J. of Ag. Eng. - Riv. di Ing. Agr. (2011), **4**, 17-23

PRELIMINARY STUDY FOR THE IMPLEMENTATION OF AN IMAGE ANALYSIS ALGORITHM TO DETECT DAIRY COW PRESENCE AT THE FEED BARRIER

Simona M.C. Porto, Claudia Arcidiacono, Giuseppe C. Guarnera, Giovanni Cascone

1. Introduction

The behaviour analysis of animals housed in livestock buildings may contribute to predict the outbreaks of major diseases (e.g., bovine spongiform encephalopathy (BSE), foot and mouth disease (FMD), and swine fever [Corkery 2007]), as well as animal discomfort due to external factors (e.g., management system, microclimatic conditions inside the building, housing layout, and building materials).

Concerning dairy cows, daily life in free-stall barns includes the following activities: lying down in stalls, eating, drinking, standing, and walking along the alleys. The analysis and assessment of such behaviours could be carried out by means of suitable indices which could be computed by adopting automated methods such as three-dimensional accelerometers, embedded sensor technology, and automatic local position systems [Müller 2003; Mattachini 2011].

The automatic analysis of digital images from video recordings represents an effective alternative to traditional techniques used to study dairy cow behaviour (e.g., completing on-site survey reports, completing survey reports by visual analysis of video recordings taken in the barn [Carreira 2009; Provolo 2009], and analyzing data collected by sensors placed on the animal body or inside it). It is comparatively inexpensive, non invasive for animals and, when real-time classification algorithms are used, it makes it possible to store the results of the elaborations relative to long observation periods, avoiding the onerous storage process of video recordings [Cangar 2008].

When the images of the observed animals are not enough in contrast with the background, the methodologies developed in several studies [Shao 1998; Cangar 2008; Shao 2008] generally require the application of a number of image pre-processing steps to segment out the animal features in the frame.

To overcome these problems, this research, that is still in progress, aims at assessing the application of an image analysis algorithm to detect dairy cows housed in a free-stall barn with open sides in order to perform the analysis of their behaviours (e.g., eating, standing, walking, and lying). The detection model developed in this study was based on the algorithm of Paul Viola and Michael Jones [2001; 2004], which was originally implemented for human face detection. The algorithm does not work directly on the pixels of the image to be processed, but on Haar-like feature values. To compute these feature values very rapidly, Viola and Jones used integral images, obtained by means of a few operations carried out on the pixels of the images to be processed. The overall form of the detection process is a 'cascade' of strong classifiers (stages), each of them generated by using the Adaboost algorithm [Freund 1995]. From literature, it results that the robustness of this algorithm could provide accurate classifications also when significant brightness and background variations occur in the sequence of the analyzed images. Furthermore, since this algorithm is suitable for real-time elaborations, it would avoid the onerousness of video-recording storage.

This paper shows the preliminary results of the implementation of the above mentioned image analysis algorithm for the detection of dairy cow presence at the feed barrier. In this work, as in previous studies [De Vries 2003; Wilson 2005], this kind of behaviour is operationally defined as "feeding behaviour". The scenes for the training, the testing and the validation of the classifier were extracted from the video recordings of a video camera placed above the feeding area providing the plan view of a region of the feeding alley. Plan views of the observed animals have been used in several experimental trials in order to develop real-time monitoring systems [Shao 1998; Cangar 2008]. Plan views of the cows make it possible to dis-

Paper received 01.06.2011; accepted 08.11.2011

Simona M.C. Porto, Phd, siporto@unict.it, Claudia Arcidiacono, associate professor, carcidi@unict.it, Giovanni Cascone, full professor, gcascone@unict.it, University of Catania, Department of Agri-Food and Environmental Systems Management, Section: Building and Land Engineering, via S. Sofia n. 100, 95123, Catania, Italy. Giuseppe C. Guarnera, Phd student, guarnera@dmi.unict.it, University of Catania, Department of Mathematics and Informatics, viale A. Doria n. 6, 95123 Catania, Italy. Tel. 0039.0957147576, fax 0039.0957147600.

