DOTTORATO DI RICERCA INTERNAZIONALE
in
INGEGNERIA AGRARIA
XXIV CICLO

Umberto Anguzza

A METHOD TO DEVELOP A COMPUTER-VISION BASED SYSTEM FOR THE AUTOMATIC DAIRY COW IDENTIFICATION AND BEHAVIOUR DETECTION IN FREE STALL BARNS

Tesi per il conseguimento del titolo di Dottore di Ricerca

Tutor: Ch.mo Prof. Claudia Arcidiacono

Coordinatore: Ch.mo Prof. Claudia Arcidiacono

UNIVERSITÀ DEGLI STUDI DI CATANIA
Dipartimento di Gestione dei Sistemi Agroalimentari e Ambientali
Sezione Costruzioni e Territorio
Catania, dicembre 2012

To my wife Simona, whose unconditional love, cooperation, knowledge, availability, patience and support made the thesis work possible.

To my daughter Chiara, who will complete her four years next month. I spent little time with her, especially in these recent months. My thoughts to Chiara allowed me to find the energy to complete this dissertation.

CONTENTS

A METHOD TO DEVELOP A COMPUTER-VISION BASED SYSTEM FOR THE AUTOMATIC DAIRY COW IDENTIFICATION AND BEHAVIOUR DETECTION IN FREE STALL BARNS

ABSTRACT

In this thesis, a method to develop a computer-vision based system (CVBS) for the automatic dairy cow identification and behaviour detection in free stall barns is proposed. Two different methodologies based on digital image processing were proposed in order to achieve dairy cow identification and behaviour detection, respectively. Suitable algorithms among that used in computer vision science were chosen and adapted to the specific characteristics of the breeding environment under study.

The trial was carried out during the years 2011 and 2012 in a dairy cow free-stall barn located in the municipality of Vittoria in the province of Ragusa. A multi-camera video-recording system was designed in order to obtain sequences of panoramic top-view images coming from the multi-camera video-recording system. The two methodologies proposed in order to achieve dairy cow identification and behaviour detection, were implemented in a software component of the CVBS and tested.

Finally, the CVBS was validated by comparing the detection and identification results with those generated by an operator through visual recognition of cows in sequences of panoramic top-view images. This comparison allowed the computation of accuracy indices. The detection of the dairy cow behavioural activities in the barn provided a Cow Detection Percentage (CDP) index greater than 86% and a Quality Percentage (QP) index greater than 75%. With regard to cow identification the CVBS provided a CDP > 90% and a QP > 85%.

Keywords: cow behavioural activity, precision livestock farming, dairy farming, object recognition, computer vision techniques, digital images.

1 INTRODUCTION

1.1 Preface

The optimization of the overall life quality of animals housed in intensive farming systems, both in terms of comfort and productivity, involves the analysis of several responses that an animal produces to adapt to the environment, such as behavioural modifications and physiological mechanisms that could affect animal health status and growth.

As concerns the intensive farming of dairy cows, animal health can be investigated by the analysis of four categories of indicators: behavioural indicators, physiological indicators, pathological indicators, and productivity indicators (Smidt, 1983).

In research works aiming at assessing cow welfare, a growing attention to the analysis of cow behaviour has been paid by the scientific community since it is considered an essential indicator for both the evaluation of cow health status in different types of housing systems and the development of farming systems that make it possible to combine high production levels and animal welfare.

In literature, the most frequently analyzed behavioural activities of dairy cows are the following: ‘feeding’, which refers to the cow standing still in the feeding alley with the head in the feed barrier; ‘standing’, which describes the standing still in the alleys and the walking; ‘lying’, which is related to all the possible lying postures of the cow in the stall; and ‘perching’, which is associated to a standing still behaviour characterized by the cow hind limbs placed in the service alley and the cow forelimbs placed in the stall.

In some research studies the lying behaviour of dairy cows was analysed to determine its effect on the level of milk production and the foetal development during the pregnancy (Nishida, Hosoda, Matsuyama, & Ishida, 2004; Rulquin & Caudal, 1992). The daily incidence of lying and standing behaviours was also examined both to improve oestrus detection (Firk, Stamer, Junge, & Krieter, 2002) and perform early diagnosis of lameness (Pastell, Tiusanen, Hakojärvi, & Hänninen, 2009). The monitoring of feeding behaviour was carried out with the aim to both optimize intake under different feeding managements (DeVries, Von Keyserlingk, Weary, & Beauchemin, 2003b; Halachmi et al., 1998; O'Driscoll, Boyle, & Hanlon, 2009) and improve barn building characteristics in terms of layouts and building materials (DeVries & Von Keyserlingk, 2006; Fregonesi, Tucker, Weary, Flower, & Vittie, 2004; Fregonesi, Veira, Von Keyserlingk, & Weary, 2007).

Traditionally, dairy cow behaviour is assessed by skilled operators, such as veterinaries, who carry out the direct observation of the herd in field or inside animal houses. These operators also verify the adequacy of the breeding environment, which affects dairy cow behaviour, apply protocols to determine the presence of diseases, and control production data. When information is gathered in *situ* the use of check lists is generally adopted to investigate animal behaviours.

However, studies based on the direct observation of animal behaviour could suffer from a common problem related to the presence of the observer, which may affect and thus modify animal behaviour. Moreover, the direct observation usually involves many hours of specialized operators' work used for the visual recognition of behavioural activities and, if more than one operator carries out the detection work, it could be also subject to discordant interpretation.

Another technique involves the observation of animal behaviour by means of the visual analysis of digital images obtained from time-lapse video recordings. This technique allows the operator to perform a more adequate planning of the monitoring activities. Furthermore, video recording facilitates the collection of more frequent data over longer time periods and assures a safe and secure record for future analysis.

With regard to dairy cows, the visual recognition of cow behaviours from continuous recordings provides precise data about the time and the duration of the considered behaviour. This data can be used for carrying out statistical analyses and computing behavioural indices (Provolo & Riva, 2009), such as the cow lying index (CLI), the cow standing index (CSI), and the cow feeding index (CFI). The computation of these indices requires only the knowledge of cow behaviour whereas cow identification and tracking are needed when studying behavioural patterns and activities with the aim to perform an estimation of missed operations (e.g., milking), and to automatically observe individual animals.

Concerning large breeding environment, the visual recognition of cow behaviour in digital images coming from time-lapse video recordings can be carried out by using one or more cameras to frame all the functional areas where cow behaviours are likely to occur. Among the possible views of a barn that can be obtained from multi-camera systems, a panoramic top view of the barn is of crucial importance to detect animal positions within the barn as well as to obtain the real shape and dimensions of the body of the animals.

To obtain a panoramic top-view image of a large breeding environment, the careful design and installation of a multi-camera video-recording system is required. However, methods for the automatic detection of cows in panoramic top-view images acquired by means of multi-camera systems have not been applied yet.

The automation of both the visual recognition of cow behaviours in digital images and the individual cow identification may represent a suitable alternative to the traditional observation techniques adopted to investigate cow behaviours since it is comparatively less expensive and not invasive for animals. The design and the implementation of a computer-vision based system (CVBS) is recommended to automate cow behaviour detection and cow identification.

Though in literature several systems have been proposed in order to detect cow behaviour in digital images, few research work has been aimed at identifying each animal of the herd by using image processing techniques. Whereas, animal identification is generally obtained by using several sensors such as wireless pedometers, or collars equipped with RFID tags. The integration of these sensors within a CVBS implemented in order to detect animal behaviour would increase the overall cost of the recognition activity. Furthermore, even though the sensors themselves are apparently harmless, they may interfere with animal behaviour.

An inexpensive solution for the identification of animals is marking them with natural paints or hair dye. However, in literature this animal identification technique was not used within CVBS implemented in order to detect animal behaviour.

1.2 Objective of the study

On the basis of the issues arising from the preface, the overall objective of the thesis work is to put forward a method to develop a computer-vision based system (CVBS) for the automatic dairy cow identification and behaviour detection in free stall barns. Two different methodologies based on digital image processing are proposed in order to achieve cow behaviour detection and cow identification, respectively. Suitable algorithms among that used in computer vision science are chosen and adapted to the specific characteristics of the breeding environment under study. The method is applied to a case study, i.e., the detection of cow behavioural activities and the identification of each cow bred in a free-stall barn located in the municipality of Vittoria in the province of Ragusa.

The use of the proposed CVBS could avoid the interference in the cow behaviour due to the presence of the operator in the barn and could reduce the time-consuming operations needed to perform the visual examination of digital images in order to recognize cow behaviours and perform cow identification.

A sub-objective of the study is the design of a multi-camera video-recording system. The achievement of this aim is crucial in order to obtain sequences of panoramic top-view images coming from the multi-camera video-recording system. The panoramic top-view images make it possible to relate the position of the cow in the image with the real position of the cow in the barn. The achievement of these aims is relevant to verify if irregular behaviours occur, e.g., idle standing in the stall, perching, alternate occupancy of the stall, standing still for long time in an area of the barn, and lying in the alleys.

The overall objective of this thesis falls within the research boundaries of the precision livestock farming (PLF). Together with precision farming (PF), which has been assisting plant production for quite a long time, PLF has been developed more recently in the field of stock-breeding. PLF consists of monitoring, collecting and evaluation of data acquired by on-going processes related to the animals, modelling these data to select useful information, and then applying these models in real-time for monitoring and control purposes. Collection of data from animals and their environment, by innovative techniques, is followed by evaluation of the data by using knowledge-based computer models. Currently, considerable PLF research is directed toward the development and the validation of various techniques for data measuring and registration in livestock farms (monitoring feeding times, feed intake, and performance parameters, real-time analysis of sounds and images, live weight assessment, body condition scoring, etc.). The final aim of PLF is to achieve information on the main parameters of animal health, animal behaviour and animal performance on a continuous basis. As farmer routinely gather visual, auditory, and olfactory information from their animals to evaluate health, welfare and productivity, new technologies can be useful to perform this task, even with large herds, thanks to the evolution in sensors and sensing techniques (Frost et al., 1997). The advantage of these monitoring systems is that much information can be automatically collected without the stress of animal disturbance or handling (Provolo & Riva, 2009).

1.3 Work organization

The second section of this thesis contains a review of the methods used to observe animal behaviour. A number of research works are described in order to highlight the methods used to achieve the animal behaviour detection and animal identification. Both traditional methods and those based on the automated image analysis are described.

The third section of this thesis describes the materials and methods of the research. In detail, it describes the method proposed for the design of a multi-camera system that makes it possible to obtain a panoramic top-view image of large breeding environments; it gives a comprehensive look at the image processing algorithm used in the two methodologies proposed in order to achieve cow behaviour detection and cow identification, respectively. This section of the thesis, illustrates also the case study. In detail, a description of the free-stall barn is provided as well as the materials and methods used to develop the multi-camera system and the two methodologies.

The fourth and the fifth sections of this thesis describe the results of the application of the two methodologies to the case study. The potential applications of the CVBS are illustrated as well as its further improvements.

2 STATE OF THE ART

2.1 The observation of animal behaviour

The study of animal behaviour is based on the consideration that animals do not behave randomly. A species of animal that lives in natural environmental conditions is able to express behavioural activities that can be considered as "normal", i.e., characterized by movements highly coordinated in both time and space. The sequences of these movements can be patterned. Each pattern is characterized by the fact that the shape of animal body does not vary and that the time intervals, at which the specific sequence of movements is expressed, are constant. These patterned sequences of movements are the fundamental units in the study of animal behaviour and they can be measured just as surely as other physiological parameters, e.g., weight, length, and height.

Some people erroneously refer to behavioural observations as subjective, contrasting them with physiological measurements that are said to be objective. What distinguishes behaviour and physiology for some people, however, is that they believe that behavioural measurements are less reliable than physiological ones, e.g., it is less easy for two people to agree what, say, an aggressive behaviour is than what a level of a given hormone is. The variability of behaviour, both between individuals of the same species and also within the same individual at different times is often thought to make it difficult to come up with reliable, repeatable measures that can be used in the same way by different observers (Rushen, 1991).

Whatever the truth of those arguments, the fact remains that the reliability of both behavioural and physiological measures needs to be checked or validated from time to time. This is particularly true when more than one observer is engaged in a single study, but it also applies when there is just one observer whose criteria for different behaviours may slip as a study proceeds.

The precise and detailed list of all the observed behaviour patterns of a species is known as the "ethogram". A period of unstructured observations that allow for a systematic data collection should be used in order to build an ethogram. Furthermore, reading what has been already published on the specific animal species under study may also help the construction of the ethogram. Each behavioural pattern of the ethogram should be clearly defined by means of drawings, photographs and video. Whether observations are recorded on paper or video, electronic support, or other means of recording, each observation should be labelled by date, time, location, and other details that are important such as weather information, total number of animals visible, etc. If paper is used, a waterproof notebook, or data sheets on a clipboard are useful and it often helps to print out maps and sheets with slots for all the information to be recorded beforehand in order to avoid missing something. If video is used, it is still necessary to record on paper the details of when and where the video recordings are being made. By keeping good records of any actions that were done, including departures from the planned protocol is an essential part of good research.

The use of the appropriate tools for defining the ethogram should allow the sharing of the observations among other people in order to explain the outcomes of a specific research. Moreover, it should be useful to train other observers who may be assisting the researcher in the phase of data collection.

There are several kinds of questions that can be asked about animal behaviour and there are several ways in which observation can contribute to all of them. The right question determines both the kind of data to collect, for example data collected over the whole lifetime of the animals or at just selected moments in time, and the type of information to be recorded, how often, how long for, and how many animals to watch to get valid results. It will determine the required equipments, whether any breaks of the observations are periodically required, or otherwise whether continuous observations are needed. It will determine whether it is important to know the investigated animals as individuals or can treat them as herd.

2.1.1 The selective observation

To study a specific behaviour of a single animal the following activities should be required:

- to record each behavioural pattern, e.g., the movements of its head, body, and limbs;
- to record the behavioural patterns of all the other animals around it;
- to record where the animal is in relation to other animals;
- to record where the animal is in relation to the building characteristics of the breeding environment:
- to record which animals are interacting with it.

Since the development of the above mentioned activities is not possible to be achieved in practice for all the possible behaviours, a selective observation is required. In detail, it is important to define:

- the behavioural patterns to be recorded;
- the suitable length of time to spend observing the animal in relation to the chosen behavioural patterns, i.e. from seconds to weeks or years;
- the appropriate observation level, i.e. from an individual to the whole herd;
- the suitable sampling regime, i.e. what, for how long, and how often you record.

When making certain hypotheses, these should be first validated by considering the suitable selection of behavioural patterns, e.g., lying, feeding, standing. Even if observations are made by using video-recordings, videos should be watched or analyzed with some hypotheses in mind. It is not the amount of data derived from the results of automatic recording devices that makes the behavioural analysis exhaustive, but databases still have to be interrogated on the basis of selected and defined hypotheses since they can only provide answers to specific questions. For example, lying and standing behaviours should be analyzed if a hypothesis predicts that a particular species of animal is more likely to sit down in cool weather than in hotter weather whereas other behaviours could be discarded, e.g., grooming itself or plying with other animals, in order to test this hypothesis.

The choice of observing individual animal behaviour or collective behaviour of the whole group of animals, characterizes the levels of observation. In a broad sense, when individuals, through local interactions, do something together is defined as collective behaviour. Examples of this behaviour are the following: bird flocks or fish schools moving coherently in the same direction, ants feeding from the same food source, cockroaches aggregating under the same shelter, etc. (Sumpter, 2006).

To clarify the concept of level of observation, let suppose that the hypothesis to be tested predicts that animals should be found in larger groups rather than being isolated. To test this, it would be necessary to record the size of the groups to be observed, but it might not be necessary to record the behaviour of each individual animal. Conversely, if the hypothesis predicts that the animals would seek food far from the place where they were, it would be necessary to observe individually where each animal eats.

To identify animals individually, it could be done in two ways: to catch the animals before the beginning of the observations and mark them in some way, such as with wing tags, leg bands, or collars; to recognize them from characteristics that make many animals naturally individually distinct. Particular attention should be paid to the first way of identifying animals because even when the tags themselves are apparently harmless, they may interfere with animal behaviour. In fact, a series of experiments on zebras (Burley, 1988) showed that the coloured leg bands which were used to identify individual male zebra finches affected not only their attractiveness to females but also their success as fathers. The females seemed to confuse the red leg bands with the red beaks that males have naturally as a sexual signal.

After having performed the selection of behavioural patterns to be recorded (e.g., preening, feeding, standing, and lying) and levels of observation, there is the problem of how many different focal animals are needed and what sampling method to use.

As regards the problem of defining the number of animals to observe, if research involves behavioural activities that are very simple to observe, it is useful to consider all the animals in the group. If, instead, the study aims at analyzing some specific behaviour, then it might be convenient to focus the attention to one animal at a time, which in this case takes the name of Focal animal sampling, whereas when the behaviour to be analyzed involves simultaneously more than one subject, can be observed a group of animals is to be observed and it is defined as Focal group sampling (Albertini et al., 2008; Martin & Bateson, 2007). For both methods of focal sampling it is required to record all the activities performed by the selected animals over a set period of time. On the data collected during the observation period, some summary statistic, such as the mean number of times the focal animal looked up, is extracted from the record and becomes one data point in a whole series of observations. Since the focal animal is taken as representative of that whole group, a particular care in deciding which of a group of animals is chosen to be the focus of the observations should be taken. To avoid invalidating the results, the focal animal is randomly chosen since choosing the cutest or the most active could bias the results.

As regards the problem of choosing the type of measurement, i.e., how much of what observed is actually recorded, the use of continuous recording provides precise information about when the considered behaviour starts and stops. This technique provides very good quality data, known as interval measurements (Siegel & Castellan Jr., 1988), and can be used for all sorts of statistical analyses. If the animal behaves relatively slowly, as cattle tend to, it will probably be possible to carry out direct observation, by noting the times of any change in behaviour, but if the animal keeps changing behaviour rapidly, this may not be possible or track of the animal may be lost in the effort to write everything down. Therefore, video-recordings are a more simple way of obtaining information about the durations of behaviours but the additional support of a notebook and a stopwatch is still advisable.

The type of measurement called “zero/one” or “yes-no” sampling is used when a continuous record is not, in fact, absolutely necessary to reach the objectives of a research, and it may be sufficient to record less about each behaviour in order to test a hypothesis as effectively as with continuous recording. Video-recording just whether or not the behaviour occurred at all during a predetermined time period might be enough to some purposes. This type of measurement makes it much easier to write down what it is observed or recorded from a video and allows keeping up with the animal’s behaviour in real-time. In particular, the focal animal is observed for the same amount of time as for continuous sampling, but only information related to whether an animal did a behavioural activity, is written down. This is defined as categorical measurements (Siegel & Castellan Jr., 1988) which are absolutely valid but can only be used in certain types of statistical tests.

A third category of measurements, very commonly used in behavioural research, constitute a compromise between the interval measurements, which provide data with a high completeness, and the categorical measurements, which give more sparse data. This kind of measurements, called ordered or ranked measurements (Siegel & Castellan Jr., 1988), can be very useful when an animal does more of something than another, but not possible to say exactly how much more. Therefore, instead of trying to record all the individual behaviour patterns, the behaviour could simply be divided into categories. Ordered or rank measurements have to be analysed with non-parametric statistical tests, which are very easy to carry out and make few assumptions about the data.

When it is not possible to watch all animals continuously, but to watch all of them for some of the time, Scan sampling is used. This method of recording behaviour and is the easiest and quickest to do. Usually it is carried out by doing a quick scan of the whole group of animals, which would provide a snapshot of what animals were all doing at one moment in time, and, sometime later, another scan is done in order to capture another frozen moment in time. The elapsed time between two snapshot should be small to allow a realistic estimate of the percentage of time that the animals used in the various activities, but reducing as much as possible the amount of energy for observation (Mitlohner, Morrow-Tesch, Wilson, Dailey, & Mcglone, 2001).

When it is not possible to make the systematic observations ideally required by the research aims, other methods of sampling could be used. Behaviour sampling is used if the behaviour at all is observed and recorded and ad *libitum* sampling is used if anything the animal is doing at all is observed and recorded (Altmann, 1974). Both methods are particularly suitable for rare behaviour or rare species.

2.1.2 Some open issues on animal behaviour observation

Studies based on the observation of animal behaviour could suffer from a common problem related to the fact that the animals may be observing the observer, and they may be afraid of observer and run away, or they may be so interested in observer that they spend all their time investigating him, like dairy cattle often do. This problem could be solved by using a multi-camera video recording system (DeVries, Von Keyserlingk, & Weary, 2004; DeVries et al., 2003b; Kaihilahti, Suokannas, & Raussi, 2007; Mattachini, Riva, & Provolo, 2011; Overton, Sischo, Temple, & Moore, 2002; Provolo & Riva, 2009).

Another issue regards the localization of the animal, i.e. where the specific animal behaviour occurs. Tracking devices based, for instance, on radio frequency or GPS, could be used to collect huge amounts of data. A basic technique of recording where animals are with respect to their environment is to make a number of photocopies of the map of the area the animals are occupying and use one copy for each scan sample, noting the positions of animals or groups on each scan in relation to specific previously defined landmarks. Since the recording of the animal positions in each scan is a burdensome activity, it is possible to subset the map into areas and to count the numbers of animals in each of them.

Another quick and coarse but useful method of recording the location of animals with respect to each other, is to select a focal animal and then count the number of animals within one, two, or more body lengths of it. The main advantage of using the animal's own body length rather than real measurements is that the distance between animals can be measured, regardless of their actual distance from the observer. Moreover, there are no parallax problems since though the animals will appear to be smaller when they are further away, a body length will also be correspondingly smaller too. Also in video-recordings analysis this suggestion could be useful, because effects of camera distortion or the size of animals in different parts of the screen could be not considered if the apparent body length of the animal in the considered area of the screen was used.

Another issue regards the choice of the most appropriate system to record the data, e.g., a notebook and stopwatch, a small portable computer, or a video-camera.

Since a safe and secure record for future analysis is obtained, there are sometimes clear advantages in using a video-camera system and there are now various software packages available to extract information. These include Observer XT (Schmied, Waiblinger, Scharl, Leisch, & Boivin, 2008) and Jwatcher (Blumstein & Janice, 2007). Nevertheless, while there are some projects where video or other automatic recording is essential, i.e. day long observations that cannot be recorded directly, direct observation should also be considered as the simplest of recording methods which may have advantages in some situations. On the other hand, technology is also opening up new possibilities in the field of animal behaviour observation and recording methods.

2.2 Observation of dairy cow behaviour

2.2.1 Traditional methods

2.2.1.1 Check lists compiled in field

From the literature it came out that often dairy cow behaviour is assessed by skilled operators and veterinaries who carry out the direct observation of the herd in field or animal house in order to compile check lists suitable to verify the adequacy of the breeding environment, apply protocols to determine the presence of diseases (e.g., lameness), control data production and dairy cow health status. Often, daily scan sampling intervals of a few hours are applied.

The direct observation of the herd was used in a number of studies in order to assess dairy cow behaviour and welfare.

In 2001, two experiments were conducted to detect differences in animal responses between strawyard and cubicle systems (Fregonesi & Leaver, 2001). The welfare of the lactating dairy cows was assessed by means of behaviour, performance and health indicators. The direct observation of the herd was used in order to obtain the data for the computation of the indicators. In detail, measurements were made by a team of operators that, through direct observation of herd activities in the barn, recorded the investigated behaviours (i.e., lying down, ruminating on bed, and standing on passage and feeding) by using check lists. In the first trial the herd was observed for two weeks and, for each day, data were recorded by using 5-minute scan sampling techniques. In the second trial the herd was observed for four weeks and data were recorder by using the same scan sample. The activity related to direct observation involved the drawing up of 4032 and 8064 check-lists in the first and in the second trial respectively.

Check lists were also used to estimate the detrimental effects of lameness on calving-to-conception interval and hazard of dying or being culled in lactating Holstein cows (Bicalho, Vokey, & Guard, 2007). Trained veterinarians assigned a visual locomotion score (VLS) to 1799 lactating dairy cows present in 5 dairy farms. The VLS could take a 5-point scale ranging from 1 = normal, 2 = presence of a slightly asymmetric gait, 3 = the cow clearly favored 1 or more limbs (moderately lame), 4 = severely lame, to 5 = extremely lame (nonweight-bearing lame). The VLS was done every 14 days, and each cow received at least 2 scores and a maximum of 7 scores.

Firstly, the observers entered the cow identification number and the assigned VLS into a digital voice recorder and subsequently the data were entered in a spreadsheet. Successive analysis demonstrated a significant increase in days from calving to conception for cows detected as *lame* when compared with those considered *nonlame* during the first 70 days in milk (DIM).

