cattle when they are leaving from the milking station. Figure 1 shows an example of a pair of thermal and RGB images captured while the concerned cow was walking out from the milking station. In data pre-processing we use the distinctive body temperature of the cattle to segment it from the background and apply an inpainting algorithm to fill the gaps generated by the occluding rods. The preprocessed images are then used to learn our classification model.

3.1 Configuration

3.1.1 Data Acquisition

The data acquisition process was carried out at the Dairy campus in Leeuwarden over a period of nine days using the FLIR E6 thermal camera, which has a straightforward point and shoot operation. The camera has an image resolution of 640×480 pixels and the ability to detect temperature differences as small as 0.06°C . The camera has three main parameters to be evaluated for collection of images, namely emissivity, distance and reflected temperature. The emissivity of a surface depends not only on the material but also on the nature of the surface. In this case, the emissivity used was 0.98. The distance parameter in the camera

is required to align the infrared and RGB images, which were captured from the distance of five meters from the cattle. The reflected temperature is thermal radiation originating from other objects that reflect off the target and was set to 20°C.

3.1.2 Pre-processing

We apply two pre-processing steps to the acquired images, namely segmentation [10] and inpainting. We use the FLIR tools to binarize the thermal images with a threshold of 25.75°C. We chose this value manually after visually inspecting some images of our data set. The resulting binary images contain some components in the background and missing parts from the cows due to the presence of rods. In order to remove the gaps caused by the rods, we apply the area closing morphology operation with a line structuring element of radius 18 pixels. In this way, all parts belonging to the cow become one connected component. Finally, we use connected component analysis to determine the largest component, which represents the cow, and discard the rest. Figure 3b shows a binarized image as a result of the segmentation algorithm that we use, and Fig. 3c shows the extracted region of interest after morphological and connected component analysis.

Next, we crop the connected component representing the cow and recover the gaps caused by the rods. In order to ensure that all rods-related pixels are removed we widen slightly the gaps with a dilation operation that uses a disc structuring element of radius 3 pixels. Figure 3d illustrates the widened gaps on the corresponding cropped RGB image. Finally, we apply the inpainting algorithm based on the fast marching method proposed in [25] to replace the gaps with more relevant values determined from the local neighbourhoods, Fig. 3e.

3.1.3 Classification Model

We build a classification model by using transfer learning on a pre-trained model. Transfer learning exploits source knowledge gained from learning one task and applying it to solve another one [11], and it is suitable when there is limited availability of training images per class, such as our data set. We use the AlexNet convolutional neural network, which was pre-trained on the ImageNet Database [5] and fine tuned it with our training data. The architecture of AlexNet consists of five convolutional layers and three fully connected layers where the ReLU activation function is applied after every convolutional and fully connected layers. Dropout regularization [24] of 50% is applied before the first and the second fully connected layer. We changed the last fully connected layer of AlexNet by replacing it from the 1000 classes used for the image classification of ImageNet to 136 classes corresponding to the number of distinct cows in our data set.

3.2 Application

A given query image is evaluated by first applying the pre-processing steps described above and then fed to the AlexNet CNN classifier. The output of the network is the identifier of the cow to which the given query image is assigned.

Fig. 3. Segmentation of the region of interest. (a) Thermal image with lower and upper limits clipped at 25.7°C and 25.8°C , respectively. (b) The binary image obtained after using a threshold of 25.75°C on the image in (a). (c) The cow mask obtained by dilating the image in (b) and determining the largest connected component. (d) The corresponding cropped RGB image with dilated gaps and (e) the result of the inpainting that we use.

AlexNet CNN classifier can be deployed on various cloud platforms. It can also be updated periodically for new classes.

4 Experiments and Results

Below we provide details about the data set used, the evaluation method, hyper-parameters and the results of our experiments.

4.1 Data Set

In this study we use images with resolution of 640×480 pixels before pre-processing. The data set contains 1237 unique pairs of RGB and infrared images of 136 cows. Figure 4 shows the distribution of the number of images per class (i.e. per cow) sorted in ascending order, where the median is 9, the mean is 9.09, and the minimum and maximum number of images per class are 6 and 14, respectively.