The authors' contribution is equal in every way.

tinguish each cow, as well as to analyze its body shape which completely changes in relation to cow behaviour. The results discussed in this paper are referred only to the video recordings obtained by the considered camera.

2. Materials and methods

2.1 *The barn and the sample of observed cows*

The research has been carried out in a cubicle free-stall barn for dairy cows equipped with innovative devices for microclimate control used to improve the thermal comfort of the cows.

With regards to the layout, the barn had a row of 64 head-to-head free-stalls (cubicles) with sand as bedding, bounded on the North by the feeding alley and on the South by the service alley.

In the alleys there was presence of slurry accumulation as the cleaning was not automated but it was carried out 1-2 times a day by a scraper. As a consequence, sunlight reflection on wet floors occurs. High floor brightness was observed on the service alley along the southern open side and the resting area.

Within the barn an area was selected where a group of 15 dairy cows was housed. Within this area, two functional units have been observed by means of a video recording system constituted by 10 cameras: the feeding area, which was a rectangular area of about 16.55 m × 3.50 m, and the resting area, which was a rectangular area of about 10.4 m × 4.30 m adjacent to the feeding area.

2.2 *The video recording system and the calibration process*

Since the barn ceiling height was not adequate to obtain a broad coverage of the barn floor from above, a wide-angle camera was required. Therefore, the camera model chosen in this study was Vivotek FD7131 designed for indoor surveillance. It is a wide-angle network camera having a maximum resolution of 640 × 480 dpi and up to 30 fps image-capture capability. The maximum horizontal angle of view provided by the technical specifications is 105.1°, whereas the maximum vertical one is 77.4°.

The images acquired from the chosen camera model were subject to an image calibration process aiming at removing lens radial distortion, which is the source of the ‘fish-eye’ or ‘barrel’ effect, and lens tangential distortion, produced by manufacturing defects resulting from lens which are not exactly parallel to the plane of the scene. In this study, the calibration was performed by applying the calibration matrix obtained by using images of a known object point array, i.e., a black and white checkerboard pattern [Zhang 2000].

The image calibration process applied to camera acquisitions produced a decrease of the view angles of about 31%. Therefore, the final horizontal view angle was about 72.5° and the vertical one about 53.4°.

The video recordings were carried out by using a desktop personal computer connected to the camera local network. The acquisition frame rate of the camera was set to 1 fps and video recordings were acquired by using a specific software supplied with the chosen camera model. Furthermore, an external hard disk was connected to the system to be fully employed for data transport. An ADSL connection allowed the remote control of the video-camera recordings through the Internet.

2.3 *The detection model*

2.3.1 *The training phase*

The Viola-Jones detector algorithm requires a training phase where positive and negative image samples must be provided as input. Positive and negative image samples are constituted by rectangular image sub-windows of the frames. A positive image shows the target objects to be detected, whereas a negative image does not contain such target object. Positive images should have the same aspect ratio r, i.e., the same ratio of the width to the height. Negative images must not have smaller dimensions than those of the largest selected positive images. For face detection, a good proportion between the number of selected positive images and the negative one is equal to 1:2.

In this study the selection of positive and negative image samples was carried out from the frames acquired by one of the six cameras placed above the feeding area. In detail, image samples were costituted of rectangular sub-windows of the frame which show feeding behaviour (positive image samples), or barn background (e.g., flooring of the feeding alley, feed bunk, and feeding barrier) (negative image samples).

To facilitate the selection of the image samples, a specific functionality was implemented in order to make it possible to easily extract parts of the frames from the video recordings. If required by the user, this functionality allows also the automatic rotation of the selected positive images. The compiler Visual C++ 2008 express was used, which is an integrated environment of software programming free distributed by Microsoft, allowing for the development of applications written in C++ language and the use of all the graphical components of the operating system.