2.2.1.2 Analysis of digital images from time-lapse video recordings

The analysis of digital images from time-lapse video-recordings represents an effective tool for studying livestock behaviours in different environmental conditions. It is relatively cheap, non-invasive and facilitates the collection of more frequent data over longer time periods (Cangar et al., 2008). In this section a review of the most relevant research on the analysis of dairy cow behaviour by analyzing digital images was carried out.

Time-lapse video recordings were used to document dairy cow behavioural patterns, examine factors affecting lying behaviour, and develop guidelines for visual assessment of free-stall usage during summer conditions in a high producing dairy (Overton et al., 2002). The Authors used a multi-camera video-recordings system constituted of four video cameras placed in a free-stall pen containing 144 stalls and 129 high producing cows. The four cameras were placed about 5 m above the pen floor to allow more complete visualization of the pen. The video recordings were recorded over a 6-day period in July 1999. Operators carried out the visual interpretation of the video recordings. In detail, each daily video recording was reviewed using 60-minute scan sampling techniques, so cows were counted as lying, standing in alley without eating, standing in free-stalls, or eating in each of the four sections of the pen. As a result, 19 hourly observations were recorded for each day of the study. Therefore, the activity related to direct observation of video recordings involved the visual interpretation of 456 frames.

The analysis of video recordings was also used to validate the data generated by GrowSafe, an electronic system designed to allow for passive monitoring of feeding behaviour of individual cows housed in a free-stall barn (DeVries et al., 2003b). Two groups of six lactating cows were monitored for 24 hours using both the GrowSafe system and time-lapse video recordings. The cows were housed in two different adjacent pens in a free-stall. One video camera was positioned approximately 6 m above the feeding alley of each experimental pen. The output from the cameras was recorded with a time-lapse video recorder and a digital video multiplexer. Cows were individually identified with symbols on both sides of their body using hair dye. Also in this experimental trial, operators carried out the visual recognition of the cows within the video-recordings. In detail, raw data were summarized for each cow and for each minute of the day. Therefore 2880 video recordings were analyzed.

Another original study (DeVries et al., 2004) made use of digital image coming from video recordings with the aim to determine whether doubling the amount of feeding space from 0.5 to 1.0 m per cow leads to increased spacing between cows at the feeder, fewer aggressive social interactions among cows, and ultimately increased feeding activity. Two adjacent pens, each having a total of 6 m of accessible feeding alley space, were observed by using a video recording system. In detail, the animals were videotaped using one video camera per pen, a time-lapse videocassette recorder, and a video multiplexer. The video cameras were located 6 m above the feeding alley, and red lights were used to facilitate recording at night. Within the two 90-minute post feeding periods, an operator labelled the inter-cow distances and the number of animals present at the feeding alley by using a 5-minute scan sampling. During the two post feeding periods, four groups of cows of 6 cows each were observed for 7 days. Therefore a total amount of 1008 frames were analyzed by the operator.

The effects of two different feed barrier systems on feeding and social behaviour of dairy cows were assessed by analysing digital images coming from video recordings (Endres, DeVries, Von Keyserlingk, & Weary, 2005). Forty-eight cows were housed in 4 pens in a free-stall barn. Feeding and social behaviour data were collected using one video camera per pen, a time-lapse videocassette recorder, and a video multiplexer. The video cameras were located at 6 m above the feed bunk of each pen. Red lights, hung adjacent to the cameras, were used to facilitate recording at night. An operator observed the analyzed cow behaviour from continuous 24-hours video recordings by using 10-minute scan sampling. The number of displacements from the feed bunk per day were collected by continuous observation of 24-hours videos for the last 3 days of recording (6 days data collection) for each treatment condition (treatment of 4 groups of cows). Therefore a total amount of 1728 frames were analyzed by the operator.

Another research (Munksgaard, Jensen, Pedersen, Hansen, & Matthews, 2005) quantified the relative priorities between lying, eating and social behaviour of dairy cows in different stages of lactation by analyzing cow responses to time constraints in two experiments. In the first experiment, the behaviour of individual cows was recorded on videotapes (time lapse, two frames per second) on three days of the treatment, when cows were in their resource pen. In the second experiment, the behaviour of individual cows was recorded on videotapes for 24 hours a day on the last four days of the experimental period. For both the experiments, an operator quantified the behaviour of all animals. The position of the head of the animal was used to assess position in the pen. Activity of the animal was classified as eating (the animal has food in the mouth and/or chews when standing in front of the feed trough), lying (body resting on floor or mattress) and other patterns (performing a behaviour other than eating or lying). Whenever a cow changed position or started doing a new activity, the time was recorded; thus both frequency and duration could be calculated. Interruptions of less than 60 s were not recorded.

The qualitative assessment of dairy cows' social behaviour on farm was assessed with regard to its inter-and intra-observer reliability and its correlation to quantitative ethogram-based assessment (Rousing & Wemelsfelder, 2006). Five farms were object of the experimental trials. The social interaction of cows around a drinker was recorded with a digital camera. The camera was mounted on a pole out of cow reach to ensure undisturbed recording of social activities. Recording took place on three successive days for 2-3 hours in the morning, starting approximately 1 hour after the morning feeding, and 2-3 hours in the afternoon before and after afternoon milking. From this video footage, 25 clips of approximately 1 minute duration were selected. This selection was designed to be a representative sample of the variation of social interactions observed at the drinkers on the different farms, including agonistic and non-agonistic interactions. The 25 video clips included 25 social events of a total of 66 cows: 14 video clips of two cows, eight video clips of three cows, two video clips of four cows and one video clip of six cows. These 25 clips were then edited on to two VHS tapes at a professional studio. Qualitative assessment of these video tapes was provided by 12 observers, five of whom were researchers of animal science, five were Ph.D. students of animal science, and two were stockmen familiar with daily routines in dairy herds. All observers had practical experience in handling cows and observing cow behaviour. These observers were gathered at the start of the study, and given detailed instructions about free choice profiling experimental procedures. Observers were divided into two groups, with each group seated in front of a wide screen TV monitor to watch the recorded video tapes. A week later this procedure was repeated by showing observers the same 25 video clips in reversed order on Tape 2. Observers were informed that the clips were the same, to avoid speculation and to encourage them to get on with the task at hand.

Carreira et. al. 2009 (Carreira, Fernández, & Mariño, 2009) estimated the variation of stall occupancy by means of an apposite indicator. Three farms were object of the experimental trials. On each farm, cow behaviour and stall use were time-lapse recorded uninterruptedly for 24 h by using a video camera connected to a timer that recorded at preset time intervals. Individual 17-second intervals were recorded, followed by a disconnection time of 3 min and 30 s. By using this system, 380 observations were recorded over 24 h. The recordings were simultaneous in every three farms and were made during winter months. To analyze stall occupancy, the number of observations in which stalls were occupied (over the 380 observations) was quantified.

In a recent study (Mattachini et al., 2011) the values of different behavioural indices at different scan-sampling frequencies were compared in order to evaluate the different methods of data aggregation that are used to obtain daily behavioural indices. The lying, standing, feeding and drinking behaviours of 69 cows in a free-stall barn were recorded over 7 days using continuous video recording. Two black-and-white closed-circuit video cameras were installed in the barn. The two cameras were placed about 5 m above the pen floor to allow for the complete visualization of the pen. A 7 day video sequence pattern over a long-term (one year) recording period was used. An operator carried out the visual analysis of the video recordings and counted the number of dairy cows occupied in different behavioural activities (i.e., eating, lying, and standing). Standing was considered to be an upright posture (i.e., motionless or walking), while the lying category included only cows that were observed in total lateral or sternal

Table 1 - Review of the most relevant research on the analysis of dairy cow behaviour by analyzing digital images

Authors	Observed behaviour	Scan sampling	Observation period	Human resources	N. of cameras	N. of observed frames	N. of check lists compiled in field
Fregonesi & Leaver, 2001	Lying down, ruminating on bed, and standing on passage and feeding	5-min	6-weeks	A team of observers	-	-	12096
Overton et al., 2002	Lying, standing in alley without eating, standing in free stalls, eating	60-min	6-days	A team of observers	4	456	-
DeVries et al., 2003b	Feeding	1-min	24-hours	Not specified	2	2880	-
DeVries et al., 2004	Feeding	5-min	2×90-minutes (post feeding period)	One operator	2	1080	-
Endres et al., 2005	Feeding and social interaction	10-min	3-days	One operator	4	1728	-
Munksgaard et al., 2005	Lying, eating and social behaviour	2-sec	3-days	One operator	2	Not specified	
Rousing et al., 2006	social behaviour	1-sec	3-days	A team of observer	5	25 video clips of 1 min duration	-
Carreira et al., 2009	Lying, feeding, and standing	120-sec	3-days	One operator	2	1140 video clips of 17 sec duration	
Mattachini et al., 2011	Lying, standing, feeding and drinking	10-min	7-days	One operator	2	1848	-

recumbency within the confines of a stall. Eating was defined as actively ingesting feed or water, or standing within 0.6 m of the feed bunk and oriented toward the feed. Behavioural activities were analysed at different scan intervals of 10, 20, 30, 60 and 120 min. The entire video observation period covered 154 hourly time periods and, therefore 1848 frames were examined.

2.2.2 Automated image analysis-based methods

Since the visual recognition of cow behaviour is generally time consuming and, if more than one operator carries out the detection work, it could be also subject to discordant interpretation (Müller & Schrader, 2003), a number of studies proposed different methods to automate image analysis of animals in their breeding environment.

In 2008 a study was carried out in order to develop a fully automatic image analysis system to identify some locomotion and posture behaviours of cows prior to calving in a continuous and automated way (Cangar et al., 2008). In the research, eight individual cows representing a range of calving events from normal to difficult were selected for analysis. In order to identify posture and locomotion behaviour patterns, five individual pens with a straw-bedded surface of 4.6 m × 3.3 m were instrumented with cameras and recording equipment. Two cameras were used on each pen. The first one was an overhead camera which generated a top view of the animal. The images were used to develop the automatic real-time posture and locomotion monitoring tool based on an active shape model (Cootes, Taylor, Cooper, & Graham, 1995). The second one was a side view camera which generated a side view of the animal as seen by the stockperson. These video surveillances were used for visual interpretation on a computer screen by ethologists. The top view camera was placed 5-m high, above the centre of each pen. A group of operators observed the side views generate by the five side view cameras. These views were used for visual interpretation on a computer screen by ethologists. They labelled the images of cows approaching parturition at 10-second intervals over the last 24 hours prior to calving. Particular behaviours such as position in pen, orientation, lying or standing, type of lying, eating or drinking and calving details were recorded. However the above experiments were done in pig chambers in a research laboratory. In commercial livestock houses, image analysis for behaviour classification becomes more complicated.

A number of studies were carried out in order to detect lameness in dairy cows. A first study was carried out in 2008 with the aim to build an automatic system for continuous on-farm detection and prediction of lameness in the farm by using vision techniques (Song et al., 2008). This research proved that vision techniques have great potential to be used for continuous quantification of lameness in cows. A digital camera was fixed on a tripod 6 m far from the side of a corridor to record on video the entire body of each cow and its movement. The locomotion of all the 15 cows was scored by four trained observers. They scored cows individually as cows passed through the alley during the image acquisition.

Side-view images were recorded when cows passed an experimental set-up freely. Digital image processing such as background subtraction, binary image operations, calibration and hoof separation, were executed to obtain the trackway information containing hoof location. The accuracy of automatically captured results was checked by comparing with the output from manually labeled hoof locations. The above experiments were done in a research farm.

A further study (Poursaberi, Bahr, Pluk, Van Nuffel, & Berckmans, 2010) aimed at developing an automatic real-time algorithm suitable to on farm detection of lameness in dairy cattle. In particular, back posture analysis as a potential variable for lameness detection was investigated. Video-recordings data of 28 lactating Holstein cows were acquired on farm by a camera located 1.5 m high above the ground, centred and 8 m away from a concrete corridor (1.2-m wide and 6-m long) which takes from the barn to the pasture ground. The video-recordings were taken during the scoring procedure carried out by several observers. Scores were given when one or more of the following "lameness indicators" derived from literature were observed: tenderness, arched back, reduced speed, irregular gait in time or place, reduced tracking up, increased abduction and head bobbing. Observers gave score "1" when the cow did not show any of the 'lameness indicators', score "2" when the presence of one "lameness indicator" was observed, and score "3" when a severe "lameness indicator" or multiple "lameness indicators" were found.

Finally, a recent research (Pluk et al., 2012) described a synchronized measurement system, useful for lameness detection in dairy cattle, which combines image and pressure data to automatically record the angle of the metacarpus and metatarsus bones of the cow with respect to a vertical line.

A pressure-sensitive mat, having an active surface 0.61 m wide and 4.88 m long, was used to record the timing and position of hoof placement and release of each cow. The video-recordings were acquired by a camera installed at 2.5 m above the ground and at a distance of 3.5 m from the pressure mat which made it possible to cover the whole measurement area. The pressure mat data and the camera images were synchronized by using the timing information, stored in text files. The camera image, together with the position information, allowed automatic computation of the touch angle by using image processing.

A trained observer visually scored locomotion of the cows from the video-recordings. In two weeks of experiments carried out in September and October 2009, 400 measurements were considered. In the group analysis, the kinematic data of the sound cows were compared with the data from the groups of cows with a higher degree of lameness to assess differences in gait.

2.3 ICT applications for animal localization and identification

The research works reported in the previous section allowed cow localization in the breeding environment by using vision techniques. The identification, or more specifically automatic or machine-readable identification, is the key to the realization of effective precision livestock farming (Banhazi et al., 2012). The identification of the cow within the herd besides allowing the tracking of the animal, which is useful to study particular behavioural activities of the animal individually in comparison to those of the whole herd, may be utilized when the isolation of individual animals is required due to the risk of infections or the welfare of each animal have to be assessed taking into account the time spent in a number of specific behaviours.

Usually in outdoor the position of animals on wild animal habitats is obtained by developing systems based on the GPS technology, e.g. GPS-collar receivers (Barbari et al., 2006). This technology allows location data with a precision of about 1 m but it is not feasible for indoor applications.

The increasing improvement of wireless technologies has favoured the development of automated localization systems utilized to detect and track motion of animals inside farm buildings. A number of studies proposed different methods to automate identification of animals in their indoor breeding environment.

In 2007 a study (Gygaxa, Neisen, & Bollhalder, 2007) was carried out in order to set-up a local position measurement system based on radar technology to be used to track cows and analyze how they use the different areas of the barn. In addition, authors tested the system suitability for monitoring and quantifying social interactions. A data set of measurements at fixed positions and on dynamic circular showed that estimates of the location of a transponder were obtained with an error within 0.5 m. To validate the automatic positioning system was used to identify dairy cows at the feed rack. Data recorded by the LPM were compared to those collected by operator's works which identified the dairy cows inside digital images of the feed rack acquired for 9 hours with a sample rate of 10 min. Results of comparison showed that animal positioning can reliably be obtained with a precision of about 0.5 m.

Another research (Huhtala, Suhonen, Mäkelä, Hakojärvi, & Ahokas, 2007) aimed at developing an automatic tracking system based on WLAN suitable for cow tracking inside a building. The system was installed in a cowshed with a special section for 10 cows milked with a milking robot. The tags were installed first in the precise places to get accurate coordinates and subsequently, for the validation of the system, the tags were placed on top in a cow's neckband and also on a cow's back with a special band. Data recorded by the tracking system were compared to those collected by operator's works which identified the dairy cows inside digital images provided by a web camera installed above the room. In undisturbed conditions (no moving cows, clear sights between antennas and tags, etc.) the result of the manufacturer's analysis was 30% inside 65 cm, 70% inside 100 cm and 90% inside 200 cm. Instead when a tag was fastened on the cow and all the 10 cows were in the section, the results were not so good; especially when a cow was lying, the stability was very poor.

In 2009 a number of experiments (Simonini, 2009) conducted in different breeding environments have shown that active RFID systems based on the use of position markers can detect the presence of animals equipped with active tags within areas bounded by inductive loops. However, from these experiences, a number of disadvantages, mainly attributable to the cost of the position markers and the complexity of the installation of its inductive loop in the breeding environment, have emerged.

In a later search carried out in the 2011 (Porto et al., 2012) an automatic detection system (ADS) was developed. The ADS defined a configuration of an active RFID system alternative to that making use of position markers for the automatic detection of tags. The ADS was initially tested in the laboratory and then applied to a group of pigs housed in a pen consisting of a built and completely roofed resting area and an enclosed open-air area. Reference tags were properly placed in the pen and a specifically developed software was executed to elaborate RFID data. The work highlighted the possibility of using active RFID tags to detect the pigs resting in the indoor area, thus avoiding the cost of the position markers and the onerousness of the induction loops installation.

3 MATERIALS AND METHODS

With regards to the automatic detection of dairy cows housed in free-stall barns, the reliability of the image processing methods described in section 2.2.2 could be threatened by a number of conditions, i.e., clean and dirty straw bedding; dry and wet sand bedding; rubber mats; rubber flooring; slurry in the alleys; higher variability of floor brightness near the openings; and higher sunlight reflection of wet floors.

Therefore, the need to adapt the described methods and to assess the effectiveness of their results arises when coping with these specific conditions.

Another issue is the need to obtain a panoramic top-view image of the free-stall barn under study. This is crucial in order to detect animal positions within the barn as well as to obtain the real shape and dimensions of the body of the cows. To obtain such a panoramic top-view image, a careful design of a multi-camera video-recording system is required because of the large dimensions of the breeding environment to be monitored. A literature review revealed that methods for the detection of animals in panoramic top-view images acquired by means of multi-camera systems have not been yet applied to the field of precision livestock farming.

As a consequence of the above considerations, the present study puts forward a novel approach for the design of a computer vision-based system (CVBS) for:

- The automatic detection of a number of cow behavioural activities in free stall barns (objective 1);
- The automatic identification of dairy cows in free-stall barns (objective 2).

To acquire the images which are utilized in the CVBS, a method for the design of the multi-camera video-recording system is proposed and described in section 3.1.

Concerning the objective 1, a first methodology proposed in the present research aimed at detecting the following cow behavioural activities: feeding, lying, walking/standing still, and perching. With regard to some of these, more precise information is required for the comprehension of the proposed methodology. In this work, the behavioural activity feeding is referred, as in previous studies (DeVries, Von Keyserlingk, Weary, & Beauchemin, 2003a; Wilson, 2005), only to animal having its head through the feed barrier. The behavioural activity lying is referred to cow resting in the stall in one of its natural postures, i.e., long, short, narrow or wide.

The selection of these behavioural activities among all those possible is justified by the growing attention for their analysis as it is considered that their modifications, caused by social and physical problems as consequence of the breeding environment, could be associated with changes in the health status and reproductive efficiency of dairy cows. Some studies analysed the permanence of dairy cows lying down in the stalls because it affects the level of milk production and the foetal development during the pregnancy (Nishida et al., 2004; Rulquin & Caudal, 1992). Other studies focused on the daily incidence of lying and standing behaviours for oestrus detection (Firk et al., 2002) and early diagnosis of lameness (Pastell et al., 2009). It has been observed (Galindo, Broom, & Jackson, 2000) that standing still and perching predispose cows to lameness. The detection of these behavioural activities along with the analysis of the walking posture (Maertens et al., 2011), could be applied to obtain an early diagnosis of lameness.

Other research focused on the monitoring and analysis of feeding behaviour with the aim to optimize intake under different feeding managements (DeVries et al., 2003b; Halachmi et al., 1998; O'Driscoll et al., 2009) and some other studies analyzed the influence of barn building characteristics on the social and feeding behaviour of dairy cows (DeVries & Von Keyserlingk, 2006; Fregonesi et al., 2004; Fregonesi et al., 2007).

With regard to the objective 1, in section 3.2 the main characteristics of the cow detection methodology, that includes an algorithm originally proposed by Viola and Jones (Viola & Jones, 2001, 2004) for the human face detection and applied only in few cases for the animal detection (Burghardt, 2004; Burghardt & Càlìc, 2006), is described. From literature, it resulted 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.

Concerning the objective 2, an improvement of the CVBS functionalities was achieved by putting forward a second methodology for the identification and consequent positioning of each cow in the functional areas of the barn. Such improvement is required when:

- models of animal behaviours must be developed on the basis of continuous behavioural observations;
- the isolation of individual animals is required due to risk of infection;
- the welfare of each animal have to be assessed taking into account the time spent in a number of specific behaviours (Huhtala et al., 2007).

Though several systems could be used for animal identification, e.g., RFID tags and position markers (Barbari, Conti, & Simonini, 2008) and wireless technology (Huhtala et al., 2007), their integration within the proposed CVBS is not recommended because of the increasing of the overall cost.

Therefore, section 3.3 describes the main characteristics of an identification methodology based on the contours extraction in digital images and the normalized product scalar method for contours matching.

3.1 The multi-camera video-recording system

Among the possible views obtainable from a video-recording system, those providing plan views of the barn are the most suitable for the recognition of the cow behavioural activities analysed in this study. In particular, plan views of the barn make it possible to distinguish each cow from the others and to determine the real position of each cow within the barn. These potentials are relevant to verify if irregular behaviours occur, e.g., idle standing in the stall, perching, alternate occupancy of the stall, standing still for long time in an area of the barn, and lying in the alleys.

To obtain a broad coverage of the barn from above, a multi-camera video-recording system has to be designed. This kind of system must provide synchronized and rectified camera images and also panoramic rectified top-view images of the barn. Both rectified plan views are needed to obtain real dimensions of cows, physical spaces, and equipments. This characteristic of rectified plan views allows the selection of homogeneous training samples to be used by the classifiers in terms of cow body proportions showed in the video recordings.

3.1.1 Computation of the number of cameras to be installed in the barn

The exact number of cameras must be established by following the steps described below:

a) *Direct metric survey to produce the plan and two or more sections of the barn*. This phase is required to determine the installation height of the cameras above the floor of the barn (h_{cam}) as well as the height above the floor of the foreground plan (h_{forg}) that is the height of the animal body, measured at the withers, when it is standing (h'_{forg}) or lying (h''_{forg}) (Figure 1).

b) *Camera model selection*. The difficulty of installation that could occur, primarily due to the geometric and dimensional characteristics of the barn, as well as the need to contain the costs of the CVBS, favour the selection of network camera models powered on Ethernet, that does not require power

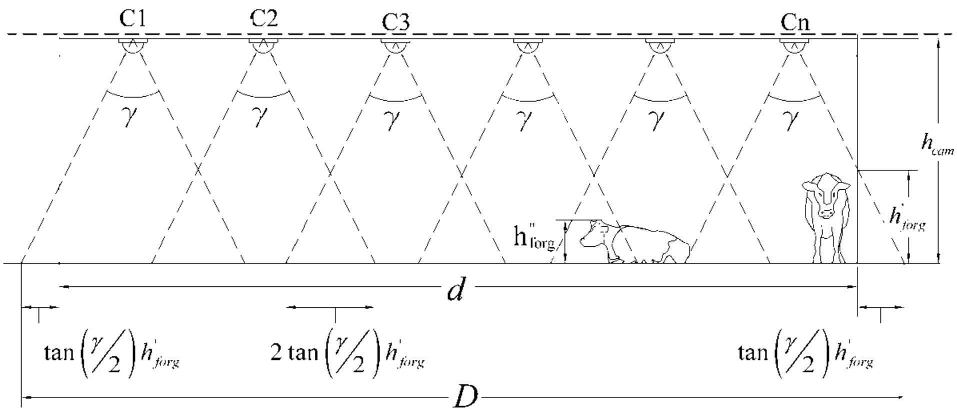


Figure 1 – A schematic representation of the section of the functional areas of barn to be monitored by means of a multi-camera system.

supply and prevent the barn from being subject to the installation of power cables. Moreover, the selected camera model has to be equipped with the hypertext transfer protocol (HTTP) based interface which provides the functionality suitable to request a single image (snapshot), to control camera functions and to get and set values of the camera parameters. The HTTP-based interface is required to obtain camera image synchronization, which will be described in the next section.

c) *Camera image calibration.* The images acquired from the cameras have to be subject to an image calibration process to remove lens radial distortion, which is the source of the 'fish-eye' or 'barrel' effect, and lens tangential distortion, produced by manufacturing defects consisting in lens not exactly parallel to the plane of the scene. For each camera, calibration is carried by using the Zhang's calibration method (Z. Zhang, 2000) that requires the use of a planar chessboard to be placed on planes with different inclination in front of each camera to be calibrated. In details the method calculates for each camera the distortion matrix and the camera matrix which characterize the process of transformation between the filmed reality and the images generated by the camera. A set of equations are defined by matching the known Cartesian coordinates of square corners in a chessboard, with the corresponding coordinates of points automatically identified in a distorted image. The solution of the defined equation are collected to build the required distortion matrix and the camera matrix.

e) *Evaluation of the maximum horizontal and the maximum vertical view angles after the calibration process*. This phase is required to establish the exact dimensions of the viewable region (π) that will be acquired by the cameras. Camera view angles have been evaluated by means of laboratory tests. The horizontal view angle (α) is the angle defined by the midpoint of left edge of the viewable region, the camera location, and the midpoint of the right edge (Figure 2). Likewise, the vertical view angle (β) is the angle defined by the midpoint of the top edge, the camera location, and the midpoint of the bottom edge.