4.2 Evaluation Methods

We evaluate the proposed approach by using five-fold cross validation with stratified sampling, and report the average performance measurements. The metrics

Fig. 4. Distribution of the number of images per class (i.e. per cow), with a mean of 9.09 and standard deviation of 1.49.

used for the evaluation of the results are accuracy, confusion matrix, precision, recall and F1-score [12].

4.3 Training Parameters

According to the requirement of AlexNet we resize the images to 227×227 pixels and normalise them by subtracting the mean. We use the stochastic gradient descent with momentum as the optimizer to re-train the last layer of AlexNet classification model with 50 epochs for each of the five folds. The mini-batch size for training is 32, the global learning rate is set to 0.001 and momentum of 0.9. All experiments, including training and evaluation, were carried out with a computer equipped by NVIDIA GeForce GTX 1050 and Intel Core i7-7700HQ CPU @ 2.80 GHz with 24 GB RAM.

4.4 Results

Table 1 reports the average results obtained over five folds of experiments and Fig. 5 shows the sum of confusion matrices obtained over the five folds. There were a total of 31 images that were misclassified across all folds. For each fold, we fine-tuned a new AlexNet classifier by transfer learning using the training images available in that fold and we calculated the macro F1-score, which is based on the average precision and recall computed for each class in a one-vs-all setup.

Table 1. Mean results obtained for individual identification from five-fold cross validation

Accuracy	Macro precision	Macro recall	Macro F1-score
0.9754 ± 0.0097	0.9808 ± 0.0060	0.9765 ± 0.0082	0.9786 ± 0.0062

Fig. 5. Sum of the confusion matrices across the five folds. The circles indicate the positions of the false positives, which all have a value of 1.

5 Discussion

The contributions of this work are two-fold. Firstly, we propose an effective pipeline of image acquisition, image processing, computer vision and machine learning components, which can be deployed in farms for the identification of cattle from the side-view in milking stations. To the best of our knowledge, this is the first study that addresses this challenge and proposes an effective solution that can be easily deployed. Secondly, we make available¹ a data set of 1237 pairs of thermal and RGB images of 136 cows, which can be used for benchmarking.

In this work we use the thermal images only for segmentation purposes and then we apply the AlexNet CNN on the corresponding RGB data. This configuration expects all cows to have sufficient amount of coat pattern and, therefore, it lacks robustness to the recognition of cattle in the rare cases where there is no or very little coat pattern, Fig. 6. In order to make the system more robust, in future work we plan to investigate a fusion approach that uses features from both RGB and thermal images for the final classification of a given query image. Furthermore, retraining AlexNet with limited number of images per class may be prone to overfitting. In order to improve the generalization, we plan to implement different data augmentation techniques.

While the aisle in the milking station is restricted to allow one cow at a time, it may happen that two cows appear in the same image. The scope of this study did not consider such situations and, in fact, some of the images were misclassified due to this problem, as shown in the examples given in Fig. 7. In order to overcome this issue, in future work we plan to investigate techniques

¹ The data set is available here: <http://infosysdemos.cs.rug.nl>.

that can identify multiple cattle in the same image and segment only the one that is closest to the center, for instance.

Fig. 6. (a) Example of an incorrectly classified query image due to insufficient coat pattern. In (b–d) and (e–g) there are samples of the training images from the class (ID-0675) with which the query image was supposed to be classified and samples of the training images from the class (ID-7600) with which it was incorrectly classified, respectively. All instances are represented by (top) the preprocessed image and (bottom) the original RGB image.

Another direction for future research is to explore the capturing of multiple images of the same cow while it is walking through aisle. In that case, we may investigate the processing of all images through the classifier and finally aggregate all decisions into a stronger one. Such a solution would circumvent the situations where the occluded part of the cattle contains one or more of the most defining features, which cannot be recovered by the inpainting algorithm. Figure 8 shows one such example. The consideration of multiple images of the same query cow would substantially reduce the risk of such situations.