The frames used to extract positive and negative image samples were selected from video recordings of two different days, characterized by different brightness conditions, i.e., a sunny day and a cloudy one of the experimental trial carried out during December 2010. This choice was made with the aim to train the classifier in different brightness conditions inside the barn. Though the camera was provided with a built-in white-light illuminator that has been activated when the environment lacked a sufficient light source, the sequence of the frames acquired during the late afternoons and the nights were not of good quality for the purpose of the present study and, thus, were discarded. Therefore, for the behaviour to be analyzed the frames used to extract the image samples were selected within the time interval between 7:00 a.m. and 4:00 p.m. for each day considered. In detail, the sampling was systematically extracted by selecting 6 frames within one hour of trial to obtain a total of 54 frames for each day considered.

For each of the two days considered, about 3-4 positive images were extracted from the selected frames. Therefore, a total number of 352 positive images contained the shape of the body of the cow obtained from the plan view of the monitored area. The value of the aspect ratio *r* of positive images was 2:1 with image widths between 180 pixels and 288 pixels, and image heights between 90 pixels and 144 pixels. These pixel ranges derived from the dimensions of the positive images that were affected by animal positions at the feeding barrier. For instance, if animals were perpendicular to the feeding barrier the minimum value of the image height was found.

A total amount of 134 negative images were obtained from the selected frames. The width of the negative images ranged between 338 pixels and 640 pixels, whereas height ranged between 220 pixels and 480 pixels.

An extended version of the Viola-Jones detector algorithm contained in the OpenCV library was used. In such a version of the algorithm the face-detection technique was later extended by Lienhart and Maydt [2002] to use diagonal features.

The training of the Viola-Jones algorithm requires the input of the number of the stages that will be constructed by using the AdaBoost algorithm [Freund and Schapire 1995] and the threshold values for false positive and true positive detection rates to be assigned in the same way for all the stages.

Each positive image must be resized in the same way for all the sample in order to calculate the related integral image. The algorithm computes the integral image also for each image sub-set of the negative images with the same dimensions of the resized positive images.

For each stage of the classifier, the algorithm applies the Haar-like features to each integral image computed for the whole number of positive images and finds a small set of features giving a positive detection rate greater than the previously assigned threshold value. Afterwards, the algorithm applies such a set of features to the integral images of the negative sub-windows to verify that the false positive rate is lower than the assigned threshold value.

The final aim of the training is to dramatically reduce the global false positive rate of the cascade yet preserving a high number of true positives, i.e., achieving a high value of the global detection rate [Viola 2004].

Following the above mentioned procedure, in this study 25 stages were input and, for each of them, the thresholds for false positive and true positive detection rates were fixed at 0.50 and 0.997, respectively. Furthermore, each positive image was resized to 30 × 15 pixels. These dimensions were chosen to limit the number of the Haar-like features to be used that increases along with the image dimensions. The higher the number of features, the higher the time required for the training of the algorithm as well as for the detection of the feeding behaviour during the algorithm execution. As an example, Viola and Jones [2001; 2004] used positive images resized to 24 × 24 pixels, involving the use of 261,600 features. Among them 6,061 were used by the classifier for detecting the human face.

2.3.2. *The testing phase*

The aim of this phase was to test the detection model for a large dataset, yet without making use of operator visual recognitions to validate the detection results.

The testing was carried out by using test images coming from the video recordings of the same camera used for the training. Starting from 150 positive images that were not used for the training, a new set of images to be tested was obtained by overlapping each positive image on a set of negative images (Fig. 1). This last was constituted by four background images captured when cows were away from the feeding area, i.e., during the milking of the morning, before and after the cleaning of the feeding alley, and before the milking of the afternoon. This choice made it possible to test the classifier in different background conditions determined by the brightness of the breeding environment and the quantities of slurry accumulation on the feeding alley. By means of a software developed for carrying out this testing phase, the position of each positive image within the negative one was determined randomly (Fig. 1), and stored in a database. The obtained images were then altered by applying a group of image processing operations, i.e., smoothing (blur, median, Gaussian), erosion and dilation. By using this procedure, a total of 3,600 test images were obtained. Among them 618 were excluded because 103 positive images were overlapped by the negative ones in unfeasible positions, i.e., in front of the pillar or between two posts of the feed barrier.