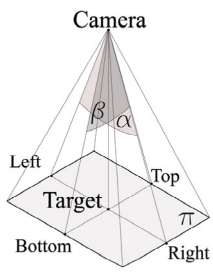


Figure 2 -Representations of the horizontal and vertical view angles of a camera.

To monitor a rectangular plan area of length l and width w, the total number N of cameras to be installed in the direction of l, or in the direction of w, can be determined by computing the value of n as:

$$n = \frac{d}{2\tan\left(\frac{\gamma}{2}\right)\left(h_{cam} - h_{forg}\right)} \qquad (1)$$

where d is equal to the length l or to the width w, γ is equal to the view angle α or to the view angle β. The number N is equal to the next whole integer of n. Once fixed the value of d, the selection of the γ value should minimize the number N.

The resulting length, or width, D of the rectangular plan area monitored by the N cameras computed by means of the equation (1) is greater than d. This makes it possible to obtain the rectified plan view of objects having height equal to h_{forg} and located near the perimeter of the area.

3.1.2 Image synchronization and mosaicing

Video acquisition from the multi-camera system must be synchronized because the N camera images characterized by the same acquisition time must be used to compose a panoramic top-view image of the barn. If camera images are not synchronized among them the image matching of the cow activities in the overlapping region of two adjacent camera images would not be assured.

By making use of the HTTP based interface, an algorithm that allows for the synchronization of snapshots coming from the cameras must be designed and implemented in a software tool. Firstly the algorithm has to make an asynchronous request to each camera to download one snapshot, and then each camera web server returns the most up-to-date snapshot in JPEG format. The algorithm must wait until all the snapshots are available before making the subsequent request, and thus the accumulation of delay times in the video sequences is avoided.

Image mosaicing refers to the combination of two or more camera images into a single composite one. This phase is required to detect cows which leave one camera scene and enter another one. In this way one single frame which will be the input of the algorithm used to detect cow behavioural activities is obtained, and the selection of the samples that will be used for the training of the classifiers is facilitated. Many algorithms today are able to take overlapping regions of camera scenes and autostitch them together to create a panoramic top-view image. However, such algorithms have to satisfy a number of requirements to obtain good results (Mills & Dudek, 2009): limited camera translations, limited lighting variation, similar exposure settings of camera images, and limited motion of objects in the scenes. Attempts have been made to autostitch images containing a number of moving objects with the objective of identifying and separating them, and assuring that they were fully included or excluded from the optimal image seam (Mills & Dudek, 2009). Yet results are not straight forwardly applicable.

Autostitch methods cannot be used within the CVBS proposed in this study because they require a running time higher than that expected for the detection model. Furthermore, also the registration process, which is the first step of the autostitch process (Gledhill, Tian, Taylor, & Clarke, 2003), represents a limit to the aim of this study because it may cause that unsuitable pixels are used when rotating and/or translating images of each camera scene, producing cow image duplication in the image seam or other undesired effects. Therefore, the methodology proposed in this study suggests a hand-made image registration process that produces the panoramic top-view image of the barn by matching, for each pair of image scenes, a number of foreground pixels of the image seam belonging to the body of the animals. This choice would assure that duplications of animal bodies do not occur in the panoramic top-view image and body shapes of the animals are maintained in the image seams (Figure 3).

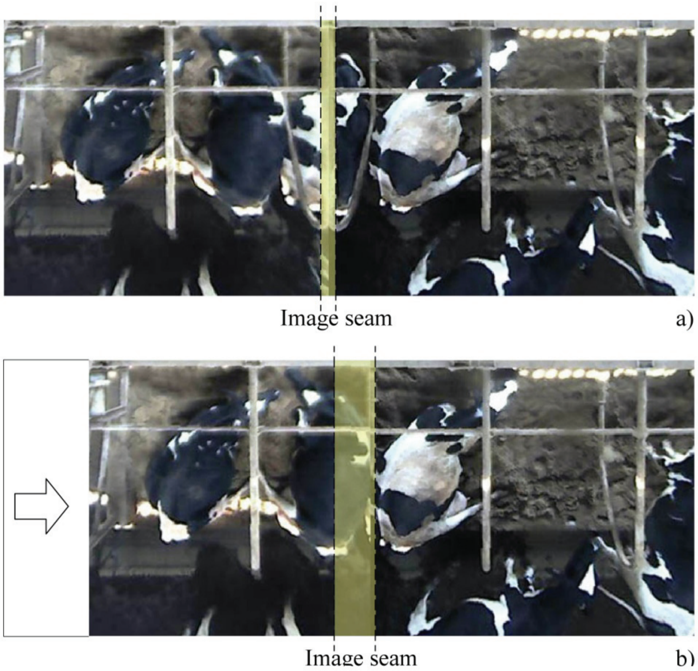


Figure 3 - Hand-made image registration process that produces the mosaicing of two cameras images.

The hand-made registration process requires that an operator who carries out, by using an image processing software, a number of image geometric transformations, i.e., rotation and translation, on a group of *n* image scenes having the same acquisition time. Since this step has to be executed only once for only one group of camera scenes, it is not a time-consuming operation. The parameters characterizing such transformations for each calibrated camera image, i.e., rotation angles, scale factors, and displacement, have to be stored and then used to automate the image mosaicing for all the acquisitions, by encoding them in a software tool. Since the proposed methodology does not require the merging and the blending of the camera images to obtain the plan view of the barn, the visualization of the whole area at a low number of frames per seconds is allowed. If required, this characteristic assures the utilization of the system for real-time analyses.

3.2 Object recognition in digital images with the Viola & Jones algorithm

3.2.1 Main concepts of the Viola & Jones Algorithm

A classifier for object detection in an image or a video sequence based on the use of the algorithm proposed by Viola & Jones (2001; 2004) is capable of processing images extremely rapidly and achieving high detection rates. The classifier used by Viola & Jones for human face detection on 384 × 288 pixels images was 15 times quicker than any technique at the time of release with 95% accuracy at around 17 fps. In general, a classifier based on the methodology proposed by Viola & Jones is a "view - based object detector" which must determine whether a sub window of an image belongs to the set of the object images to be detected.

The algorithm proposed by Viola & Jones relies on the use of simple Haar features (Papageorgiou, Oren, & Poggio, 1998) that are evaluated quickly through the use of integral images. A modified version of AdaBoost algorithm is used to find a number of best features within a large comprehensive set. Finally, the speed of detection is achieved by building a cascade of stages which are strong classifiers obtained by the combination of weaker classifiers.

The Viola & Jones algorithm is based on the following four concepts:

- Rectangular Haar features
- Integral Image
- The AdaBoost machine-learning method
- A cascaded classifier

3.2.1.1 Haar features

The Viola & Jones algorithm classifies images based on the value of simple features called Haar-like because of their similarity to Haar-basis functions.

A Haar-like feature is a real-valued function of a matrix containing the intensity values of the pixels of an image. The Haar-like features are composed of black and white regions having the same size and shape and are horizontally or vertically adjacent. A Haar-like feature value is the sum of the levels of brightness of the pixels in the white regions subtracted from the sum of the levels of brightness in the remaining black regions.

There are many reasons for using Haar-like features rather than the pixels directly. The most common reason is that these features can act to encode ad-hoc domain knowledge that is difficult to achieve using a finite quantity of training data. For this system there is also a second critical motivation for features: the feature-based system operates much faster than a pixel-based system.

A weak classifier utilizes only one feature to classify the information contained in a region of the image within an image. Within the overall set of

Haar-like features prototype Viola & Jones chose four them, the first with a vertical division, the second with a horizontal one, the third containing two vertical divisions and the last containing both the horizontal and the vertical divisions (Figure 4a).

Lienhart & Maydit (Lienhart & Maydt, 2002) introduced an extended set of Haar-like features starting from the standard Haar-like features that have been twisted by 45 degrees (Figure 4b). These twisted Haar-like features can also be fast and efficiently calculated using an integral image that has been twisted 45 degrees.

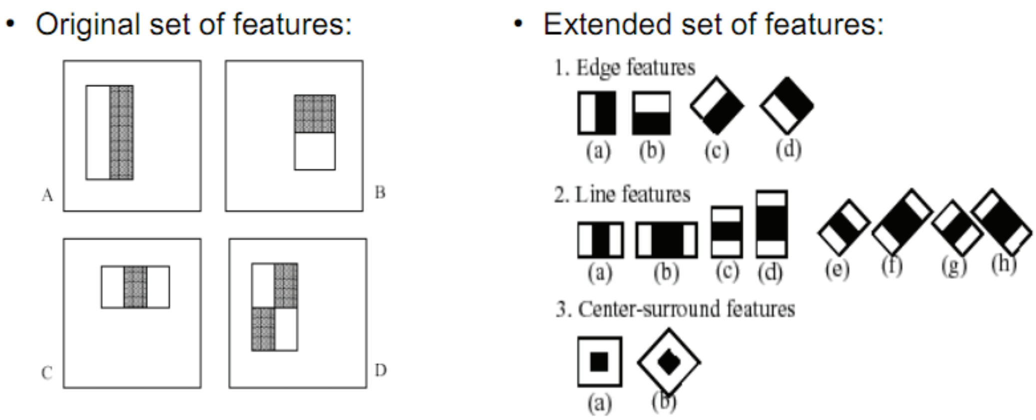


Figure 4 - a) The set of Haar-like feature prototypes used by Viola & Jones; b) The set of extended Haar-like feature prototypes**.**

3.2.1.2 Integral image

The Viola & Jones algorithm does not use the original input image of $w \times h$ pixels but computes a new image representation, the integral image, by making the level of brightness of each pixel equal to the entire sum of the level of brightness of the pixels above and at the left of the considered pixel of the original image. The use of such integral image allows for a very fast Haar-like features evaluation and consequently a faster object detection than that achievable by using other sets of features.

The brightness level of a pixel of the integral image at location *(x, y)* is:

$$ii(x, y) = \sum_{x' \le x, y' \le y} i(x', y') \qquad 0 \le x < w; \qquad 0 \le y < h; \tag{2}$$

where $ii(x, y)$ contains the level of brightness of the pixel at location *(x, y)* of the integral image and $i(x', y')$ contains the level of brightness of the pixels at location *(x', y')* of the original image (Figure 6).

Defining the cumulative row sum $s(x, y)$ by using the following formulas:

$$\begin{cases} s(x, y) = s(x, y-1) + i(x, y) \\ s(x, -1) = 0 \end{cases} \quad 0 \le x < w; \quad 0 \le y < h; \tag{3}$$

the integral image can be computed in one pass over the original image by using the formulas:

$$\begin{cases} ii(x, y) = ii(x-1, y) + s(x, y) \\ ii(-1, y) = 0 \end{cases} \tag{4}$$

Given a rectangular sub-window *D* (Figure 5), defined by means of the four coordinates *(x1,y1)*, *(x2,y2)*, *(x3,y3)*, *(x4,y4)* in an image, the sum of the level of brightness of the pixels within rectangle *D* is computed by means of the four reference values computed in the integral image reported below:

ii(1) = Area(A);

ii(2) =Area(A) + Area(B);

ii(3) =Area(A) + Area(C)

ii(4) = area(A) + area(B) + area(C) + area(D)

where:

Area (A) = sum of the brightness level of the pixels in rectangle A

Area (B) = sum of the brightness level of the pixels in rectangle B

Area (C) = sum of the brightness level of the pixels in rectangle C

Area (D) = sum of the brightness level of the pixels in rectangle D

Consequently,

the sum of the brightness level of the pixels in rectangle *D* can be computed as:

Area(D)= *ii*(4)+*ii*(1)−(*ii*(2)+*ii*(3))

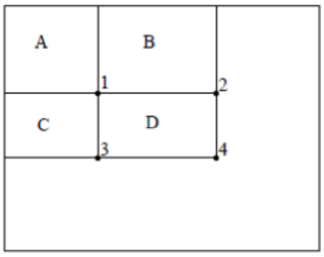


Figure 5 - A sub window D and the three sub windows A, B, and C considered for the calculation of the sum of the level of brightness of the pixels in rectangle D by using integral image.

For example, the sum of the brightness level of the pixels within the dark area of *i(x,y)* in Figure 6 can be computed by considering the highlighted values of *ii(x,y)* as:

$$36 + 6 - (19+9) = 14$$

i(x,y)

4	2	4	5
2	2	5	1
2	3	5	1
1	4	2	4
3	4	3	4
4	5	3	4

(a)

s(x,y)

4	6	10	15
2	4	9	10
2	5	10	11
1	5	7	11
3	7	10	14
4	9	12	16

(b)

ii(x,y)

4	6	10	15
6	10	19	25
8	15	29	36
9	20	36	47
12	27	46	61
16	36	58	77

(c)

Figure 6 – Example of an integral image computation by considering: a) level of pixel brightness of the input image; b) level of pixel brightness of cumulative row sum calculated on the input image; c) level of pixel brightness of the integral image.

The computation of the integral image is useful when some basic arithmetical operations, i.e., sums and differences of the level of the pixel brightness within rectangular areas, are to be carried out in the original image. In fact, any sum of the level of pixel brightness inside any rectangular area in the original image can be computed by means of four values of the integral image. Whereas, the difference of the sums of the pixel brightness of two different rectangular areas can be computed by knowing eight values of the integral image.

Since the Haar-like features defined by Viola & Jones are constituted by adjacent rectangles, their computation can be carried by using six, eight or nine values of the integral image in relation to the type of the considered Haar-like features (Figure 4).

3.2.1.3 The AdaBoost machine-learning method

In practical applications the use of only one weak classifier, i.e., one feature, is unsuitable to allow the classification of the information contained in an image. Therefore, the building of a strong classifier composed of a combination of weak classifiers is needed.

The number n of Haar-like features obtainable in a rectangular sub-window of $w \times h$ pixels can be computed by means of the following relation:

$$n \approx \frac{117941}{24^4} \times \left(w \times h \right)^2 \tag{5}$$

where 117941 is the number of all possible sizes and positions of the of Haar-like features obtainable in a rectangular sub-window of 24 × 24 pixels (Lienhart & Maydt, 2002).

By considering for example a rectangular sub-window of 30 × 40 pixels, the amount of possible Haar-like features (approximately 0.5×10^6) greatly exceeds the number of pixels in the sub-window, i.e. 1200 pixels. Even though each feature can be computed very efficiently, computing the complete set is prohibitively time-consuming. However, the experience of Viola & Jones showed a very small number of these Haar-like features can be combined to form an effective classifier. Infact, among all the Haar-like features obtainable within a rectangular sub-window of fixed size (Figure 7), only a few of them are expected to give the highest values when they overlap the object to be detected.

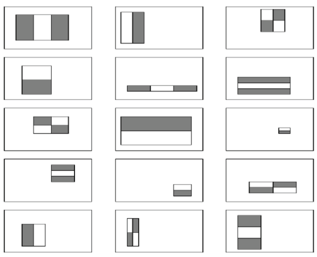


Figure 7 - Localization of Haar-like features within a rectangular sub-window of fixed size.

Therefore Viola & Jones proposed a modified version of the AdaBoost algorithm (Freund & Schapire, 1997) to find weak classifiers, using only significant Haar-like features, and to build a strong classifier as a combination of them. AdaBoost is a machine learning boosting algorithm capable of constructing a strong classifier through a weighted combination of weak classifiers. To match this terminology to the presented theory each Haar-like feature is considered to be a potential weak classifier. A weak classifier is mathematically described as:

$$h(x, f, p, \theta) = \begin{cases} 1 & \text{if} \quad pf(x) > p\theta \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Where x is a $w \times h$ pixel rectangular sub-window, f is the applied feature, p the polarity and θ the threshold that decides whether x should be classified as a positive or a negative value. Since only a small amount of the possible Haa-like feature values are expected to be potential weak classifiers, the AdaBoost algoritrhm is modified to select only the best features. The pseudo code contained in Figure 8 describes the modified AdaBoost algorithm. It takes a Haar-like feature set and a training set of positive and negative images, and finds T weak classifiers, each of them using a single Haar-like feature. The final strong classifier is a weighted linear combination of the T weak classifiers where the weights are inversely proportional to the training errors.

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively
- For $t = 1, \ldots, T$:

  1) Normalize the weights, $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$

  2) Select the best weak classifier with respect to the weighted error:

  $$\varepsilon_t = \min_{f,p,\theta} \sum_i w_i \left| h(x_i, f, p, \theta) - y_i \right|$$

  3) Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where f_t, p_t, θ_t are the minimizers of ε_t

  4) Update the weights:

  $$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

  where $e_i = 0$ if example x_i is classified correctly and $e_i = 1$ otherwise,

  and $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$
- The final strong classifier is:

  $$C(x) = \begin{cases} 1 \text{ if } \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 \text{ otherwise} \end{cases}$$

  where $\alpha_t = \log \frac{1}{\beta_t}$

Figure 8 - The pseudo code of the modified AdaBoost algorithm.

3.2.1.4 *The cascade of strong classifiers*

The reduced set of Haar-like features identified through the use of the modified version of AdaBoost is not sufficient to reduce the large amount of computation in the detection process. In fact, the final shape of the classifier proposed by Viola & Jones is a 'cascade' of 'stages' (Quinlan, 1986) which are strong classifiers.

It was also found that even if an image contained one or more searched objects, there is a large amount of evaluated sub-windows that would not contain objects. Starting from this consideration, the problem of finding objects within an image can be reformulated as follows: instead of finding objects, the classifier must discard the "non-objects", so at the end of the process the "not discarded" sub-windows contain the objects to be recognized.

A cascade of stages suited well to solve the above problem. The job of each stage is to determine whether a given sub-window is definitely not "an object to be recognized" or maybe "an object to be recognized". When a sub-window is classified to be a non-object by one of the stages, it is immediately discarded. Conversely a sub-window classified as a maybe-object is passed to the next stage in the cascade. It follows that the higher is the number of the stages a given sub-window passes, the higher the chance the sub-window actually contains the object to be recognized. Furthermore, the first stages of the cascade are very simple (using a few features) and are used to reject the majority of negative sub-windows, then more complex classifiers are used to refine the search by eliminating gradually the remaining negative sub windows and allowing only positive sub-windows to continue the path of detection until the end of the cascade. The flowchart illustrated in Figure 9 describes the algorithm that builds a cascade of classifiers, which constitutes the final classifier used to scan all sub-windows of the image.

3.2.2 Modelling and execution of the Viola & Jones classifier

The classifier based on the use of the algorithm proposed by Viola & Jones must be modelled by means of a training phase. In this phase, an operator must prepare a set of training data, and must set the training parameters. Subsequently the training algorithm, described in Figure 9, can be executed. The training process produces as output a series of summary results that will be analyzed by an operator to determine whether to accept the trained detector or restart the training by changing the set of training data or the values of the training parameters. The data set preparation may take several hours, whereas the execution of the training process may take from a few hours to several days on a conventional computer, e.g. 2.66 GHz Intel Dual Core.

In the execution phase, the trained detector searches for objects to be recognized in images which are provided by the user or come from a system of automatic image acquisition.

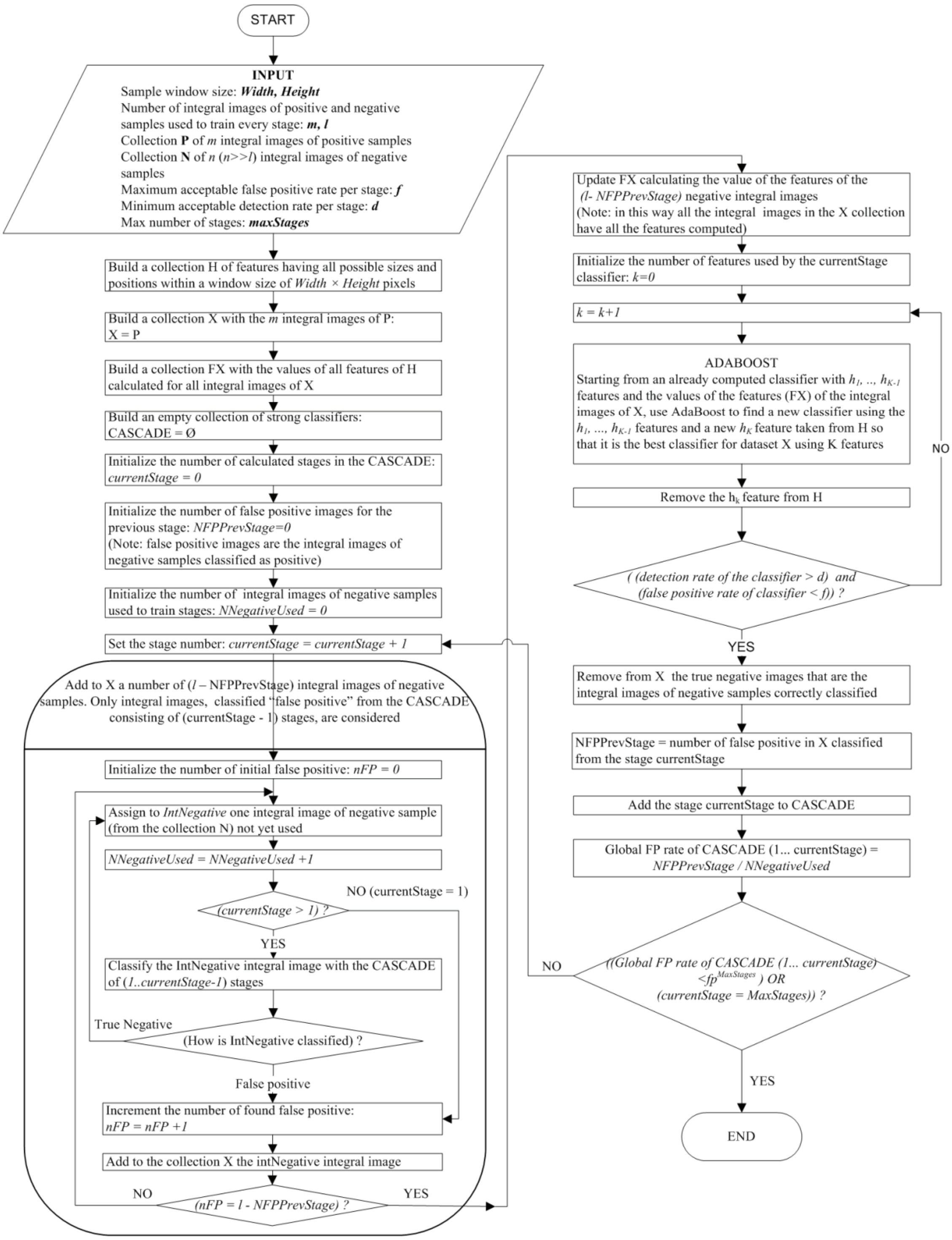


Figure 9 - Flowchart of the algorithm suitable to build the Viola & Jones classifier.

The execution of the detector produces for each image the description of a set of sub-windows surrounding each detected object. The execution of the detection process must be done in real-time, i.e., at least ten images of size 640×480 pixels must be processed in one second on a conventional computer.

3.2.2.1 Training

In the training phase, the classifier requires positive image samples and negative image samples, both having a size of $w \times h$ pixels. A positive image is defined as an image which contains the object to be recognized, whereas a negative image contains only other elements which constitute the image background. To obtain the image samples, an operator must extract a set of positive images having a size of $W \times H$ pixels and a set of negative images having a size of $W' \times H'$ pixels, from the frames acquired by the cameras. In the extraction process, all the positive images must maintain the same aspect ratio w/h. The sizes of the negative images, instead, should not be lower than those of the largest positive image.

The positive image sample is obtained by the algorithm reducing all the positive images selected by the operator to the size $w \times h$ pixels. The negative image sample, instead, is generated by the algorithm through the automatic execution of multiple scans on the negative images selected by the operator. In detail, a first scan is performed by using a sliding window of $w \times h$ pixels. For each row of the image, the algorithm puts the top left corner of the sliding window in each pixel of the considered row, starting from the first pixel on the left. At each new location of the sliding window, obtained by moving the window corner of one pixel to the right, the algorithm executes the extraction of the part of the image contained in the window.

Therefore, the number of negative images obtained by the first scan is given by the following relation:

$$N = \sum_{i=1}^{n} \left(W'_i - w + 1 \right) \times \left(H'_i - h + 1 \right) \tag{7}$$

in which n is the number of negative images selected by the operator, and W'_i and H'_i represent the width and the height of each of them, respectively. At each next scan, the sizes w and h of the window are increased by a scale factor z. The process ends when one of the sizes of the sliding window becomes higher than the corresponding size of the negative image analyzed.