Fig. 7. The images in (a–b) are the original RGB and preprocessed images with two cattle, and the images in (c–d) show another instance of the main cow, with ID equals to 9884, alone.

Fig. 8. Two pairs of RGB and preprocessed images of the cow with ID equals to 0091. In the images in (c–d) the cow has one of the defining coat features occluded, and the inpainting algorithm fails to reconstruct it properly.

6 Conclusions

The proposed pipeline for the recognition of cattle from the side-view in milking stations is very effective. Our approach is completely automatic, non-invasive, and does not require the farms to change anything in their infrastructure or in the way they handle cattle in the milking stations. The pipeline consists of image acquisition with a thermal camera, segmentation, inpainting and the AlexNet CNN with transfer learning. The proposed solution is flexible to be integrated with systems already present in the farms and can be deployed on cloud platforms.

Acknowledgements. We thank the Dairy campus in Leeuwarden for permitting the data collection used in this project and for approving its availability for academic use.

References

1. Andrew, W., Greatwood, C., Burghardt, T.: Visual localisation and individual identification of Holstein Friesian cattle via deep learning. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2850–2859 (2017)

2. Awad, A.I., Zawbaa, H.M., Mahmoud, H.A., Nabi, E.H.H.A., Fayed, R.H., Hassanien, A.E.: A robust cattle identification scheme using muzzle print images. In: 2013 Federated Conference on Computer Science and Information Systems, pp. 529–534. IEEE (2013)
3. Choi, D., An, T.H., Ahn, K., Choi, J.: Driving experience transfer method for end-to-end control of self-driving cars. arXiv preprint [arXiv:1809.01822](https://arxiv.org/abs/1809.01822) (2018)
4. Daugman, J.G.: High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(11), 1148–1161 (1993)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: CVPR09 (2009)
6. Edwards, D., Johnston, A., Pfeiffer, D.: A comparison of commonly used ear tags on the ear damage of sheep. *Anim. Welf.* **10**(2), 141–151 (2001)
7. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**(7639), 115 (2017)
8. Feng, J., Fu, Z., Wang, Z., Xu, M., Zhang, X.: Development and evaluation on a RFID-based traceability system for cattle/beef quality safety in China. *Food control* **31**(2), 314–325 (2013)
9. Fosgate, G., Adesiyun, A., Hird, D.: Ear-tag retention and identification methods for extensively managed water buffalo (*bubalus bubalis*) in trinidad. *Prev. Vet. Med.* **73**(4), 287–296 (2006)
10. Fu, K.S., Mui, J.: A survey on image segmentation. *Pattern Recognit.* **13**(1), 3–16 (1981)
11. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016). <http://www.deeplearningbook.org>
12. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York (2001). <https://doi.org/10.1007/978-0-387-21606-5>
13. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. arXiv preprint [arXiv:1801.06146](https://arxiv.org/abs/1801.06146) (2018)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
15. Kumar, S., et al.: Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* **116**, 1–17 (2018)
16. LeCun, Y., et al.: Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1**(4), 541–551 (1989)
17. Lowe, D.G., et al.: Object recognition from local scale-invariant features
18. Lu, Y., He, X., Wen, Y., Wang, P.S.: A new cow identification system based on iris analysis and recognition. *Int. J. Biom.* **6**(1), 18–32 (2014)
19. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010)
20. Petersen, W.: The identification of the bovine by means of nose-prints. *J. Dairy Sci.* **5**(3), 249–258 (1922)
21. Phillips, C.: Cattle Behaviour and Welfare. Wiley, Hoboken (2008)
22. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
23. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015)

24. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014). <http://jmlr.org/papers/v15/srivastava14a.html>
25. Telea, A.: An image inpainting technique based on the fast marching method. *J. Graph. Tools* **9**(1), 23–34 (2004)
26. Wamba, S.F., Anand, A., Carter, L.: RFID applications, issues, methods and theory: a review of the AIS basket of TOP journals. *Procedia Technol.* **9**, 421–430 (2013)
27. Wieslander, H., et al.: Deep convolutional neural networks for detecting cellular changes due to malignancy. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 82–89 (2017)