The algorithm searched the cow images within each test image and assigned the value 1 (hit) to the tested image if the cow was detected in the right position, i.e., present at the feed barrier, otherwise assigned the value 0 (missed). For each test image, the algorithm counted the hit cows and the missed ones. The hit rate (HR) was defined as the ratio between the total number of hit cows and the total number of test images, whereas the miss rate (MR) was computed as the ratio between the total number of missed cows and the total number of test images. The number of false positives contained in each tested image was computed by counting the number of the regions of the frames incorrectly detected as feeding behaviour. The false positive rate (FPR) was obtained by computing the ratio between the total number of false positives and the total number of test images.

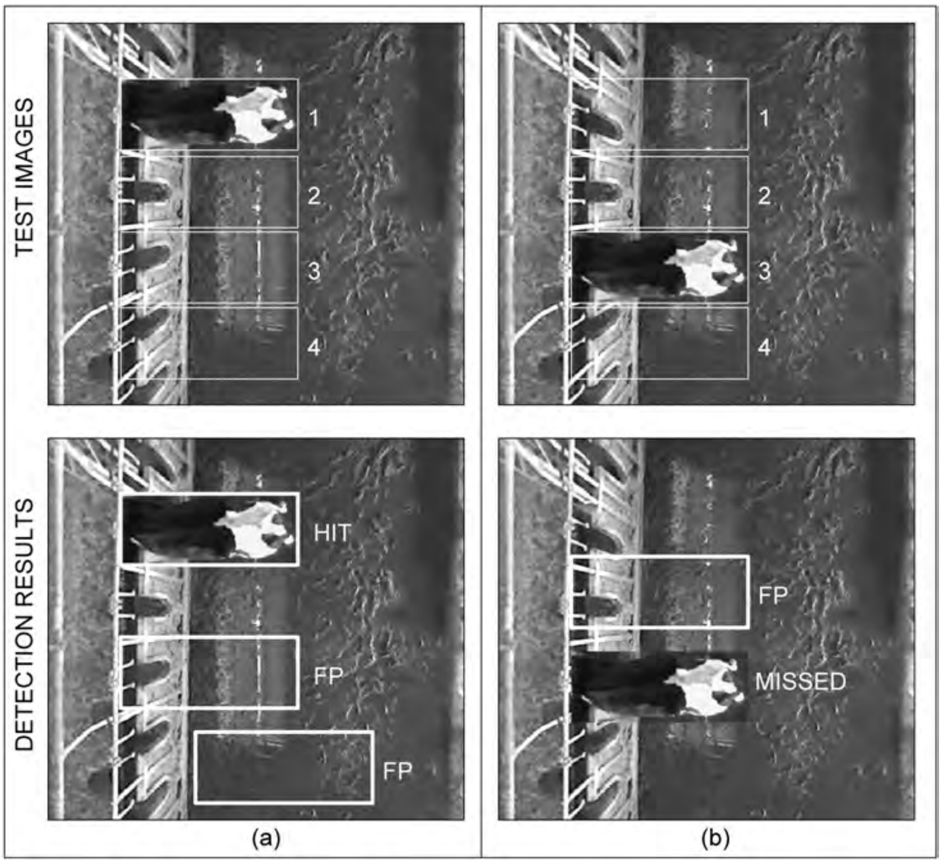

Fig. 1 - Two test images (above), obtained by overlapping one positive image to one negative one, and two examples of possible detection results (down). a) (above) The cow occupies position 1 among the 4 that are possible; (down) the cow is detected by the classifier with a sub-window placed in the right position (HIT), yet 2 false positives (FP) are detected in the frame. b) (above) The cow occupies position 3 among the 4 that are possible; (down) the classifier missed the cow (MISSED) and detected 1 false positive (FP).