During the training process, the algorithm classifies a negative image as "false positive" if it has wrongly recognized the target object in the image. Likewise, the algorithm classifies a positive image as "true positive" if it has correctly recognized the target object in the image. Therefore, besides the maximum number of stages (ns) which compose the cascade, the training phase requires the definition of the maximum value for the false positive rate

(*MaxFPR*), and the minimum value for the true positive rate (*MinTPR*). The fixed values of *MaxFPR* e *MinTPR* must be equal for all the stages. The parameter *MinTPR* is obtained from the relation:

$$MinTPR = \left(TPR_{cas}\right)^{1/ns} \tag{8}$$

by fixing the minimum value of the true positive rate that the cascade must produce (TPR_{cas}), and the value of *ns*.

During the training phase, the positive image sample is the same for every stage, whereas the negative image sample for an intermediate stage is made by all the negative images which contain false positives deriving from the previous stages and from other negative images that were not yet classified in the previous stages. At the end of the training process, when all the stages of the cascade have been built, the minimum false positive rate of the cascade (FPR_{cas}) is obtained. If it is of the order of magnitude 10^{-6} (Viola & Jones, 2001; Viola & Jones, 2004), then it shows that the classifier has a good ability to distinguish the background from the object to be detected.

3.2.2.2 Execution

In the execution phase, the algorithm of Viola & Jones utilizes a scanning process and a sliding window similar to those used in the training phase. Each stage of the cascade establishes if the area of the image contained within the sliding window has to be classified as “not an object” or as “probably an object” (Figure 10). If one of the stages establishes the content is classified as “not an object” that part of the image is definitively discarded and classified as background, otherwise the next stage will analyze the content. The higher the number of stages which evaluate the content of the sliding window, the higher the possibility that it actually would contain the object to be detected.

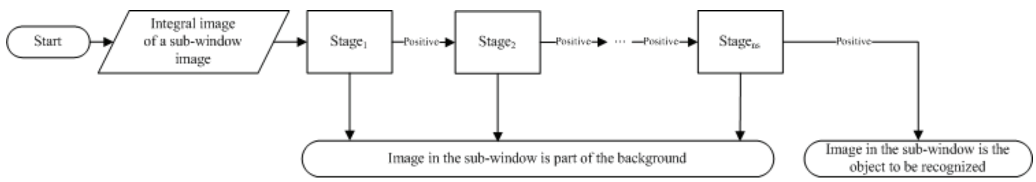


Figure 10 - Classification process of an area of the image.

The classifier analyzes the image at multiples scales. In fact, at each next scan, the sizes *w* and *h* of the sliding window are increased by a scale factor *z*. The process ends when one of the sizes of the sliding window becomes higher than the corresponding size of the analyzed image.

The resizing of the sliding window, rather than resizing the analyzed image, allows the computation of Haar-like features resized by a scale factor z without increasing the computational cost.

The effect of the sliding window is obtained by shifting the sub-window of a number $z \times \Delta$ pixels, where Δ is a parameter of the algorithm.

The choice of z and Δ affects both the speed of the classifier as well as the classification accuracy. Viola & Jones proposed a scale factor $z = 1{,}25$ and $\Delta = 1{,}5$ to allow significant speedup and good detection results.

3.3 Recognition of object contours in digital images with the normalized product scalar method

The search for an object image in a database can be done by comparing certain features extracted from the image with the corresponding features extracted from known images and stored in a database. An important visual feature used to describe an object within the image is the shape of the object. The automatic search of object shapes in digital images and the comparison between the shapes identified with those stored in a database has involved a great deal of scientific works (D. Zhang & Lu, 2004).

Before discussing the details of the proposed methodology, it is appropriate to distinguish three types of image processing on digital images (Gonzales & Woods, 2002): low level processing, mid level processing, and higher level processing.

Low level processing is characterized by the fact that both inputs and outputs are images and involves primitive operations such as image preprocessing to reduce noise, contrast enhancement, and image sharpening.

Mid level processing is characterized by the fact that inputs generally are images, but outputs are attributes extracted from those images, e.g., edges, contours, and involves tasks such as segmentation, i.e., partitioning an image into regions or objects, description of those objects to reduce them to a form suitable for computer processing, and classification, i.e., recognition of individual objects.

Finally, higher level processing involves the implementation of cognitive functions that allow the assignment of a meaning to the scene in the image.

3.3.1 Main concepts of object contour detection

3.3.1.1 Image segmentation

The interpretation of the contents of an image through the digital image processing is based on human interpretation that considers an image as an arrangement of regions and objects (Malik, Belongie, Leung, & Shi, 2001). Human visual grouping was studied extensively by the Gestalt psychologists in the early part of the 20th century (Wertheimer, 1938). They identified several factors that lead to human perceptual grouping: similarity, proximity, continuity, symmetry, parallelism, closure and familiarity.

In computer vision, the image segmentation is the process of grouping pixels having similar visual shapes, and separating them from the background. The segmentation process first produces elementary graphical elements, e.g. areas, lines, curves, edges, that belong to or surround real objects found in the image, and afterwards uses them to obtain the shape of each object in the image.

The methods used for the automatic search for shapes can be grouped into three categories (Bradsky & Kaehler, 2008): pixel based methods only use the gray values of the individual pixels and do not even consider the local neighbourhood; region based methods analyze the gray values in larger areas in order to find homogeneous regions; contour based methods detect edges and then try to follow them by looking only for discontinuities.

The contour based method for the automatic search for shapes was used in this study because it best suits to the case study described later in section 3.4 where shape characteristics of cows are required.

3.3.1.2 Edge detection and contour findings

Edge detection is a low-level image processing tool which identifies edges, i.e. pixels in which the pixel brightness presents substantial local variations of intensity. Edges typically occur on the boundary between two different regions of the image. The process of edge detection consists of four steps: smoothing, enhancement, detection, and localization. Smoothing suppresses as much noise as possible, without destroying the true edges. Enhancement enhances the quality of the edges in the image increasing the contrast between each pixel and its neighbours. Detection determines which edge pixels should be discarded as noise and which should be maintained. It involves convolving the image with filters, which are constructed to be sensitive to local changes of intensity in the image whereas they return zero values in uniform regions. Localization estimates the edge that best fits the identified edge pixels.

The algorithm used in this work to find for edges in images is the Canny edge detector (Canny, 1986) which is widely considered as the optimal edge detector (Gonzales & Woods, 2002). The algorithm is not fully described in this thesis, however, as other edge detectors, it requires a greyscale input image, and provides as output a binary image which consists only of detected edges that separate different areas in the image. Finally, the binary image provided by the edge detection process is used within a well consolidated algorithm (Suzuki, 1985) that analyzes the topological structure of a binary image by following the edges in the image. For each edge detected, the Suzuki's algorithm produces as output a descriptor also known with the name of "contours", i.e., a sequence of coordinates in the image that delimits the object from the background.

3.3.1.3 Contour representation

All operations discussed in the previous paragraph extract objects contour from images and describe them through an ordinate sequence of points. Since similar objects may appear different if they are framed at different distances or if they are rotated, the scientific community has focused the search on contour representations that make it possible to obtain scale and rotation invariant features of the contours (D. Zhang & Lu, 2004). Various representations of contours and relative based contour descriptors have been developed: chain codes, autoregressive models, wavelet descriptors, curvature scale space, moment, and Fourier descriptors (Yadav, Nishchal, Gupta, & Rastogi, 2007).

In this work the contours representation elaborated by Zahn and Roskies (Zahn & Roskies, 1972) was used. These Authors elaborated a Fourier descriptors-based technique for contour matching. In this technique is assumed that the contour of an object is represented by a closed polyline which does not have self intersections. A set of *L* points representing the contour of the object obtained by sampling contour pixels in clockwise or anticlockwise way, is available in the form of Cartesian image coordinates:

$$c(t):(x(t),y(t)) \qquad t=0,1,\ldots,L-1 \tag{9}$$

A complex coordinate function *z(t)* is obtained calculating the offsets on *x* axis and *y* axis between two consecutive points:

$$z(t)=u(t)+jv(t)=\begin{cases}(x(t+1)-x(t))+j(y(t+1)-y(t)) & t=0\ldots L-2\\ (x(0)-x(L-1))+j(y(0)-y(L-1)) & t=L-1\end{cases} \tag{10}$$

where the symbol of the imaginary unit is denoted with *j*.

It is also valid to extend *z(t)* domain to values greater than *L-1* by putting:

$$z(t)=z\left(t+L\right)$$

and, therefore, *z(t)* is a complex periodic function of *L* period .

Figure 11 shows the exemplification of the *z(t)* function, related to the highlighted contour, as a result of the conversion of Cartesian image coordinates in complex coordinates.

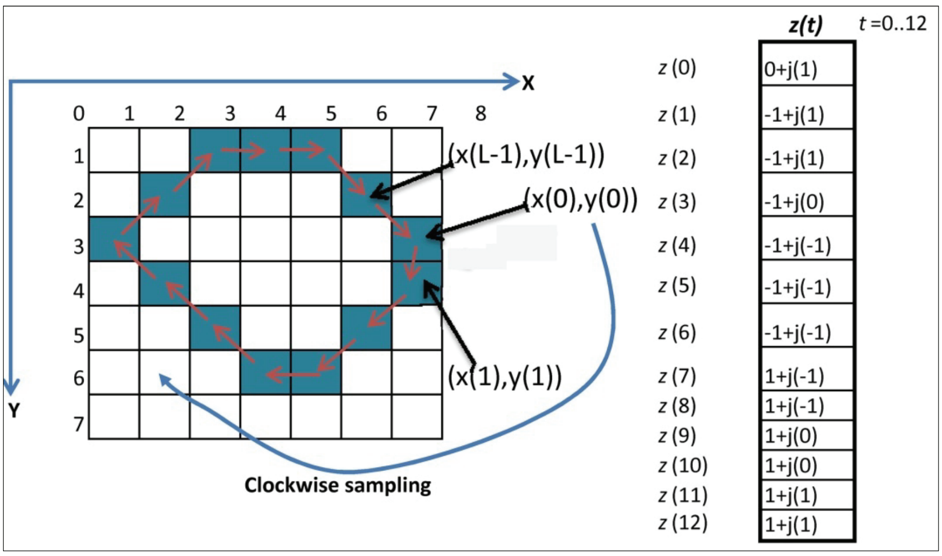


Figure 11 - Examples of a graphical contour constituted by 13 points and its representation with complex coordinates.

The representation of the contours by using the function *z(t)* has some common properties when considering geometrical transformations of the shapes in the Cartesian space (Li & Lee, 2005). In particular *z(t)* is translation invariant along *x* and *y* axis, i.e., if *c(t)* is moved of Δx in the *x* direction and Δy in the *y* direction, then *z(t)* takes the same values. Furthermore, a geometrical rotation of *c(t)* in the Cartesian space around any point (also different from the centroid point) corresponds to a number of cycle displacements of the elements of the vector function *z(t)*. Finally, zoom in or zoom out of the shape in Cartesian space with *s* scale factor corresponds to an arithmetical multiplication or division of each element of the vector function *z(t)* by the factor *s*.

3.3.1.4 The proposed contours matching methodology

In this study the matching between the contour representations obtained in the analyzed images and those contained in the database were achieved by applying the normalized scalar product method (Sivic & Zisserman, 2009).

The representation of the contours through the complex function *z(t)* is used to obtain a synthetic index that indicates the level of similarity between two contours. The index used for contours matching refers to the definition of the scalar product between two vectors of complex numbers. It reaches its maximum value when correspondent elements of the two vectors are aligned (Sivic & Zisserman, 2009).

Let consider $z_1(t)$ and $z_2(t)$ as representations of two contours having the same number *L* of boundary points and obtained by sampling their contour pixels in the same way, i.e., clockwise or anticlockwise:

$$\begin{cases} z_1(t) = \left[z_1(0), z_1(1), ..., z_1(L-1)\right] = \left[(u_1(0) + jv_1(0)), (u_1(1) + jy_1(1)), ..., (u_1(L-1) + jv_1(L-1))\right] \\ z_2(t) = \left[z_2(0), z_2(1), ..., z_2(L-1)\right] = \left[(u_2(0) + jv_2(0)), (u_2(1) + jv_2(1)), ..., (u_2(L-1) + jv_2(L-1))\right] \end{cases} \quad (11)$$

The *Like* index is a complex number defined as follows:

$$Like\left(z_1(t), z_2(t)\right) = \frac{\sum_{i=0}^{L-1}\left(z_1(i); z_2(i)\right)}{\left\|z_1(t)\right\|\left\|z_2(t)\right\|} \quad (12)$$

where $\left(z_1(i); z_2(i)\right)$ is the Euclidean inner product of complex numbers $z_1(t)$ and $z_2(t)$:

$$\left(z_1(i); z_2(i)\right) = z_1(i)\overline{z_2(i)} = \left[x_1(i) + jy_1(i)\right]\left[x_2(i) - jy_2(i)\right] \quad (13)$$

Where $\left\|z_1(t)\right\|$ and $\left\|z_2(t)\right\|$ are the norms of the two vectors of complex numbers $z_1(t)$ and $z_2(t)$, and $\overline{z_2(i)}$ is the complex conjugate of the complex number $z_2(t)$.

For instance the norm of $z_1(t)$ is written as follows:

$$\left\|z_1(t)\right\| = \sqrt{\sum_{i=0}^{L-1}\left(z_1(i);z_1(i)\right)} = \sqrt{\sum_{i=0}^{L-1}\left(z_1(i)\overline{z_1(i)}\right)} =$$

$$= \sqrt{\sum_{i=0}^{L-1}\left[u_1(i)+jv_1(i)\right]\left[u_1(i)-jv_1(i)\right]} =$$

$$= \sqrt{\left[u_1^2(0)+v_1^2(0)\right]+\left[u_1^2(1)+v_1^2(1)\right]+\ldots+\left[u_1^2(L-1)+v_1^2(L-1)\right]}$$

The *Like* index has the following invariance properties:

1. $\left|Like\right| = \sqrt{\mathrm{Re}^2\left(Like\right)+\mathrm{Im}^2\left(Like\right)}$ is moving invariant, rotation invariant, and resizing invariant, i.e., if in the Cartesian space $z_1(t)$ or $z_2(t)$ are independently moved, or if $z_1(t)$ or $z_2(t)$ are independently rotated around any point, or if $z_1(t)$ or $z_2(t)$ are independently resized then $\left|Like\right|$ index is the same.

2. If $z_1(t)$ and $z_2(t)$ are two representations of the contour of the same object, which can be moved, resized or rotated in the two representations then it is $\left|Like\right| = 1$, which is the maximum obtainable value. If, instead, $z_1(t)$ and $z_2(t)$ are two representations of the contours of two different objects, it results that $0 < \left|Like\right| < 1$, and $\left|Like\right| \to 0$, at increasing of the diversity between the shapes of the two objects.

3. If $z_1(t)$ and $z_2(t)$ are two representations of the contour of the same object and $z_1(t)$ or $z_2(t)$ is rotated by an angle α around any point then is:

$$\mathrm{Re}(Like) = Cos\alpha$$

However, the invariance properties of the *Like* index calculated for two contour representations $z_1(t)$ and $z_2(t)$ of the same object are valid only if the starting points of the two contours, $z_1(0)$ and $z_2(0)$, correspond to the same part of the object. This is unlikely to occur because only one contour is known. This problem is solved by calculating the *Like* function *L* times between $z_1(t)$ and other *L* functions $Z_{2k}(t);\ k=0\ldots L-1$ obtained through *L* cycle displacements of $Z_2(t)$.

E.g.:

$$Z_2(t) = \begin{bmatrix} u_1 + jv_1 \\ u_2 + jv_2 \\ u_3 + jv_3 \\ u_4 + jv_4 \end{bmatrix} \Rightarrow Z_{20}(t);\ Z_{21}(t);\ Z_{22}(t);\ Z_{23}(t)$$

where:

$$Z_{20}(t) = \begin{bmatrix} u_1 + jv_1 \\ u_2 + jv_2 \\ u_3 + jv_3 \\ u_4 + jv_4 \end{bmatrix};\ Z_{21}(t) = \begin{bmatrix} u_2 + jv_2 \\ u_3 + jv_3 \\ u_4 + jv_4 \\ u_1 + jv_1 \end{bmatrix};\ Z_{22}(t) = \begin{bmatrix} u_3 + jv_3 \\ u_4 + jv_4 \\ u_1 + jv_1 \\ u_2 + jv_2 \end{bmatrix};\ Z_{23}(t) = \begin{bmatrix} u_4 + jv_4 \\ u_1 + jv_1 \\ u_2 + jv_2 \\ u_3 + jv_3 \end{bmatrix}$$

The contour $Z_{2\tilde{k}}(t)$ that maximizes $\left|Like\left(Z_1(t), Z_{2k}(t)\right)\right|$; $k = 0, \dots, L-1$ is used for the comparison with $z_1(t)$. The *Like* index is therefore modified as follows:

$$\begin{cases} \widetilde{Like}\left(z_1(t), z_2(t)\right) = Like\left(z_1(t), z_{2\tilde{k}}(t)\right) \\ \text{where } \tilde{k} \text{ is chosen so that } \left|Like\left(z_1(t), z_{2\tilde{k}}(t)\right)\right| = \max\limits_{k=0,\dots,L-1} \left|Like\left(z_1(t), z_{2k}(t)\right)\right| \end{cases} \tag{14}$$

The $\widetilde{Like}$ index has the same invariance properties as the $Like$ index and can be used for any choice of the starting points for the contours $z_1(t)$ and $z_2(t)$.

3.3.2 Modelling and execution of the proposed contour matching algorithm

The contour classifier based on the use of the complex function $z(t)$ for the contour representations and the normalized product scalar method for contours matching involves three different phases:

- Definition of the classes of the objects to be detected;
- Construction of a database of known contours related to the object classes;
- Execution phase in which the contour classifier searches for each unknown contour among the known contours stored in the database and associates it to the appropriate object class.

First of all some parameters must be defined: the direction of movement used for the search of contour of objects (clockwise or anticlockwise), the number L of points which constitute the contour of the objects, and the geometric constraints for the contours, e.g., minimum and maximum values for perimeter and area. These parameters are used for both the processing of images used for the construction of the database and the processing of unknown images in the execution phase.

3.3.2.1 Construction of a database of known contours

In this phase an operator must collect an input data set of images that contains the objects to be recognized. Multiple images are required because the object could be framed with different perspectives. It is possible to insert also images that contain only some portion of the object with the aim to recognize it whenever it is partially visible in the image.

Each image in the input dataset is processed as described in Figure 12 by implementing a software tool that allows for the automatic extraction of contours and the next operator visual check required in order to associate among the extracted contours that corresponding to the object to be recognized. In detail, the operator must assign to the selected contour a unique label that associates it to the object. The results of this check must be stored in a database.

The data set preparation may take several hours, and it does not depend on computer speed, but on operator's ability to select meaningful images that contain objects or symbols to be recognized. The population of the database ends when it is not possible to add new contours because the index of similarity $\widetilde{Like}$ computed between the extracted contour and those already stored in the database was lower than a pre-established threshold.

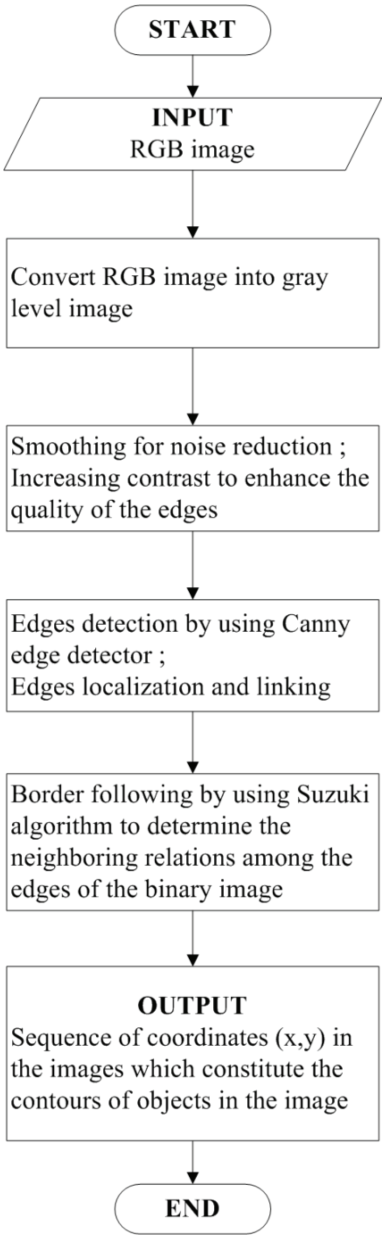


Figure 12 – Process used to extract different object contours in an image.

3.3.2.2 Execution

In the execution phase, the known contours stored in the database are used for object identifications in images which were provided by the operator or came from a system of automatic image acquisition. For each image to be analyzed and for each object to be detected, an automatic tool, which must be implemented, executes the activities described below:

1. Finds a set of unknown contours (*SC)* by processing the input image as described above (Figure 12);

2. Removes from *SC* those contours that do not meet the geometric constraints;
3. For each unknown contour in SC compares it with the known contours stored in database by using the $\widetilde{Like}$ index;
4. Associates each unknown contour to the right object class if the $\widetilde{Like}$ index is greater than a pre-established threshold $Th_{\widetilde{Like}}$.

The definition of the value of the threshold $Th_{\widetilde{Like}}$ is crucial to obtain a good level of contour detection rate and, at the same time, obtain low the false positives rates.

3.4 The case study

3.4.1 The area of the barn under study

The objective of the activities described in this paragraph which refers to the application of the above described methodologies to a case study, was to demonstrate the effectiveness of the proposed CVBS for the automatic detection of dairy cow behavioural activities in free-stall barns and for the automatic identification of dairy cows in free-stall barns.

The trial was carried out during the years 2011 and 2012 in a dairy cow free-stall barn located in the municipality of Vittoria in the province of Ragusa.

First of all, a direct metric survey of the barn was carried out to obtain the plan and two sections of the building. The barn was characterized by a rectangular plan of about 55.6 m × 20.7 m with three sides completely open, i.e. without outside walls (Figure 13). The roof was symmetric and covered by fibro-cement sheets supported by a bearing structure made of steel trusses and purlins. The feeding alley of about 55.75 m × 3.50 m was adjacent to the resting area that was arranged with two rows of 64 stalls faced head-to-head and filled with sand. Service alleys allowed the easy access of the cows from the feeding alley to the service alley for the second row of stalls. The side of the barn at the back of the second row of stalls was completely open.

An interview with the breeder and the direct observation of the breeding environment allowed the knowledge of the management activities carried out in the barn. Feed was delivered to the cows once a day at approximately 06:30 a.m. and was moved closer to the feed barrier later in the day at 4:30 p.m.. Milking occurred twice a day at 06:00 a.m. and 05:00 p.m. Furthermore, in the alleys there is presence of slurry accumulation as the cleaning is not automated but it is carried out 1-2 times a day by a scraper.

Within the barn a reduced but complete breeding area of about 15.40 m × 11.50 m was considered for the experimental trial (Figure 14). In this area a group of 15 Holstein dairy cows was housed. The breeding environment was composed of a resting area of about 10.40 m × 4.30 m with two rows of stalls arranged head-to-head and sand beds, a feeding alley of about 15.40 m × 3.50 m adjacent to the resting area, a service passage and two service alleys.

Considerable variability of the brightness conditions during the day in the analyzed breeding environment was observed. Factors which could make the detection of cow behavioural activities and the identification of the cows within the stalls more difficult, are: colour variations of the sand beds; high brightness variation in areas close to the open side of the barn; high levels of solar radiation reflected by sand, wet floors and metal surfaces of the stable cross-bars; lack of colour homogeneity; alley surface reflection caused by manure and presence of shadows.