2.3.3 *The validation of the detection model*

The validation was carried out by the following accuracy assessment procedure. Within the time interval between 7:00 a.m. and 4:00 p.m. of one day different from those considered in the training and testing phases, a systematic sample of frames was selected. In detail, 2 frames per minute were extracted and then checked to eliminate those not significant for the accuracy assessment, i.e., consecutive frames showing no substantial differences of their contents. If the frames were taken consecutively from the video sequence to 1 fps, the values of the accuracy would have been higher, or lower, than those obtained using the adopted sampling method, because of the consecutiveness of frames too similar to each other. Therefore, a total amount of 715 frames were used for the accuracy assessment.

By a visual examination of the selected frames an operator labelled within the database the position of each cow present at the feed barrier, i.e., about the point of intersection between the withers-to-pinbone axis and the diameter of the heart girth.

The execution of the classifier produced for each selected frame a set of sub-windows surrounding each cow that accessed the feed bunk. The system should not detect cows staying along the feeding alley without accessing the feed barrier.

Finally, the output results of the classifier were compared to those produced by the operator.

3. Results and discussion

The setting of the minimum image dimensions (30 × 15 pixels) required the use of 150,250 extended Haar features to be applied to the 352 resized positive images and to the 4.8×10^7 negative sub-windows obtained from the 134 negative images. After the training of the classifier, 110 features were selected among those available and subdivided into 25 stages.

The global detection rate reached about 98.88% and the global false positive rate of 2.84×10^{-4} %, i.e., 1 in every 3.5×10^5 negative sub-windows. The training time was 8 hours using an Intel® Core (TM) 2 Quad CPU Q6700.

The results of the training phase demonstrated that a restrained number of positive and negative images, i.e., 352 and 134 respectively, in comparison to those considered for human face detection, produced high values of the detection rate and low values of the false positive rate. Furthermore, the number of negative images required to perform the training was smaller than that of the positive ones. This result highlights that in this research the variability of the background was moderate compared to that found by Viola and Jones. Moreover, the training results revealed that though the presence of manure in the feeding alley caused the continuous changing of the background, these variations did not affect the quality of the classification.

The test and the validation of the classifier could be carried with a lower number of stages than those used during the training. This could improve the performance of the classifier in terms of speed of the detection process and number of false positives. The selection of the optimal number of the stages composing the classifier was based on the analysis of the Receiver Operator Characteristic (ROC) curve. This kind of curve relates the FPR with the HR obtained from a test carried out by varying the number of the stages from 1 to that fixed during the training phase. In this work the ROC curve of the classifier was obtained varying from 1 to 25 the number of the stages (Fig. 2). The maximum value of the FPR, obtained for the classifier composed of one stage, was equal to 1.68% corresponding to an HR of 99.24%, whereas the minimum FPR value obtained for the classifier composed of 25 stages, was equal to 0.03% corresponding to an HR of 94.01%.

The reduction of the FPR of 1.65% implied also the reduction of HR and, as a consequence, an increase of the MR of 5.23%. Though the number of the false positives could be reduced by considering a number of geometrical constraints deriving from the layout of the barn as well as some characteristics of the observed behaviour (e.g., time spent for eating), an analogous result cannot be achieved for false negatives. Therefore, for the aim of this study 6 stages are considered suitable because they produced a low number of false positives of about 0.67%, yet maintaining a good HR of about 97.85%.

The testing of the classifier, carried out on the 2,982 test images, produced the following results: 2,918 hit, 64 missed, and 20 false positives. These results show that 1 false positive was detected about every 149 analyzed frames whereas 1 false negative about every 47 analyzed frames.

The obtained classifier was used to detect the feeding behaviour within the 715 frames selected for the accuracy assessment of the classification results. The elapsed time for the processing of one frame was 80 msec using an Intel® Core (TM) 2 Quad CPU Q6700. This result demonstrates that the algorithm is suitable for real-time elaborations.