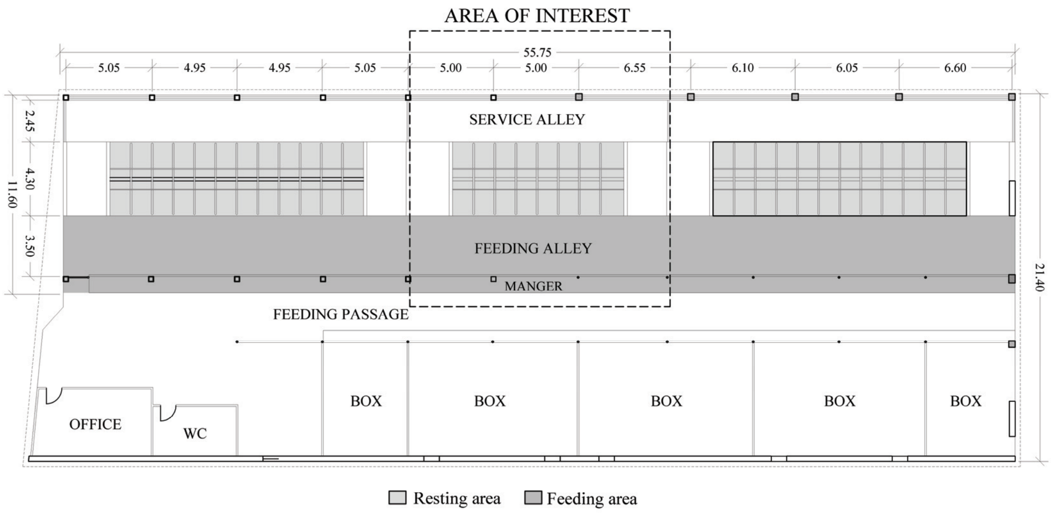
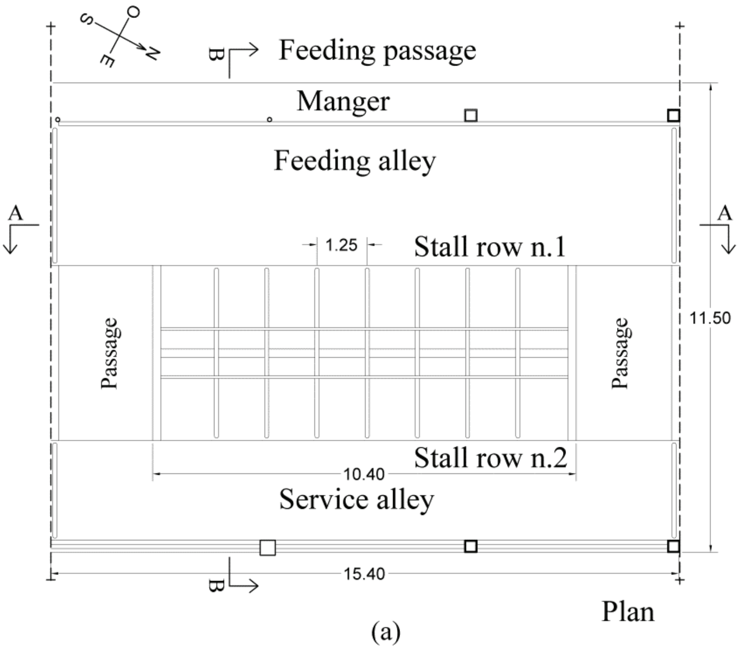


Figure 13 - Plan of the free-stall barn showing the area object of the experimental trial.



(a)

Figure 14 - a) Plan of the study area.

3.4.2 Developed software tools

Several specific software tools have been developed to perform the automatic image processing activities required both to obtain the panoramic top-view images of the barn and to develop and apply the two methodologies proposed for cow behavioural activities' detection and cow identification. All the software tools developed in this study were implemented in Visual C++ 2008 express edition, and Visual C# 2008 express edition within two integrated environment of software programming free distributed by Microsoft®, which allow the development of applications written in C++ language and C# language and the use of all the graphical components of the Microsoft® Windows operating system.

Furthermore, the OpenCV (Open Source Computer Vision Library), an open-source BSD-licensed library (Bradsky & Kaehler, 2008) was used because it provides the user with a computer vision infrastructure that helps software developers to build sophisticated vision applications providing functions for low level image processing and mid level image processing.

3.4.3 The design and installation of the video recording system

The design and installation of the multi-camera video recording system in the free-stall barn were firstly carried out because of their relevance for the attainment of the two main objectives of the study, i.e., the automatic detection of dairy cow behavioural activities following the first methodology and the automatic identification of dairy cows following the second methodology.

After the direct metric survey which made it possible to produce the plan and the sections of the area under study (Figure 19), the activities reported in the following of this section have been carried out to obtain the calibrated camera images and the panoramic top-views of the barn in real-time.

3.4.3.1 Camera model selection

The height of the barn and the bearing structure of the roof have constrained the maximum value for the cameras installation height.

The cameras installation height above the floor of the feeding area (h'_{cam}) was about 4.40 m, whereas that above the floor of the resting area (h''_{cam}) was about 4.00 m. The height from the floor of the two foreground plan was (h'_{forg}) about 1.40 m and (h''_{forg}) about 0.70 m (Figure 19). The small values of h'_{cam} and h''_{cam} had prompted the use of a camera model, Vivotek FD7131, equipped with wide angle having the maximum horizontal angle of view provided by the technical specifications of 105.1°, whereas the maximum vertical one was of 77.4°. This choice led to reduce the number of the cameras required to obtain the plan view of the barn. The network camera model had a maximum resolution of 640 × 480 dpi, up to 30 fps image-capture capability, and was equipped with light-emitting diodes (LEDs) for night illumination and a HTTP based interface through which the developed software can read or set various media control functions for camera and require a snapshot.

Though the camera model was provided with a built-in LED illuminator, it did not provide sufficient lighting in the barn in the evening and night-time hours. Therefore, the panoramic images used in the training phase refer to the time interval between 6:00 a.m. and 8:00 p.m. The time interval which was chosen did not affect the modelling of the detectors because it included the most significant herd management tasks and illumination condition variations. In particular, it included the feeding distribution, the two milking operations during which the animals left the resting area, the activation of the evaporative cooling devices and the direct dripping system which produced a significant variation of the colour of the sand beds in the stalls, the cleaning of the feeding alley which determined a variation of the appearance of the floor, and light intensity variations in the different areas of the breeding environment throughout the day.

3.4.3.2 Camera image calibration

A specific software component tool was implemented in C++ to perform the automatic rectification of distorted camera images. Rectified camera images have been available for subsequent processing, e.g., to compose each camera images in panoramic top-view images.

The setting of the calibration process for each camera was firstly carried out in laboratory. A drawing of chessboard was placed at a distance of 4 m that was about the distance between the camera installation position and the foreground plane in the barn. Fifteen snapshots (Figure 15) per camera were obtained filming the chessboard from different angles.

Figure 15 – A set of chessboard images used for the camera calibration process.

The software firstly uses the *cvFindChessboardCorners* function of the OpenCV library to find each corner of the chessboard, then the *cvCalibrateCamera2* function of the OpenCV library is used to generate the set of equations, to solve them, and to get the distortion matrix and the camera matrix. The matrices are then stored in XML files to be used by the *cvInitUndistoryMap* and *cvRemap* functions of the OpenCV library to rectify the corresponding distorted camera images.

Calibration software tool was executed for each set of chessboard snapshots captured by each camera. The outputs of the tool were the camera matrices and the distortion matrices of each camera. Finally, some test were performed to verify the goodness of the calibration process: the ceiling of the laboratory was filmed by pairs of cameras and the calibrated images were used to create manually a panoramic rectified view image of the ceiling. Examples of images of the ceiling of the laboratory used for test the calibration process are shown in Figure 16.

3.4.3.3 Evaluation of the maximum horizontal and the maximum vertical view angles after the calibration process

The calibration process produced for each camera image a decrease of the view angles of about 31%. Therefore the horizontal view angle was about 72.5°, whereas the vertical one of about 53.4°.

To obtain the full rectified panoramic top-view (Figure 19) of the two selected functional units i.e., the feeding area and the resting area, the installation of four cameras above the resting area and of six cameras above the feeding area, respectively, has been determined by solving equation (1).

Two parallel steel beams anchored to the bottom chords of the steel trusses were used as supports for the installation of cameras. The distance between these steel beans was 3.86 m and the images of each scene overlap each other in the row direction by approximately 37%, and in the direction perpendicular to the row by approximately 40%. These overlapping areas were used to obtain the panoramic top-view of the two functional units. Figure 17 shows the results of the calibration process carried out for one of the cameras installed above the resting area. The ten calibrated camera images are subsequently used in the mosaicing process.

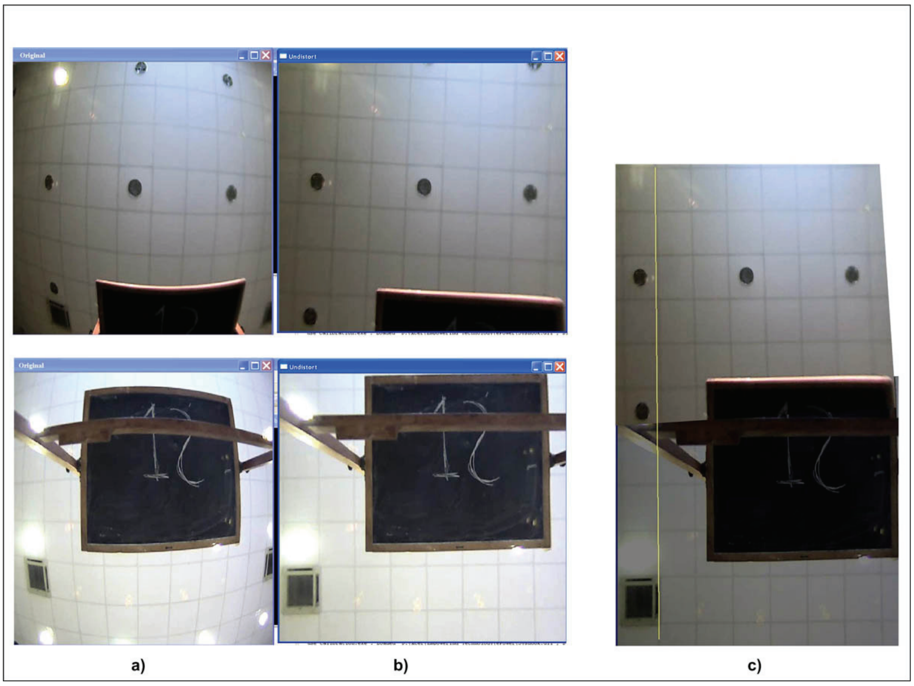
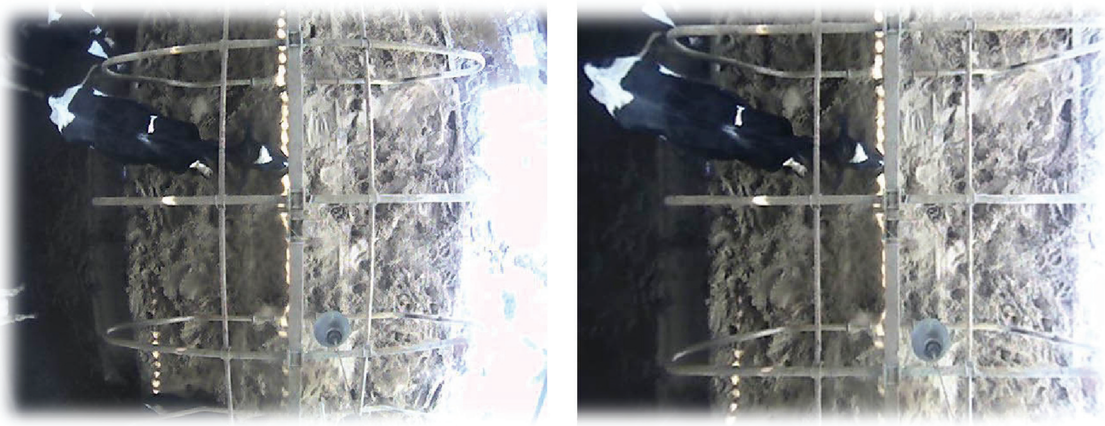


Figure 16 - a) Uncalibrated image; b) Calibrated image; c) Composition of calibrated images.

The multi-camera image recordings were carried out by using a desktop personal computer with a processor Intel® Core™2 Quad CPU Q6700 at 2.66 Ghz, 3 Giga bytes of RAM and Windows Vista™ Business operative system. The software tools for synchronous cameras images acquisition and camera image calibration were installed in the personal computer and were activated when all the hardware was connected and turned on (Figure 18). Furthermore, an external hard disk was connected to the personal computer to be fully employed for transport of acquired images. An ADSL connection allowed the remote control of the multi-camera recordings through the Internet.

3.4.4 Full rectified panoramic top-view image of the area under study

A specific software component was implemented in C++ language to perform the acquisition of digital images from cameras in synchronous mode. Firstly the software makes asynchronous and simultaneous requests to all cameras HTTP interfaces to download one snapshot; then each camera web server returned the most up-to-date snapshot in JPEG format.



a) b)

Figure 17 - a) Uncalibrated image; b) Calibrated image.

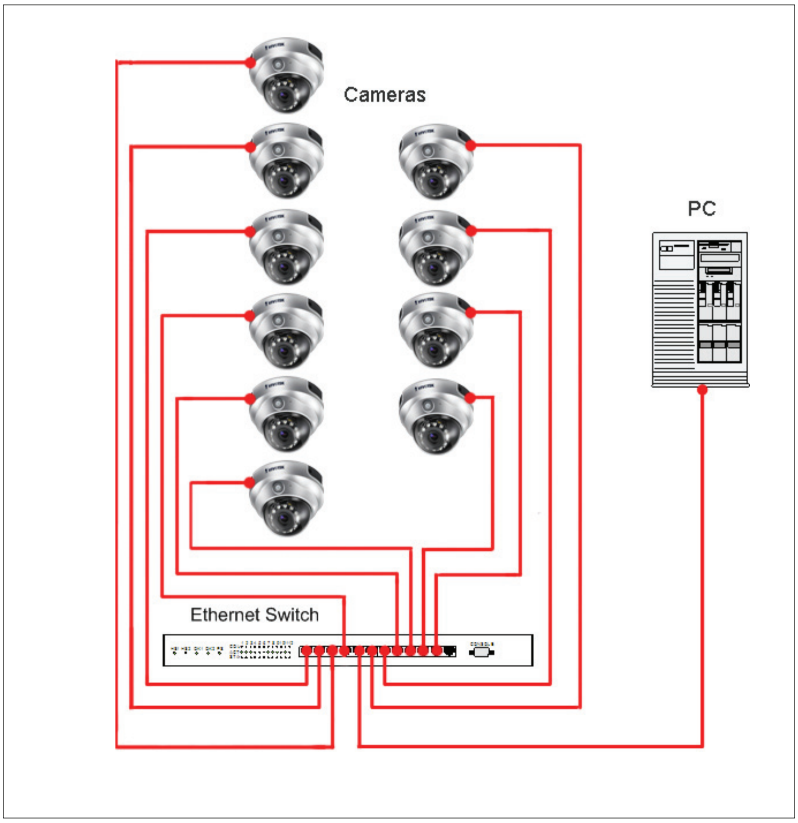


Figure 18 - Hardware components installed in the free-stall barn and link schema.

The software waits until all the snapshots are available so that to avoid the accumulation of delay times in the video sequences.

The panoramic top-view of the two selected functional units was obtained by mosaicing the set of calibrated and synchronised images coming from the 10 cameras. In detail, for each pair of images coming from two contiguous cameras, those showing the body of the cow in the image seam were selected and overlapped by using pixels belonging to the body of the cow. By means of an image processing free-software, an operator carried out a total amount of 25 image geometric transformations which parameters were stored in order to automate the procedure for all the sequence of the video recordings.

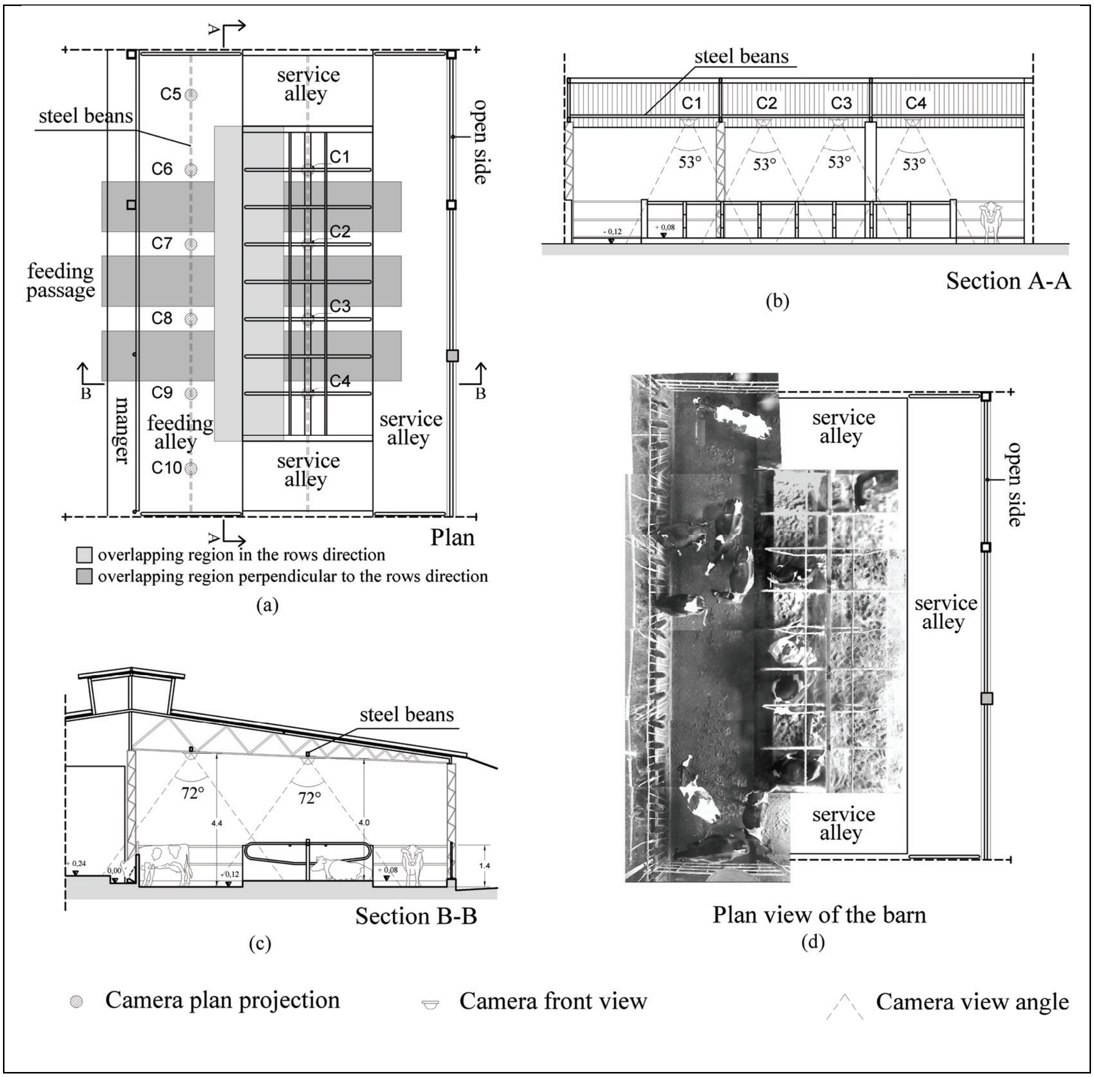


Figure 19 - a) Plan of the area of interest showing the plan projection of the 10 cameras and the overlapping region of the camera scenes; b) Longitudinal section A-A showing the position of the first rows of cameras and the camera vertical view angle; c) Transverse section B-B of the area of interest showing the position of the two rows of cameras, their height from the floors, and the camera horizontal view angle; d) mosaic images obtained from the 10 camera scene.

The execution of the image synchronization and image mosaicing on each set of calibrated and synchronised images produced a video sequence characterized by frames with resolution of 1044 × 1920 pixels (Figure 19) at 1 frame every 2 seconds. This value of the frame rate represented a constraint for the detection and for the identification phases because the time required for the automatic detection of all the cow behavioural activities and for the identification of the signed cows, considered in this study, must not be greater than the time required for obtaining both the calibrated camera images and the panoramic top-view of the barn, i.e. 2 seconds.

3.4.5 First methodology: the automatic detection of cow behavioural activities in free-stall barns (objective 1)

The method involves the simultaneous application of several classifiers to panoramic top-view images of the free-stall barn to detect each behavioural activities analyzed in this study, i.e., lying, feeding, standing, perching. For the automatic construction of the top-view panoramic images, the method assumes that a set of synchronized images, collected from the cameras installed in the free-stall barn, is available, and also an operator has identified a set of geometric transformations, i.e., rotation, translation, and resizing, on the group of calibrated camera images installed in the free-stall barn that were used to produce the panoramic top-view image of the barn.

The method is constituted by four sequential phases that involve the use of the CVBS for the detection of dairy cow behavioural activities in free-stall barns:

a) Training phase which allows the construction of the classifiers for each cow behavioural activity;
b) Development phase that involves the development of the software component of the CVBS used for the detection of the cow behavioural activities in the free-stall barn;
c) Test of classifier functioning for a large image dataset, yet without making use of operator’s visual recognitions;
d) Accuracy assessment procedure that allows the validation of the proposed method by comparing the results of the detection obtained using the automatic software tool with the detection results obtained through an images observation activity performed by an operator.

In the following, the phases of the proposed detection method are described in detail.

3.4.5.1 Training phase

A training software tool was developed by implementing the algorithm for the building of the Viola & Jones classifier described in paragraph 3.2.1.4. It required as input: positive image samples that contain cows behavioural activities, negative image samples, the values w and h of the rectangular sub-window, the values of *MaxFPR*, *MinTPR*, and *ns*.

The positive image samples with behavioural activities were extracted manually by an operator from panoramic top-view images by using generic tools for image processing, e.g. GIMP (GNU Image Manipulation Program) which is a free and open-source software.

Positive images should be subdivided into behavioural classes. Images belonging to the same behavioural class must have the same aspect ratio, i.e. w/h, to be resized by the algorithm to $w \times h$ without distortions. Each behavioural class could be subdivided into more sub-classes in relation to the specific requirements of the behavioural activity analysed. The number of the sub-classes mainly depends on the positions occupied by the body of the animals within the functional units of the free-stall barn, i.e., resting area, service alley, and feeding area.

The following behavioural classes, with their sub-classes, were defined:

a) 'Lying': this class shows cows lying in the stalls. A number of sub-classes could be required in relation to the stall layout, i.e., head-to-head or back-to-back. This subdivision is required because the algorithm is not invariant to the rotation of the body of the animal. The sub-windows used for the sample selection must coincide with the perimeter of the stall visible in the frame (Figure 20a).
b) 'Standing': this class shows cows standing still or walking in the barn. Since positive images of 'standing still' behavioural activity are very similar to 'walking' ones the detection method cannot distinguish between them. Therefore, walking and standing still (bunching) behaviours must be grouped in the same class named standing. At runtime, the algorithm could differentiate the two behaviours on the basis of the analysis of consecutive frames. Since each cow can occupy any position within the service alleys at least three sub-classes have to be considered: one sub-class shows cows standing toward the direction of the longitudinal axis of the barn; the second sub-class shows animals standing at right angles to the longitudinal axis of the feeding alley; and the third sub-class shows cows oblique to the longitudinal direction. As considered for the lying class, this subdivision is required because the algorithm is not invariant to the rotation of the body of the animal. The sub-windows for the first two sub-classes must coincide with rectangles including the body of the animal (Figure 20b to Figure 20d). The sub-windows of the third sub-class should have square aspect ratio (Figure 20c). The head of the cow should be excluded from the selection because its inclusion in the

sub-windows would enlarge the number of pixels belonging to the background, increasing the occurrence of false positives during the detection.

c) ‘Feeding’: this class shows cows having their head through the feed barrier. The dimensions of the sub-windows have the same characteristics of those described for the second sub-class of the standing class (Figure 20e).

d) ‘Perching’: this class shows cows standing half in the stall and half in the service alley or in the feeding alley. Two sub-classes were adequate for head-to-head stalls. The sub-windows used for sample selection show cows from the head to the tail standing half in the stall and half in the alley (Figure 20f).

Negative images show barn background elements, such as floorings of the alleys and the resting area, operators, and equipments. In detail, for lying and perching behaviours negative images are sub-windows of the images showing empty stalls (Figure 21a); for the other behavioural classes, negative images contain details of floorings as well as other equipments (Figure 21c).

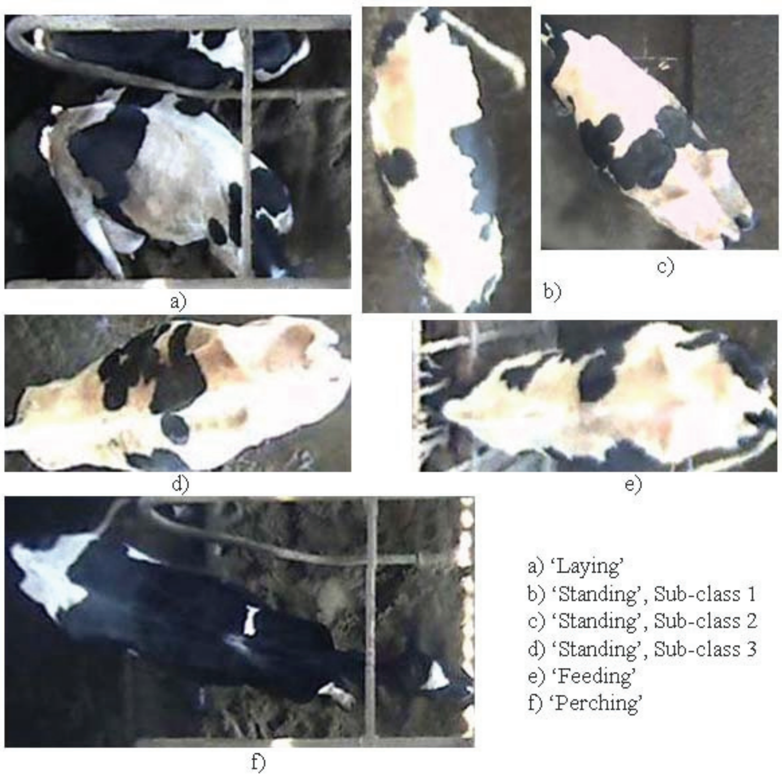


Figure 20 – Examples of positive images used to train Viola & Jones classifiers for different behavioural classes.