The visual recognition required about 8 hours of work by an operator. Afterwards, the information collected by the operator was compared with that produced by the execution of the algorithm. Cows present in the feeding area without eating were considered false positives, if detected from the classifier. The final results of this procedure were grouped according to the number of cows present in each analyzed frame, which ranged between 0 and 4. For each group of cows, Table 1 shows the total number of frames in which cows were detected without errors, i.e., showing the feeding behaviour in the right position, and those which contained one or two false negatives. The accuracy assessment revealed that among the 715 frames about 90.63% contained only true positives, whereas about 9.37% were affected by underestimation, i.e., contained also one or two false negatives. Among the 715 frames, only 21 false positives occurred (about 2.93 %).

By comparing these results with those obtained by the testing, it came out that the detection rate of the classification decreased by about 6%. This result was affected by the number of false negatives that was recorded for groups of two and three cows (Tab.1). In fact, the classifier missed some cows which were in close contact with each other during the feeding. A

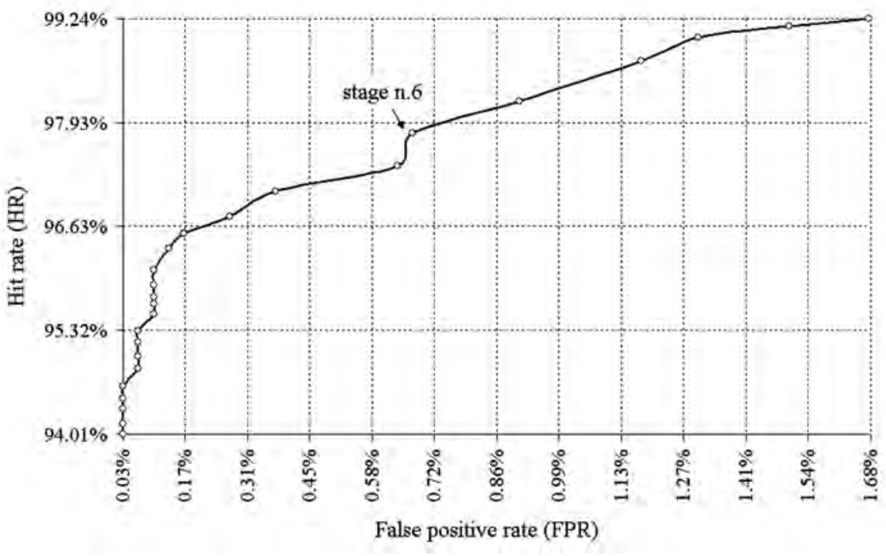

Fig. 2 - ROC curve of the classifier obtained by varying the number of the stages from 1 to 25.

Visual recognition		Automatic detection			
Number of cows	Number of frames	Containing only true positives	Containing also 1 false negative	Containing also 2 false negatives	Containing also 1 false positive
0	286	286	0	0	5
1	148	138	10	0	0
2	153	128	18	7	13
3	114	83	31	0	3
4	14	13	1	0	0
	Tot. 715	Tot. 648	Tot. 60	Tot. 7	Tot. 21
		% 90.63	% 8.39	% 0.98	% 2.93

TABLE 1 - Total number of frames in which cows were detected by the classifier without errors, frames containing 1 false negative, frames containing 2 false negatives, and frame containing 1 false positive.

better performance of the classifier could be achieved by incrementing, during the training phase, the number of positive images.

4. Conclusions

Since the proposed methodology did not require any image pre-processing step for segmenting out the animal features in the video recordings, its application overcomes some difficulties that could occur when images of the observed animals are not in contrast with the background.

As the results of the training phase demonstrated that moderate quantities of positive and negative images can produce high values of the detection rate and low values of the false positive rate, the application of the proposed methodology would not lead to an excessive amount of time for the selection of the image samples.

During both the testing and the validation phases, the obtained classifier gave good classification results also when significant brightness and background variations occurred in the sequence of the analyzed images.

Finally, since the validation phase demonstrated that the real-time execution of the classifier is feasible, the application of the methodology would avoid the onerous activity of video-recording storage.