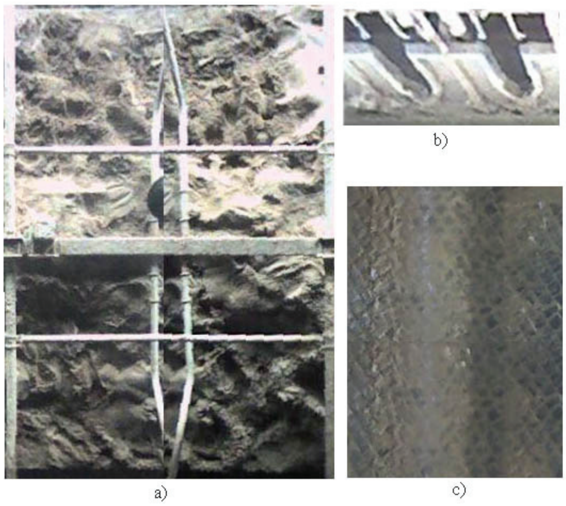


Figure 21 – Examples of negative images used in the training of the Viola & Jones classifiers.

The positive image sample regarding lying and standing behaviours were systematically extracted from the panoramic top-view images by applying 10-min instantaneous scan sampling that previous studies demonstrated to be suitable for the analysis of these behaviours (Mattachini et al., 2011). With regards to the feeding behaviour, the sampling would be limited to specific hours of the day, when the feed was delivered to the animals in the manger. For perching behaviour scan sampling interval should be the same as that of the standing behaviour as it shows cows standing in a particular area of the barn, i.e., the stall.

In the proposed methodology the extended set of features prototype was used (Figure 4).

With regard to the values of *w* and *h*, they should be chosen taking into account that the higher the number of Haar-like features the higher the time required for the training and execution of the classifier. Therefore, by considering the average performance of the CPUs which are currently installed in personal computers (e.g., Intel Core, i3, i5, i7), the number of Haar-like features should not exceed 2×10^6 to obtain a good performance of the CVBS.

The total number *ns* to be built during the training phase should be obtained by means of a trial-and-error technique where an initial small number of stages should be incremented until the desired final values for TPR and FPR were achieved.

For each behavioural class considered, the training tool is executed by providing as input the set of positive images belonging to the considered behavioural class, the set of negative images and the values of the parameters *w*, *h*, *MaxFPR*, *MinTPR*, and *ns*. The output of the training is a formal description of the cascade of stages each constituted by a combination of weak classifiers. The trained classifiers for all considered behavioural class are stored as *XML* files and are subsequently incorporated in the software component of the CVBS used for the detection of cow behavioural activities in the free-stall barn.

3.4.5.2 Software tool for the detection of dairy cow behavioural activities in free-stall barns

A software component of the CVBS must be developed to allow for the simultaneous execution of the modelled classifiers of each considered behavioural class and the sub-classes, which makes it possible to detect cow behavioural activities in the panoramic top-view images of the barn.

The input data required for the execution of the software tool are:

a) one or more set of synchronized images, obtained from the cameras installed in the free-stall barn;
b) a number of geometric transformations, to be applied to the calibrated images obtained from the cameras, to produce the panoramic top-view image of the barn.
c) the trained classifiers for each considered behavioural class which were obtained in the training phase and stored as *XML* files.

Once the input data have been provided, the software tool performs the image mosaicing, by combining camera images into a single composite panoramic top-view. This is the same activity previously performed by the operator but now it is executed in an automatic way because the geometric transformations to be applied to each camera image and the rules for their placement in the new composite image are known.

Subsequently each trained classifier is executed by using as input the panoramic top-view image, as described in the section “execution phase” of the paragraph 3.2.2.2.

When a behavioural class is subdivided into more sub-classes, the outputs of the related classifiers are merged by using geometrical considerations.

Results of the detection process are shown to the CVBS user by superimposing on the panoramic top-view image with coloured rectangular sub-windows surrounding each detected cow (Figure 22).

Different colours are used to distinguish between the types of cow behaviours considered. Furthermore, the software should mainly allow for the optimization of the detection process by selecting the regions of interest where each behaviour should be detected, e.g., stalls for the lying and perching behaviours, feeding alley for the feeding, and feeding and service alleys for standing behaviours.

Finally, the developed software tool must allow the storage of the detection results in a database, a statistical analysis for the detection results, the computation of mostly used behavioural indices, such as cow lying index, cow standing index, and free-stall utilization index.

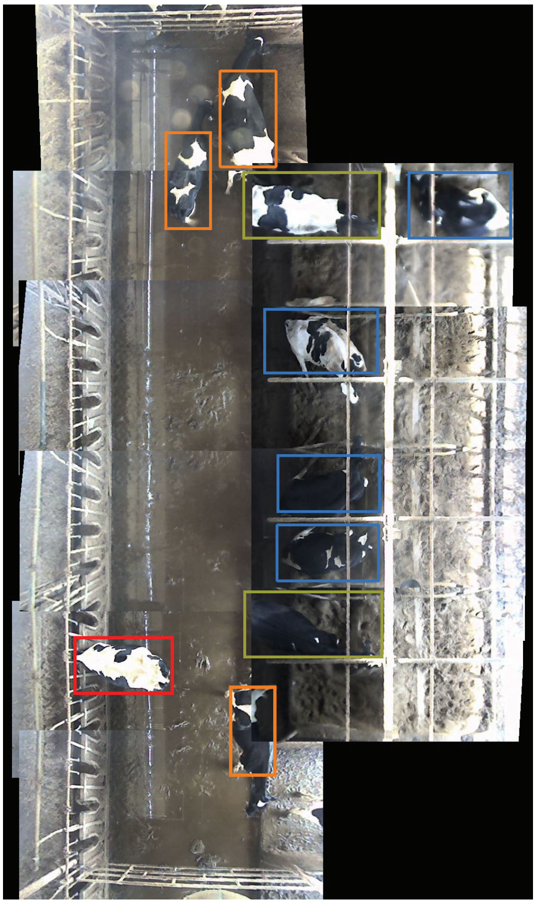


Figure 22 - Results of the detection process obtained by superimposing on the panoramic top-view image of the barn the coloured sub-windows that surround each detected cow behaviour.

3.4.5.3 Test phase

The test phase has to be carried out to assess the accuracy of each classifier. The panoramic top-view images used in this phase must be different from those used for the training and must be selected by following the same method adopted for the sample of the training images. For examples, the panoramic top-view test images could be extracted from the same video sequence used for the training, but staggered of 5 min.

For each behavioural class, starting from a number of positive images, a set of test images must be obtained by overlapping each positive image on pre-established regions of a group of negative images (Figure 23). This last set of images must be chosen with the aim to be significant of the different background conditions such as, for example, those determined by direct sunlight illumination of the breeding environment or other factors that could cause sunlight reflections.

By means of a software that should be specifically developed for this phase, the position of each positive image within the negative one should be determined randomly and stored in a text file. To simulate different image noises that could occur in the breeding environment, each obtained test image should be altered by applying a group of image processing operations (Lefkovits, 2009), i.e., smoothing (blur, median, Gaussian), erosion and dilation.

The software tool for the detection of dairy cow behavioural activities is then executed to search for the cow images within each test image. It compared the detection outputs with the information contained in the text file and assigned the value 1 (hit) to the tested image if the cow image was detected in the right position, i.e., present at the feed barrier, standing, lying, and perching, otherwise it assigned the value 0 (missed). The hit rate (HR) is defined as the ratio between the total number of hit cows and the total number of test images, whereas the miss rate (MR) is 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 is computed by counting cows that were detected out of the pre-established regions. The false positive rate (FPR) is obtained by computing the ratio between the total number of false positives and the total number of test images.

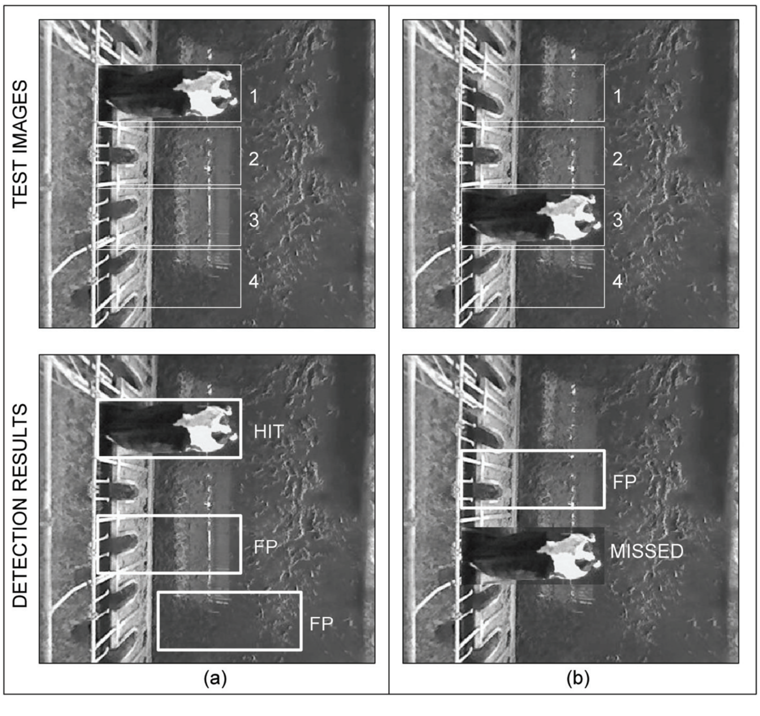


Figure 23 – Example of test images for the test of the feeding classifier.

3.4.5.4 *The validation of the CVBS for the automatic detection of dairy cow behavioural activities in free-stall barns*

The validation of the CVBS for the automatic detection of dairy cow behavioural activities in free-stall barns must be carried out by an accuracy assessment procedure composed of four steps.

The first step of the procedure is the selection of a number of panoramic top-view images that were not used for the training and the test of the classifiers. More than one day of observations must be considered.

A specific software tool (Figure 24) is designed to facilitate the work of an operator, who indicated the locations of all the visible cows and their behaviours through a visual examination of the selected panoramic top-view images.

The software must provide the operator with the following functionalities:

1. The selection of one panoramic top-view image at a time;
2. The display of all the calibrated camera images that constitute the selected panoramic top-view image;
3. The positioning of a graphic element, such as a label, on every calibrated camera image to highlight the cow identifier (if the cow was marked by a symbol), the location of each cow and the corresponding behavioural activity;

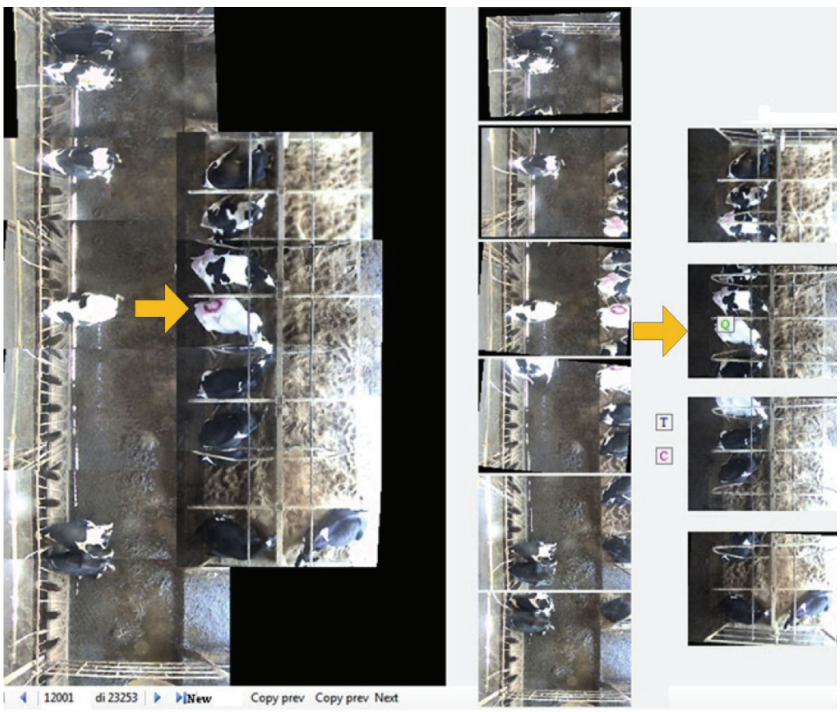


Figure 24 - Interface of the software tool that makes it possible for the CVBS operator to highlight the cow identifier (if the cow was marked by a symbol), the location of each cow and the correspondent behavioural activity.

The software stores in a table of a database both the information provided by the operator and the Cartesian coordinates related to the position of the cow in the panoramic top-view image.

The second step of the accuracy assessment procedure is the execution of the modelled classifiers of each considered behavioural class and related sub-classes to detect cow behavioural activities in the panoramic top-view images of the barn.

Clearly, the input of the software tool must be composed of the same panoramic top-view images examined by the operator in the first step. The outputs of the software tool are automatically stored in a database using, for each detected cow behavioural activity the following attributes: panoramic top-view image number, cow behavioural class, Cartesian coordinates of the corner of the rectangular sub-windows surrounding the detected cow.

Third step of the accuracy assessment procedure is the comparison between the results obtained by using the software tool for the automatic cow behaviour detection and those produced by the operator. In detail, for each cow which was examined by the operator a corresponding cow is searched among the outputs of the CVBS.

A cow behavioural activity that has been recognized by the classifiers as a specific behavioural class and has been correctly associated by the operator to the same behavioural class constitutes a True Positive (TP).

A cow behavioural activity that has been recognized by the classifiers as a specific behavioural class but that has not been associated by the operator to the same class is defined as a False Positive (FP).

A cow behavioural activity that has not been recognized by the classifiers but that has been associated by the operator to a specific behavioural class constitutes a False Negative (FN).

Fourth step of the accuracy assessment procedure involves the use of the comparison results, achieved in the previous step, for the computation of the indices reported below:

- Branching factor (BF): defined as the ratio between FP and TP, it provides information on the number of FPs generated for every TP. Low values of the index show the high capability of the CVBS to distinguish the considered cow behaviour from other behaviours and from background objects.

$$BF = \frac{FP}{TP} \tag{15}$$

- Miss factor (MF): defined as the ratio between FN and TP, it provides information on the number of FNs generated for every TP. Low values of the index show the high capability of the CVBS to detect the considered cow behaviour.

$$MF = \frac{FN}{TP} \tag{16}$$

- Cow detection percentage (CDP):

$$CDP = \frac{TP}{TP + FN} \times 100 \tag{17}$$

yields the percentage of cow behaviours correctly detected over the whole number of the observed ones. It makes it possible to evaluate the ability of the CVBS to detect cow behaviours, yet it does not provide information on the capability to distinguish cow behaviours from the barn background.

- Quality percentage (QP):

$$QP = \frac{TP}{TP + FN + FP} \times 100 \tag{18}$$

provides additional information to that of CDP by also considering the presence of FPs in the barn background.

3.4.6 Second methodology: the automatic identification of dairy cows in free-stall barns (objective 2)

The proposed methodology for cow identification is based on the automatic image recognition of a feature that identifies each cow of the herd. The method assumes the hypothesis that for every set of calibrated and synchronized images coming from the multi-camera system there is at most one cow characterized by that feature. Moreover, each feature is represented by the contour of a shape. If more than one feature of the same cow are detected an identification error occurs.

The methodology involves five sequential phases:

a) Selection of the most appropriate feature to adopt for cow visual identification;
b) Building of the database of the contours of the known feature shapes;
c) Development of a software component of the CVBS for the execution of both the edge detection algorithm and the proposed contour matching;
d) Functioning test of the identification tool for a large image dataset, yet without making use of operator's visual recognitions;
e) Accuracy assessment procedures that allows for the validation of the methodology by comparing the identification data obtained from the application of the automatic software tool with the identification data obtained by an operator.

3.4.6.1 Cow visual identifier

It is theoretically possible to distinguish each cow from the others by using as feature the pattern given by the colour of its coat (Dawkins, 2007). However, this method is not always feasible for all cow behavioural activities because such a pattern might be not completely visible from the cameras, e.g., during the lying activity. Furthermore, the body of some cows might lack of coat pattern, e.g., cow having uniform colour of the coat (Figure 28), or might not have highly distinguishable coat pattern from that of the other cows. This last case occurs for the Holstein dairy cows and, thus, they were marked with hair dye in some research works (DeVries & Von Keyserlingk, 2006; DeVries et al., 2004; DeVries et al., 2003a).

Since the cows considered in this case study are Holstein dairy cows having very similar coat patterns and the aim of the research is to investigate their behavioural activities during daytime, artificial symbols were marked on the coat of each animal by using a natural paint. In general, the selection of the kind of symbol to be used depends on the number of cows to be monitored. Geometrical symbols, e.g., circle, triangle, and square, should be used when a few cows must be identified (Figure 29); otherwise letters or numbers should be adopted (Figure 28).

In this case study, geometrical symbols were used and two different products were tested in order to mark the coat of the cows, i.e., a pencil RAIDL Maxi produced by RAIDEX® and a spray colour produced by Ghislandi & Ghislandi s.r.l. for the signature of livestock.

3.4.6.2 Software tool for the automatic extraction and matching of contours

A software tool for the automatic extraction and matching of contours was developed in order to build the contours database and execute the contour classifier.

Cow identification was carried out by processing the calibrated and synchronized images acquired from the multi-camera system. To avoid the scarce visibility of the symbols due to the mosaicing of the camera images (cf. 3.1.2), in this second methodology the panoramic top-view images were not used. To clarify, Figure 25 shows how the union of the two images (panoramic view) determines the scarce visibility of the symbols.

Consequently, the developed software tool allowed for the simultaneous execution of the contour extraction and contour matching algorithm separately for each camera image acquired by the multi-camera system.

As mentioned in the paragraph 3.3.1.2 the contours extraction in a digital image is done after a number of image processing steps, i.e., smoothing and increasing contrast that aims at the enhancement of the image.

The image smoothing and contrast increasing algorithms are carried out several times on the same image acquired by the same camera. The process is repeated for each camera of the multi-camera systems. In order to achieve good results for the contours extraction, a number of parameters of the algorithms should be selected and adopted in relation to the quality of the image being processed (Table 2). The search for values of the adjustable parameters described in the table is performed through the execution of the following two activities:

a) Collection of images containing at least one marked cow and representing the different quality of the framed image;

b) For each image, an operator manually changes the three parameters reported in the Table 2 until the software tool detects the contour of the symbol. For this reason, the software tool for contour identification must allow the changing of the values of the parameters at runtime as well as display their effects.

Figure 26 shows two input images and the output binary images obtained by applying three different quality improvement processes.

This image enhancement process produces an image data set constituted of a greater number of images than those acquired by the multi-camera system. This involves a slowdown of the identification process, but a better cow identification and increased robustness to false positives.

After creating the image data set, the software tool performs the extraction of contours from each image, compares it with the contours stored in the database, and associates it with one of the geometrical symbols marking the cows. If the software tool identifies several times the same symbol in the image data set, it selects the contour that has the highest index of similarity $\widetilde{Like}$ with a contour stored in the database.

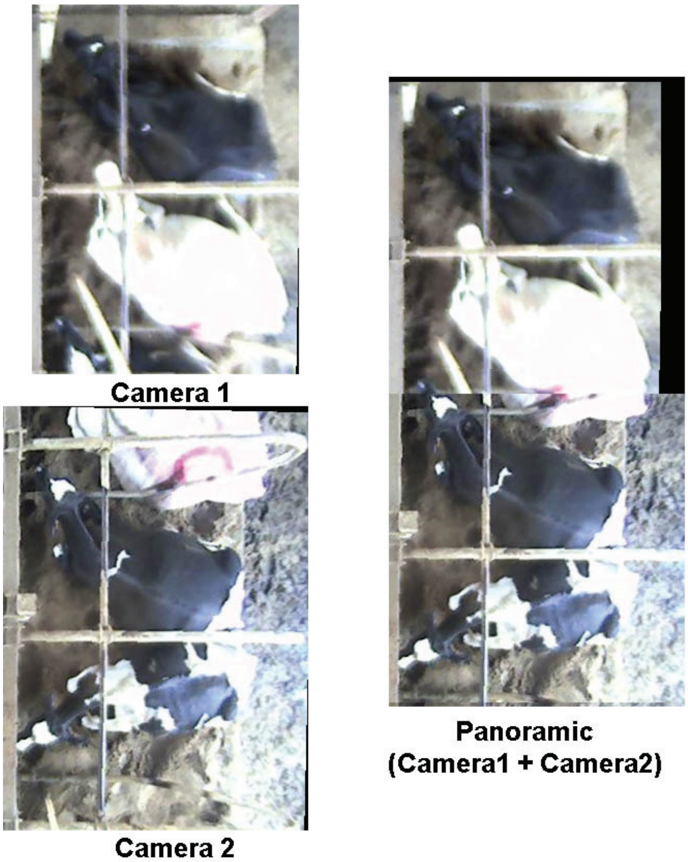


Figure 25 - Mosaic of two camera images acquired from Camera 1 and Camera 2. The union of the two images (panoramic view) determines the scarce visibility of the symbols.

As a result of the cow identification process, for each set of synchronized images, the software tool displays a labelled graphical window surrounding each identified cow. A fully panoramic top-view image shows the composition of the set of camera images to provide the operator with a better vision.

Finally, the software tool for the identification of cows in the barn must allow the storage of the identification results in a database.

Table 2 - Algorithm parameters used to modify input image in order to improve the process of contour detection.

Parameter	Values	Description
Size of the Gaussian filter in the Canny edge detector algorithm	*n* × *n* pixels *n*=3, 5, 7, 9, 11, …	Smaller values, i.e., *n*=3, 5, allow detection of small edges whereas a larger filter, i.e., *n*=9, 11, allow detection of larger edges.
Blur	'Yes', 'No'	Only if Blur is set to 'Yes' then smoothing process is performed to reduce the image noise with the consequent reduction of the details.
Histogram equalization	'Yes', 'No'	Only if histogram equalization is set to 'Yes' then the transformation of the gray levels of the image so that the histogram of the resulting image is equalized to become a constant is performed and as a result the overall contrast of the images is increased.

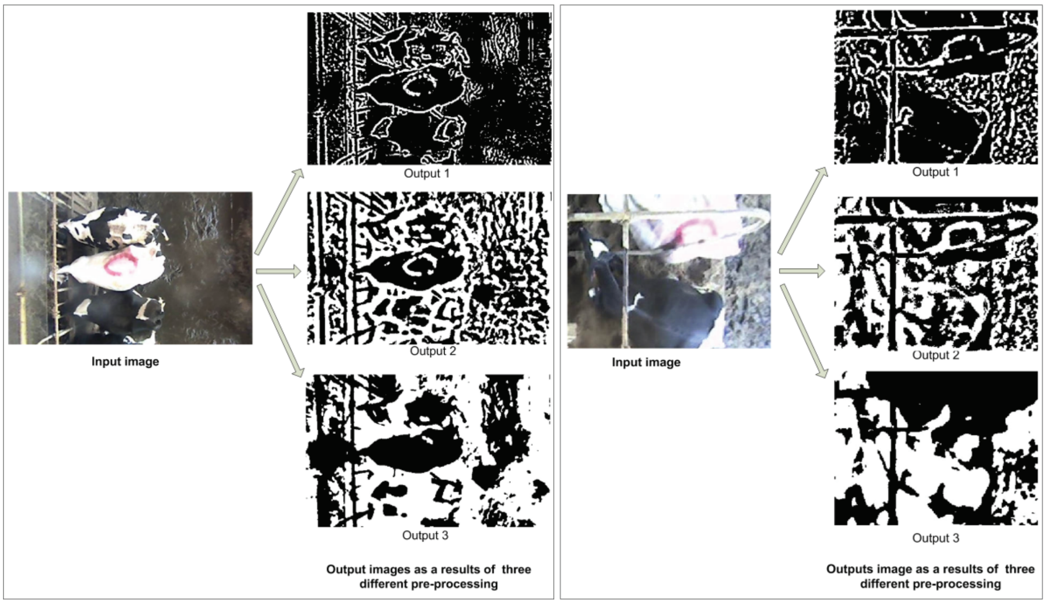


Figure 26 -Two input images and output binary images obtained by applying three different quality improvement processes.

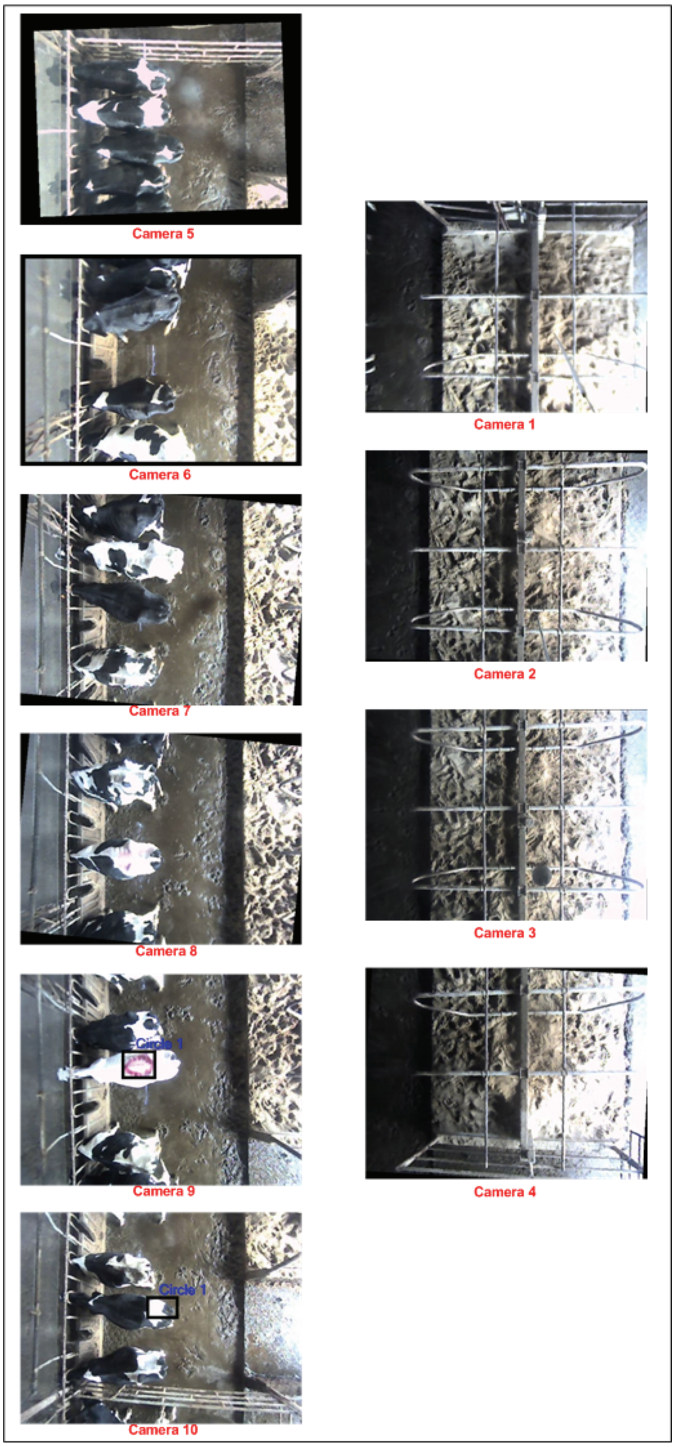


Figure 27 – Results of the identification process obtained by superimposing on each camera image the coloured sub-windows that surround the identified cow. In camera 9 a true positive was detected, whereas in camera 10 a false positive was detected.

3.4.6.3 Construction of the database constituted by known contours of cow visual identifier

A set of positive images, i.e., images where marked cows are present, must be collected. Each positive image has to contain one marked cow among those considered. In the selection of the positive images, the different behavioural activities of the cows during the day, i.e., feeding, lying, perching, standing should be considered (Figure 30) because the symbol applied in the coat might assume a different shape, or might not be completely visible, at varying of the behavioural activity.

Figure 28 - Beef cattle marked by using numbers.

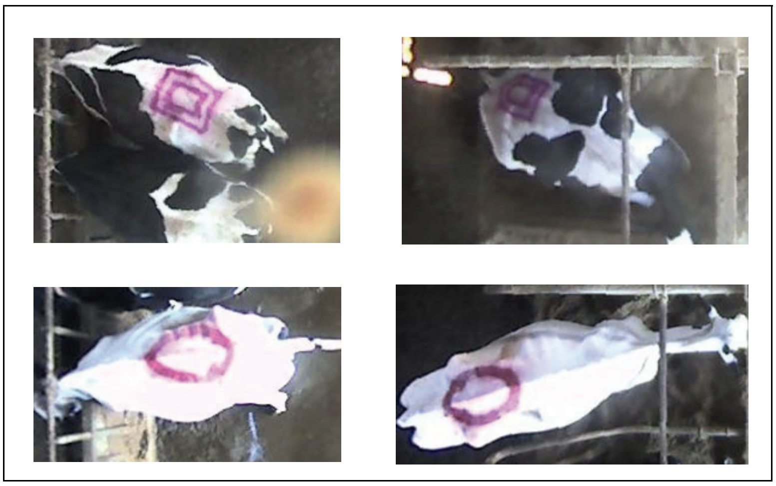


Figure 29 - Example of geometrical symbols marked on the coats of the dairy cows monitored in the case study.

The positive image sample has to be systematically extracted from the frames by applying 10-min instantaneous scan sampling that previous studies demonstrated to be suitable for the analysis of these behaviours (Mattachini et al., 2011). With regards to the feeding behaviour, the sampling would be limited to specific hours of the day, when the feed was delivered to the animals in the manger. For perching behaviour scan sampling interval should be the same as that of the standing behaviour since it shows cows standing in a particular area of the barn, i.e., the stall.

The obtained set of positive images must be subdivided into a number of classes that corresponds to the number of symbols used to mark the cows. Each class may be subdivided into sub-classes because of the different shape assumed in the camera images by the marked symbols during the different behavioural activities (Figure 30). Subsequently, each positive image is processed by the software tool described in section 3.4.6.2 and accordingly to the guidelines reported in 3.3.2.1.

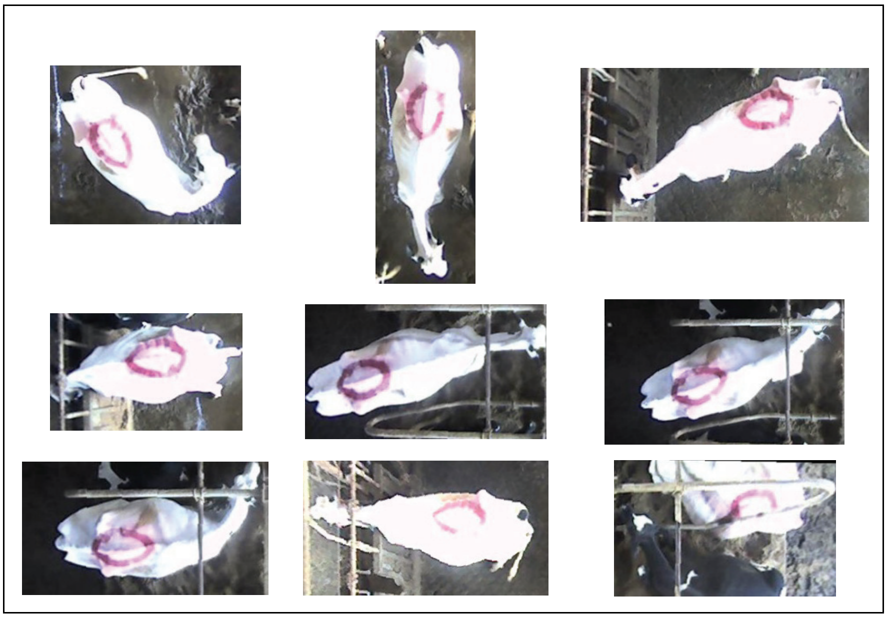


Figure 30 – Examples of positive images belonging to the same class. The figure shows the different shapes assumed by the adopted symbols in relation to the different behavioural activities.

The list of Cartesian coordinates obtained from the contour extraction (Figure 31) together with a numerical identification code of the class is stored in the database. In general, the database could be composed of a simple text file up to the most sophisticated DBMS. In the case study, a structured Extensible Markup Language (XML) file was used because it is a simple standard database that does not require the use of proprietary applications.

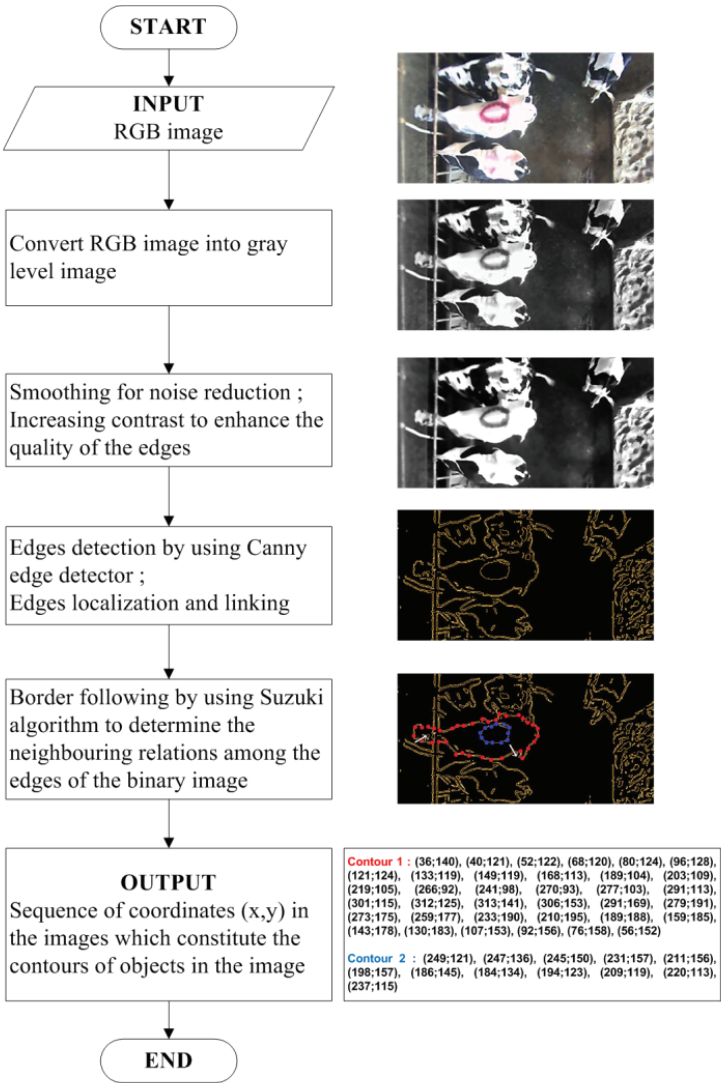


Figure 31 – Example of contour extraction processing in a positive image

3.4.6.4 Test phase

This phase of the second methodology is analogous to that described in section 3.4.5.3. For each cow to be identified, starting from a number of positive images, a set of test images was obtained by overlapping each positive image on pre-established regions of a group of negative images (Figure 32). This last set of images must be chosen with the aim to be representative of the different background conditions such as, for example, those determined by direct sunlight illumination of the breeding environment or other factors that could cause sunlight reflections.

The test images were processed by using the software for the automatic extraction and matching of contours described in section 3.4.6.2. The results of this phase were used in order to compute the HR, MR and FPR indices.

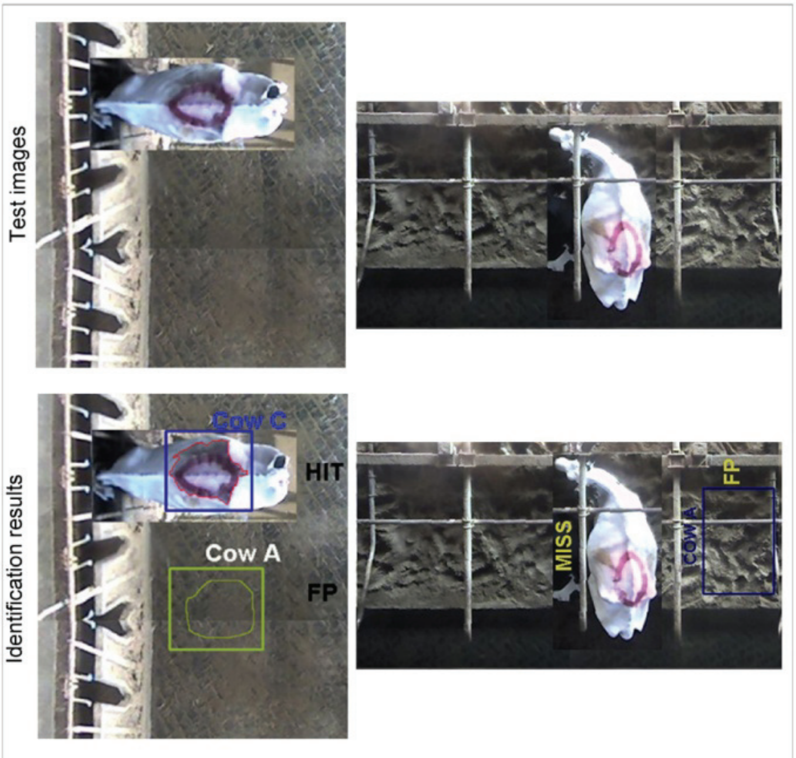


Figure 32 – Example of test images for testing the cow identification software.

3.4.6.5 *The validation of the CVBS for the automatic identification of dairy cows in free-stall barns*

In the first step of the accuracy assessment procedure an operator uses the software tool described in section 3.4.5.2 with the aim to store information on marked cows, i.e., the cow identifier, the location and the behavioural activity. With regard to the selection of panoramic top-view images, they must be chosen among those not used for the test and for the building of the contour database of symbols. More than one day of observations must be considered in the selection of the panoramic top-view images.

The second step of the accuracy assessment procedure is the automatic identification of marked cows by using the automatic identification tool described in section 3.4.6.2. Obviously, the input of the software tool must be the same calibrated camera images used by the operator in the first step. The outputs of the software tool are automatically stored in the database by using for each detected contour the following attributes: representation of the contour through a sequence of Cartesian coordinates and the identifier of the cow to which the contour has been assigned.

Third step of the accuracy assessment procedure is the comparison between the results obtained by using the software described in section 3.4.6.2 and those produced by the operator.

In detail, for each cow identified within a panoramic top-view image by the operator in the first step of the procedure, a contour C is searched among the outputs of the software tool for the automatic cow identification (cf. 3.4.6.2). The properties required for C are the following:

1. The cow identifier associated to C by the software tool must correspond to the cow identifier established by the operator;
2. The Euclidean distance between the centroid of C and the position of the cow established by the operator must be lower than a fixed threshold.

The centroid $G:(G_x, G_y)$ of a closed contour available in the form of Cartesian coordinates $c(t):(x(t), y(t))\ \ t = 0,1,\ldots,L-1$ is calculated with the following formulas (Yang, Kidiyo, & Rosin, 2008):

$$\begin{cases} G_x = \dfrac{1}{6A}\displaystyle\sum_{i=0}^{L-1}\left[x(i)+x(i+1)\right]\left[x(i)\,y(i+1)-x(i+1)\,y(i)\right] \\ G_y = \dfrac{1}{6A}\displaystyle\sum_{i=0}^{L-1}\left[y(i)+y(i+1)\right]\left[x(i)\,y(i+1)-x(i+1)\,y(i)\right] \end{cases} \tag{19}$$

where

– the point $\left(x(L), y(L)\right)$ is assumed to be the same as $\left(x(0), y(0)\right)$ since the contour is a closed polyline

– A is the contour's area given by: $A = \dfrac{1}{2}\displaystyle\sum_{i=0}^{L-1}\left[x(i)\,y(i+1)-x(i+1)\,y(i)\right]$

If a contour with the two mentioned properties is identified then C is labelled as a TP otherwise it is noted as not found and, therefore, a FN is obtained for the accuracy assessment.

The contours that were identified by the software for the automatic identification but that were not been associated with any cow identified by the operator constitute the FPs of the accuracy assessment.

The fourth step of the accuracy assessment procedure involves the computation of the quality indices reported in section 3.4.5.4.

4 RESULTS

4.1 First methodology: the automatic detection of cow behavioural activities in free-stall barns (objective 1)

The availability of top-view panoramic images of the study area allowed an operator to perform the manual extraction of positive images that showed the cow body shape and negative image samples that showed background elements of the barn such as alleys and resting area flooring, operators and equipment. Samples have been subdivided into behavioural classes and subclasses to be used for training and testing each classifier.

The building of the classifiers used for the detection of the dairy cow behavioural activities was obtained through the execution of the training tool for each considered behavioural activity.

The defined behavioural classes and subclasses have lead to the modelling of the following classifiers:

a) Two classifiers, called classifier *1a* and classifier *1b*, to detect cow lying behaviour;
b) One classifier, called classifier *2*, to detect cow feeding behaviour;
c) Three classifiers, called classifier *3a*, classifier 3b, and classifier *3c*, to detect cow standing behaviour;
d) Two classifiers, called classifier *4a*, and classifier *4b*, to detect cow perching behaviour.

In the training phases the same values of *MaxFPR, ns,* and TPR_{cas} were assigned to each classifier. The maximum number of the stages, *ns*, was fixed to 30. The value of *MaxFPR* was set equal to 0.5 because it made it possible to obtain a rapid decrease of the FPR_{cas} value as the number of stages increased. TPR_{cas} value was set at least equal to 0.90 because it was assumed a maximum value of the false negative rate, i.e., percentage of cows that were not detected, equal to 0.10. Taking into account these common settings of the training parameters, the computation of the *MinTPR* by means of the relation (8) yielded a value of 0.9965 for each stage of the classifiers. The *z* factor used to increase the dimensions of the rectangular sub-window was set to 5%, for each classifier.

Other specific training parameters were chosen in relation to the specificity of the cow behavioural activity to be detected, i.e., the number of positive images, the dimensions of the rectangular window used to cut out each cow image from the panoramic top-view images, the number of negative images, and the dimensions of the sliding window.

The panoramic images used to train the classifiers were extracted from video-recordings acquired between the 1st and the 7th of August, 2011, from the 6:00 a.m. and the 7:00 p.m.

By applying 10-min instantaneous scan sampling, a total amount of 84 frames were extracted each day for lying, standing, and perching behaviours. For feeding behaviour, the sampling regarded only the two hours after the first milking and one hour in the first afternoon, when the feed was moved closer to the cows. Consequently, for such behaviour 18 frames each day were extracted from the video sequence.

In the test phase of each classifier, panoramic top-view images were extracted from video-recordings acquired between the 1st and the 7th of August, 2011 from 6:00 a.m. to 8:00 p.m., by applying a 10 minutes sampling, as it was done in the training phase. However, the instant of acquisition was delayed by 5 minutes. Furthermore, in order to test the quality of the classifications in different conditions of ambient light, the average daily trend of the pixel brightness average values in the resting area and in the feeding alley was obtained (Figure 33 and Figure 34). The pixel brightness average values in the resting area and in the feeding alley were computed every 10 minutes of each day of the considered week. The computation was carried out by using the commercial software ERDAS Imagine®.

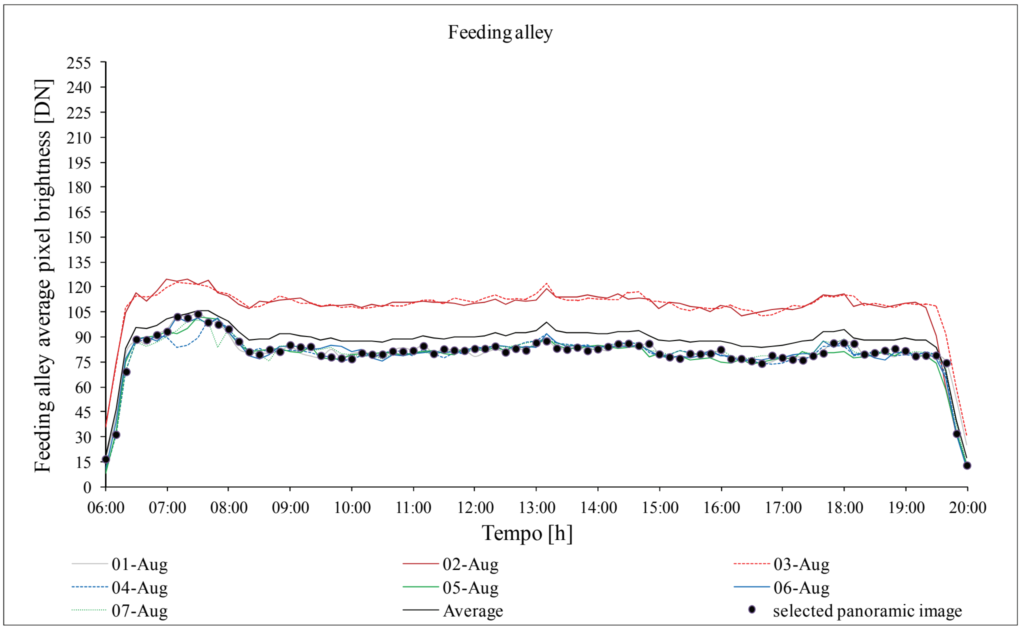
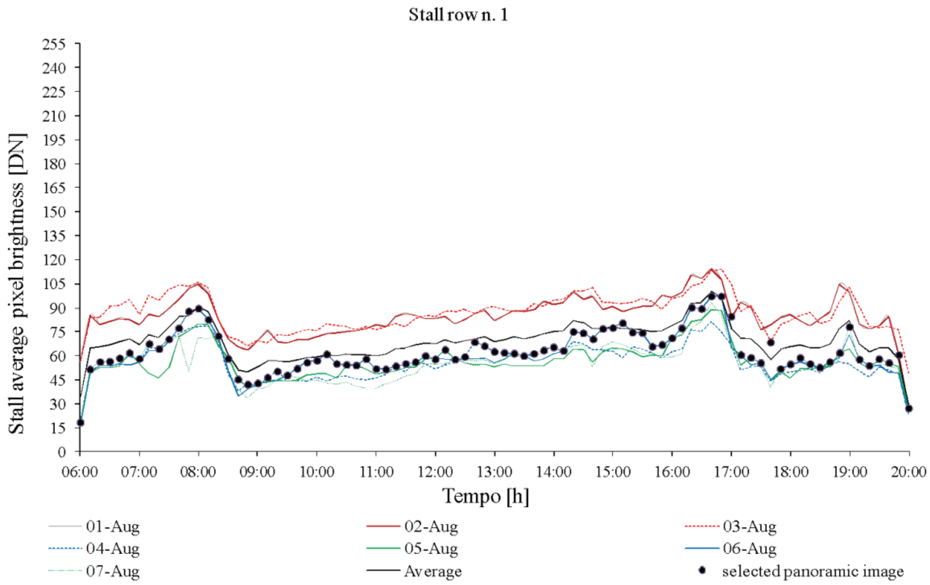
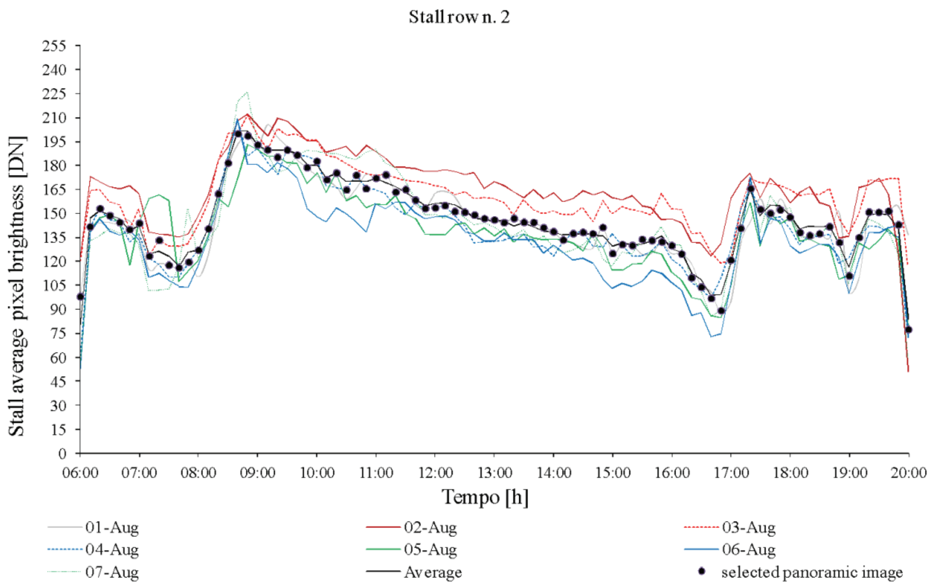


Figure 33 - Averages of the pixel brightness values of the feeding alley images computed for each acquisition instant of the seven days considered; weekly means of the averages of the pixel brightness values; panoramic images selected to produce the sequence of images.