Acknowledgements

This research was financed by the Italian Ministry of University and Research within the PRIN 2008 project on "Innovative techniques for behaviour analysis of dairy cows in barns equipped with automatic milking systems" - Local coordinator of the research unit: prof. Giovanni Cascone.

References

Cangar Ö., Leroy T., Guarino M., Vranken E., Fallon R., Lenehan J., Mee J., Berckmans D., Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis. Computers and electronics in agriculture, 2008, 64, 53-60.

Carreira X.C., Fernández M.E., Mariño R.A., Indices for estimation of dairy free-stall occupancy. Applied Animal Behaviour Science, 2009, 119, 23-31.

Corkery G.P., Gonzales-Barron U.A., Butler F., Mc Donnell K., Ward S., A preliminary investigation on face recognition as a biometric identifier of sheep. Transaction on the ASABE, 2007, 50 (1), 313-320.

DeVries J.T., von Keyserlingk M.A.G., Weary D.M., Beauchemin K.A., Measuring the Feeding Behavior of Lactating Dairy Cows in Early to Peak Lactation. Journal of Dairy Science, 2003, 86, 3354-3361.

Freund Y., Schapire R.E., A decision-theoretic generalization of on-line learning and an application to boosting. In Computational Learning Theory, Springer-Verlag, 1995, 23-37.

Lienhart R., Maydit J., An extended set of haar-like features for rapid object detection. In: Proceedings of the IEEE International Conference on Image Processing, Rochester, New York, 2002, 1, 900-903.

Mattachini G., Riva E., Provolo G., The lying and standing activity indices of dairy cows in free-stall housing. Applied Animal Behaviour Science, 2011, 129, 18-27.

Müller R., Schrader L., A new method to measure behavioural activity levels in dairy cows. Applied Animal Behaviour Science, 2003, 83, 247-258.

Provolo G., Riva E., One year study of lying and standing behaviour of dairy cows in a fre-stall barn in Italy. Journal of Agricultural Engineering, 2009, 2, 22-33.

Shao B., Xin H., Harmon J.D., Comparison of image feature extraction for classification of swine thermal comfort behaviour. Computers and electronics in agriculture, 1998, 19, 223-232.

Shao B., Xin H., A real-time computer vision assessment and control of thermal comfort for group-housed pigs. Computers and electronics in agriculture, 2008, 62, 15-21.

Viola P., Jones M., Rapid object detection using a boosted cascade of simple features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, 1, 511-518.

Viola P., Jones M., Robust real-time face detection. International Journal of Computer Vision, 2004, 57(2), 137-154.

Wilson S.C., Spectral analysis of feeding and lying behavior of cattle kept under different feedlot conditions. Journal of applied animal welfare science, 2005, 8(1), 13-24.

Zhang Z., A flexible new technique for camera calibration. IEEE Trans. on Pattern analysis and machine intelligence, 2000, 22(11), 1330-1334.

Summary

The objective of this study was to investigate the applicability of the Viola-Jones algorithm for continuous detection of the feeding behaviour of dairy cows housed in an open free-stall barn. A methodology was proposed in order to train, test and validate the classifier. A lower number of positive and negative images than those used by Viola and Jones were required during the training. The testing produced the following results: hit rate of about 97.85%, missed rate of about 2.15%, and false positive rate of about 0.67%. The validation was carried out by an accuracy assessment procedure which required the time-consuming work of an operator who labelled the true position of the cows within the barn and their behaviours. The accuracy assessment revealed that among the 715 frames about 90.63% contained only true positives, whereas about 9.37% were affected by underestimation, i.e., contained also one or two false negatives. False positives occurred only in 2.93% of the analyzed frames. Though a moderate mismatch between the testing and the validation performances was registered, the results obtained revealed the adequacy of the Viola-Jones algorithm for detecting the feeding behaviour of dairy cows housed in open free-stall barns.

This, in turn, opens up opportunities for an automatic analysis of cow behaviour.

Keywords: precision livestock farming, dairy farming, vision techniques, animal detection.