(a)



(b)

Figure 34 - Averages of the pixel brightness values of the images of stall row n. 1 (a) and stall row n.2 (b) computed for each acquisition instant of the seven days considered; weekly means of the averages of the pixel brightness values; panoramic images selected to produce the sequence of images.

For the classifiers used to detect the feeding and standing behaviours, i.e., *2*, *3a*, *3b*, and *3c*, a sequence of panoramic top-view images was produced by selecting every 10 minutes the panoramic top-view images wich had the pixel brightness average value closest to the weekly mean in the feeding alley (Figure 33).

Likewise, for the classifiers used to detect the lying and perching behaviours, i.e., *1a*, *1b*, *4a*, and *4b,* a sequence of panoramic top-view images was produced by selecting every 10 minutes the panoramic top-view images having in the resting area the pixel brightness average value closest to the weekly mean (Figure 34).

Finally, in the validation phase, the accuracy assessment procedure was carried out by using the implemented software tool that executed simultaneously the classifiers on a set of selected panoramic top-view images that were not used in the training and test phases. In detail, the accuracy assessment procedure was carried out by using the video-recordings acquired between the 8th and the 14th of August, 2011, from 6:00 a.m. to 8:00 p.m. The panoramic images were selected at 10-minute sampling intervals. A number of 589 panoramic top-view images were analyzed by all the classifiers and by the operator. The results of the training, test, and accuracy assessment phases executed for each classifier are described in the following paragraphs.

4.1.1 Lying behaviour classifiers

In the training phase of the classifiers *1a* and *1b*, 826 and 319 positive images were selected, respectively (Table 3). The positive images were obtained by extracting image sub-sets corresponding to all the rectangular areas of 224×140 pixel that showed stalls occupied by the cows in each panoramic top-view image (Table 3).

The number of negative images was equal to 600 for both the classifiers and was obtained by extracting areas identifying empty stalls, i.e., unoccupied by the cows.

The sliding window was set to 40×25 pixel in order to maintain the same aspect ratio equal to 1.6 of the positive images extracted from the panoramic top-view images. For both the classifiers the number of the negative image samples was equal to 25×10^7.

The training of the two classifiers ended at the 27th *stage* when FPR_{cas} values reached the order of magnitude 10^{-7}. The results of the training were reported in Table 4.

In the test phase, , 346 positive images of 224×140 pixels were extracted from the obtained panoramic top-view image sequence. In particular, 274 images obtained by selecting all the cows lying in the stalls in the first row, were utilized to test the classifier *1a*. The other 72 images, obtained by selecting all the cows lying in the stalls in the second row, were utilized to test classifier *1b*.

With regards to the number of negative images for each day, two were selected at 6:30 a.m. and 6:30 p.m., when the cows were in the milking area, and other two images were selected slightly before and just after the cleaning of the feeding alley, carried out at approximately 8:00 a.m., when the cows were confined within an area which includes the service alley and the second row of stalls (Figure 19).

The image overlay and the alteration operations produced 4384 test images for the classifier *1a* and 1152 for classifier *1b.* The results of the test were reported in Table 5.

In the accuracy assessment phase 2281 images of cows lying in the cubicles were present. The overall number of cows correctly detected was 2088 and, therefore, the number of unclassified cows was 193. The number of false positives given by the system was 167. The accuracy indices computed on the basis of these values were: BF=0.08; MF=0.09; CDP=0.92; QP=0.85.

Table 3 - Characteristics of the positive and negative images that were selected in order to constitute the image samples used in the training phase.

Classifier	Positive images		Negative images		Sliding windows	Samples	
		Dimensions		Average dimensions	Dimensions	Positives (w×h)	Negatives (w×h)
	n.	*W*×*H*	*n.*	*W'*×*H'*	*w*×*h*	*n.*	*n.*
1a	826	224×140	600	230×144	40×25	826	24.84×10^7
1b	319	224×140	600	230×144	40×25	319	24.84×10^7

Table 4 - Characteristics of the trained classifiers *1a* and *1b*.

Classifier	Stage	Processed negative images	TPR_{cas}	FPR_{cas}
	Results of the training phase			
	n.	n.		
1a	27	15.21×10^7	0.902	3.09×10^-7
1b	27	6.31×10^7	0.920	7.93×10^-7

Table 5 - Test results of the trained classifiers *1a* and *1b*.

Classifier	N. of images	TP	FN	FP	HR	MR	FPR
	Results of the test phase						
		n.	n.	n.			
1a	4384	3838	546	358	0.88	0.12	0.08
1b	1152	1000	152	92	0.87	0.13	0.08

4.1.2 Feeding behaviour classifier

The training of the classifier *2* required 656 positive images and 384 negative images (Table 6). The positive images were obtained by extracting the image sub-sets corresponding to all the rectangular areas of 227×102 pixel that showed single cow at the feeding barrier in each panoramic top-view image. Likewise, the negative images were obtained by extracting areas corresponding to empty feeding locations, i.e., unoccupied by the cows having dimension equal to 230×100.

The sliding window was set to 40×18 pixel in order to maintain the same aspect ratio equal to 2.2 of the positive images extracted from the panoramic ones. For the feeding classifier *2*, 656 positive image samples and about 11×10^7 negative image samples of 40×18 pixels were obtained.

Also the training of the feeding detector terminated at the 27th *stage* when FPR_{cas} values reached the order of magnitude 10^{-7}. The results of the training were reported in Table 7.

In the test phase, from the obtained panoramic top-view image sequence, 368 positive images of 227×102 pixel were extracted.

As for the classifiers *1a* and *1b*, the negative images were selected within panoramic top-view images acquired at 6:30 a.m. and 6:30 p.m., when the cows were in the milking area, and at 8:00 a.m. when the cows were confined within an area which includes the service alley and the second row of stalls (Figure 19).

The image overlay and the alteration operations involved 5888 test images for the classifier *2*. The results of the test were reported in Table 8.

In the accuracy assessment phase 2217 images of cow presence at the feed barrier were present. The overall number of cows correctly detected was 1922 and, therefore, the number of unclassified cows was 295. The number of false positives given by the system was 157. The accuracy indices computed on the basis of these values were: BF=0.08; MF=0.15; CDP=0.87; QP=0.81.

4.1.3 Standing behaviour classifiers

In the training phase of the classifiers *3a*, *3b* and *3c*, the number of positive image was equal to 646, 651 and 430, respectively (Table 9). The dimensions of the rectangular areas that showed cows standing in the feeding alley in each panoramic top-view image were 100×220 pixel, 220×100 pixel, and 185×185 pixel for the classifiers *3a*, *3b* and *3c*, respectively. The number of negative images was equal to 708 for all the classifiers and was obtained by extracting areas identifying empty areas in the feeding alley.

The sliding windows for the three classifiers were set to 18×40 pixel, 40×18 pixel, and 40×40 pixel for the classifiers *3a*, *3b* and *3c*, respectively. Such dimensions were chosen in order to maintain the aspect ratios equal to 0.45, 2.2, and 1 for the classifiers *3a*, *3b* and *3c*, respectively. In all, for the standing classifier *3a*, 646 positive image samples and about 57×10^7 negative image samples of 18×40 pixel were obtained; for the standing classifier *3b*, 651 positive image samples and about 57×10^7 negative image samples of 40×18 pixels were obtained, whereas for the standing classifier *3c*, 430 positive image samples and about 43×10^7 negative image samples of 40×40 pixels were obtained.

The training of the classifiers terminated at the 30th *stage* when the maximum number of stages was reached. The FPR_{cas} values reached the order of magnitude 10^{-5} for all three classifiers. The results of the training were reported in Table 10.

In the test phase, from the obtained panoramic top-view image sequence, 174 positive images were extracted. In particular, 58 images obtained by selecting all the cows standing toward the direction of the longitudinal axis of the barn were utilized to test the classifier *4a*, 71 images obtained by selecting all the cows standing crosswise the direction of the longitudinal axis of the barn were utilized to test the classifier *4b,* 45 images, obtained by selecting all the cows standing oblique to the direction of the longitudinal axis of the barn were utilized to test the classifier *4c.*

As for the lying classifiers and the feeding classifier, the negative images were selected within panoramic top-view images acquired at 6:30 a.m. and 6:30 p.m., when the cows were in the milking area, and at 8:00 a.m. when the cows were confined within an area which includes the service alley and the second row of stalls (Figure 19).

Table 6 - Characteristics of the positive and negative images that were selected in order to constitute the image samples used in the training phase.

Classifier	Positive images		Negative images		Sliding windows	Samples	
		Dimensions		Average dimensions	Dimensions	Positives ($w\times h$)	Negatives ($w\times h$)
	n.	$W\times H$	*n.*	$W'\times H'$	$w\times h$	*n.*	*n.*
2	656	227×102	384	230×100	40×18	656	10.86×10^7

Table 7 - Characteristics of the trained classifier *2*.

Results of the training phase				
Classifier	Stage	Processed negative images	TPR_{cas}	FPR_{cas}
	n.	n.		
2	27	6.05×10^7	0.925	7.67×10^{-7}

Table 8 - Test results of the trained classifiers *2*.

Results of the test phase							
Classifier	N. of images	TP	FN	FP	HR	MR	FPR
		n.	n.	n.			
2	5888	5210	678	455	0.88	0.12	0.08

Image overlay and the alteration operations involved 928 test images for the classifier *3a,* 1136 test images for the classifier *3b,* and 720 test images for the classifier *3c*. The results of the test were reported in Table 11.

In the accuracy assessment phase 1,286 images of cows standing were present. The overall number of cows correctly detected was 1,102 and, therefore, the number of unclassified cows was 184. The number of false positives given by the system was 190. The accuracy indices computed on the basis of these values were: BF=0.17; MF=0.17; CDP=0.86; QP=0.75.

Table 9 - Characteristics of the positive and negative images that were selected in order to constitute the image samples used in the training phase.

Detector	Positive images		Negative images		Sliding windows	Samples	
		Dimensions		Average dimensions	Dimensions	Positives ($w×h$)	Negatives ($w×h$)
	n.	$W×H$	*n.*	$W'×H'$	$w×h$	*n.*	*n.*
3a	646	100×220	708	225×225	18×40	646	$57.43×10^7$
3b	651	220×100	708	225×225	40×18	651	$57.43×10^7$
3c	430	185×185	708	225×225	40×40	430	$43.47×10^7$

Table 10 - Characteristics of the trained classifiers *3a 3b* and *3c*.

Results of the training phase				
Detector	Stage	Processed negative images	TPR_{cas}	FPR_{cas}
	n.	n.		
3a	30	$8.51×10^7$	0.904	$4.70×10^{-5}$
3b	30	$12.85×10^7$	0.901	$3.12×10^{-5}$
3c	30	$11.32×10^7$	0.908	$1.30×10^{-5}$

Table 11 - Test results of the trained classifiers *3a 3b* and *3c*.

Results of the test phase							
Detector	N. of images	TP	FN	FP	HR	MR	FPR
		n.	n.	n.			
3a	928	805	123	98	0.87	0.13	0.11
3b	1136	980	156	101	0.86	0.14	0.09
3c	720	642	78	62	0.89	0.11	0.09

4.1.4 Perching behaviour classifiers

The training phases of the perching classifiers *4a* and *4b* required 174 and 80 positive images, respectively (Table 12). The positive images were obtained by extracting from each panoramic top-view image all the image sub-sets corresponding to the rectangular areas of 308×140 pixel that showed cows standing half in the stall and half in the feeding alley, for classifier *4a*, and cows standing half in the stall and half in the in the service alley, for classifier *4b*.

The number of negative images was equal to 600 for both the classifiers and was obtained by extracting areas identifying empty stalls, i.e., unoccupied by the cows, and portions of the adjacent alleys (Table 12).

The sliding window was set to 55×25 pixel in order to maintain the same aspect ratio equal to 2.2 of the positive images extracted from the panoramic top-view images. For both the classifier the number of the negative image samples was equal to 33×10^7 .

The training of the classifiers terminated at the 30th *stage* when the maximum allowed number of stages was reached. In correspondence the FPR_{cas} values reached the order of magnitude 10^{-6} for both classifiers (Table 13). The results of the training are reported in Table 13.

Table 12 - Characteristics of the positive and negative images that were selected in order to constitute the image samples used in the training phase.

Classifier	Positive images		Negative images		Sliding windows	Samples	
		Dimensions		Average dimensions	Dimensions	Positives ($w \times h$)	Negatives ($w \times h$)
	n.	$W \times H$	*n.*	$W' \times H'$	$w \times h$	*n.*	*n.*
4a	174	308×140	600	310×144	55×25	174	32.62×10^7
4b	80	308×140	600	310×144	55×25	80	32.62×10^7

In the test phase, from the obtained panoramic top-view image sequence, 62 positive images of 308×140 pixels were extracted. In particular, 31 images obtained by selecting all the cows perching in the stalls in the first row were utilized to test the classifier *4a*. The other 31 images, obtained by selecting all the cows perching in the stalls in the second row, were utilized to test classifier *4b*.

As for the other classifiers, the negative images were selected within panoramic top-view images acquired at 6:30 a.m. and 6:30 p.m., when the cows were in the milking area, and at 8:00 a.m. when the cows were confined within an area which includes the service alley and the second row of stalls (Figure 19).

The image overlay and the alteration operations involved 496 test images for the classifier *4a* and 496 for classifier *4b*.The results of the test are reported in Table 14.

In the accuracy assessment phase 553 images of cows perching were present. The overall number of cows correctly detected was 494 and, therefore, the number of unclassified cows was 59. The number of false positives given by the system was 49. The accuracy indices computed on the basis of these values were: BF=0.10; MF=0.12; CDP=0.89; QP=0.82.

Table 13 - Characteristics of the trained classifiers *4a* and *4b*.

Results of the training phase				
Classifier	Stage	Processed negative images	TPR_{cas}	FPR_{cas}
	n.	n.		
4a	30	11.04×10^{7}	0.903	4.53×10^{-6}
4b	30	10.21×10^{7}	0.905	7.72×10^{-6}

Table 14 - Test results of the trained classifiers *4a* and *4b*.

Results of the test phase							
Classifier	N. of images	TP	FN	FP	HR	MR	FPR
		n.	n.	n.			
4a	496	438	58	38	0.88	0.12	0.08
4b	496	430	66	43	0.87	0.13	0.09

4.2 Second methodology: the automatic identification of dairy cows in free-stall barns (objective 2)

As it has been described in paragraph 3.4.6, the proposed method for the automatic identification of dairy cows in free-stall barns requires the marking of cows that need to be identified by drawing an artificial symbol in the coat with natural paints or hair dye. Since the objective of the study was to verify the feasibility of the proposed method of identification only few animals were marked.

In a first trial, a pencil RAIDL Maxi produced by RAIDEX® was used to mark three cows with triangle, square and circle shapes (Figure 35a). However, after only two days, the marks on the coat of the cows were no longer clearly visible. This happened after that the marked cows had used the cooling system composed of showers for direct wetting or when they had used the tilting brush (Figure 35b).

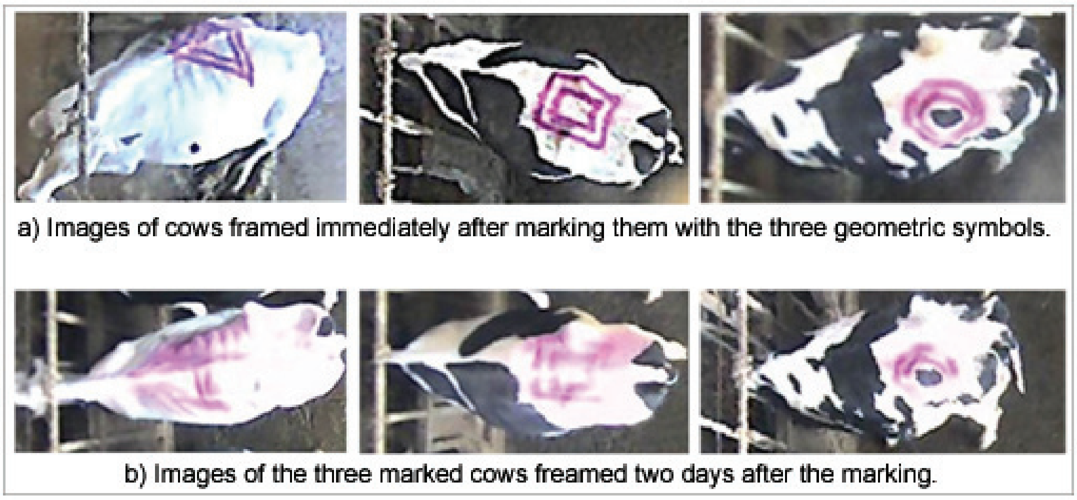


Figure 35 - Images of marked cows in the first trial.

In the second trial, a spray for the signature of livestock produced by Ghislandi & Ghislandi s.r.l. was used. Only one cow was marked by drawing a circle on the coat (Figure 36). For four week the symbol on the coat remained clearly visible.

The availability of the synchronized camera images that frame both the resting area and the feeding alley allowed the manual extraction of positive images that show the cow marked with the circle, to be performed by an operator.

For this kind of artificial symbol, two contour sub-classes were defined:

a) The sub-class *1a* that contained the contours extracted from images framing the cow during the feeding, standing and perching behavioural activities;

b) The sub-class *1b* contained the contours extracted from images framing the cow during the laying behavioural activity.

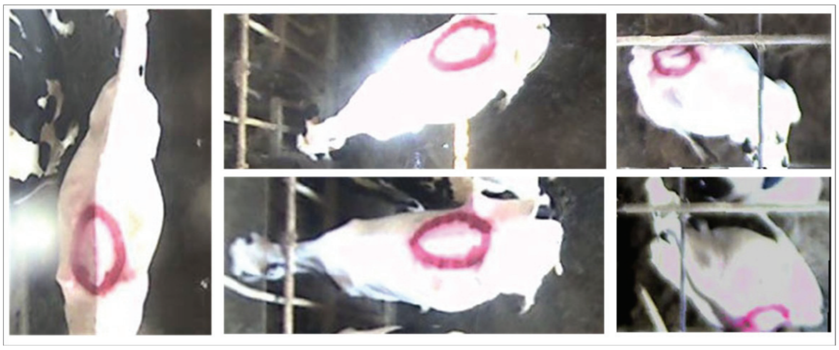


Figure 36 - Images of the marked cow in the second trial.

In the trial the definition of two sub-classes was required because when the marked cow was in feeding, standing or perching the circular symbol applied in the coat was completely visible in the camera images, whereas when the cow was lying in a stall the circular symbol was subject to deformations or was only partially visible.

The parameters of the contours detection algorithms were determinated by means of the trial and error technique. In details, the number L of points which constitute the contour of the objects was fixed to 40, the perimeter of each accepted contour was bounded between 150 and 300 pixels for the sub-class *1a* and between 80 and 200 pixels for the sub-class *1b*, whereas the area enclosed by the perimeter was bounded between 400 and 600 pixels for the sub-class *1a* and between 200 and 400 pixels for the sub-class *1b*. The threshold value $Th_{\widetilde{Like}}$ was set equal to 0.85 for both the sub-classes.

The parameters' values of the image enhancement algorithm that were tested in order to modify each input camera image with the aim of improving the contour detection process, were grouped into 16 different settings reported in Table 15.

Table 15 – Parameters settings of the of the image enhancement algorithm

Settings number	Gaussian filter	Blur	Histogram equalization	Settings number	Gaussian filter	Blur	Histogram equalization
1	3×3	'No'	'No'	9	11×11	'Yes'	'Yes'
2	3×3	'No'	'Yes'	10	13×13	'No'	'Yes'
3	3×3	'Yes'	'Yes'	11	15×15	'Yes'	'Yes'
4	5×5	'No'	'Yes'	12	21×21	'Yes'	'Yes'
5	5×5	'Yes'	'Yes'	13	25×25	'Yes'	'Yes'
6	7×7	'No'	'Yes'	14	31×31	'Yes'	'Yes'
7	7×7	'Yes'	'Yes'	15	45×45	'Yes'	'Yes'
8	9×9	'No'	'Yes'	16	61×61	'Yes'	'Yes'

The synchronized and calibrated camera images used to build the database of the known contours of the circle used to mark the cow were extracted from video-recordings acquired between the 26th and the 30th of May, 2012, from the 7:00 a.m. to the 7:00 p.m. For each day, by applying 10-min instantaneous scan sampling, a set of 72 calibrated camera images was obtained where the marked cow was present. Therefore, an amount of 360 images were extracted.

The positive images were obtained by extracting the image sub-sets corresponding to the most significant rectangular areas that showed the marked cow, from each selected image.

The records of the contour database were populated by an operator, who used the developed software tool for the automatic identification of dairy cows. For each positive image and for each parameter's value used in the image enhancement process the tool showed the detected contours. Therefore, the operator was able to select the contour related to the circular symbol marked on the coat of the cow as well as the sub-class. This contour was stored in the database if the values of the index of similarity $\widetilde{Like}$ computed between the extracted contour and those present in the database was lower than 0.95.

In Table 16 the characteristics of the contour database were reported.

Table 16 – Characteristics of the contour database

Characteristics of the contour database			
Sub-class	**Positive images**	**Positive images after the enhancement processes**	**Contours extracted**
	n.	n.	n.
1a: Cow marked with circular symbol (feeding, standing, perching)	60	960	410
1b: Cow marked with circular symbol (laying)	45	720	280

In the test phase, camera images were extracted from video-recordings acquired between the 26th and the 30th of May, 2012 from 7:00 a.m. to 7:00 p.m., by applying a 10 minute sampling, as it was done in the building of the database. However, the instant of acquisition was delayed 5 minutes in comparison to that employed in the training phase.

A number of 275 positive images were extracted. In particular, 171 images, obtained by selecting the marked cow during the feeding, standing and perching behaviours, were utilized to test the identification of the contours belonging to the sub-class *1a*. The other 104 images, obtained by selecting the marked cow during the laying behaviour, in the first or in the second stall row, were utilized to test the identification of the contours belonging to the sub-class *1b*.

With regards to the number of negative images, two sets of synchronized camera images were selected at 6:30 a.m. and 6:30 p.m., when the cows were in the milking area, and at 8:00 a.m. when the cows were confined within an area which includes the service alley and the second row of stalls (Figure 19). Forty negative images were selected. Among them 24, i.e., 4 images for each of the 6 cameras placed above the feeding alley, were used for the building of the test images regarding the contours belonging to the sub-class *1a*; whereas 16, i.e., 4 images for each of the 4 cameras placed above the resting area, were used for the building of the test images regarding the contours belonging to the sub-class *1b*. Image overlay and alteration operations involved 16,416 test images for the class *1a* and 6656 for the class *1b*. The results of the test were reported in Table 17.

Table 17 - Results of the test of cow identification for the two sub-classes.

Sub-class	N. of test images	TP	FN	FP	HR	MR	FPR
	n.	n.	n.	n.			
1a	16416	14945	1471	705	0.91	0.09	0.04
1b	6656	5776	880	495	0.87	0.13	0.07

The accuracy assessment procedure was carried out by using the video-recordings acquired between the 8^{th} and the 14^{th} of June, 2012, from 7:00 a.m. to 7:00 p.m. The camera images were selected at 10-minute sampling intervals. For each day, by applying 10-min instantaneous scan sampling, 72 calibrated camera images were obtained for each of the 10 cameras. Therefore, an amount of 5040 images were extracted and analyzed by the automatic identification tool and by an operator. In these images the presence of the marked cow would be correctly detectable by the operator 504 times if it was always in the functional areas monitored by the multi-camera system. However, the operator found the cow in 353 images which constituted about 70% of the real presences. The marked cow was in feeding, standing, and perching in 231 images, whereas in lying in the remaining images, i.e., 122 images.

The software tool found 212 TPs and 19 FNs among the symbol appertaining to the contour sub-class *1a;* whereas 120 TPs and 13 FNs were found among the symbol belonging to the contour sub-class *1b.*

The number of FPs obtained by the software tool was 13 for the sub-class *1a* and 17 for the sub-class *1b*. The accuracy indices computed on the basis of these values were: BF=0.06; MF=0.09; CDP=0.92; QP=0.87 for the sub-class *1a*, and BF=0.14; MF=0.11; CDP=0.90; QP=0.80 for the sub-class *1b*.

5 DISCUSSION

5.1 Further utilization of the designed multi-camera video recording system

Apart from its crucial role for the development and application of the two methodologies proposed in this study, the installation of the multi-camera video recording system represented a benefit for the breeder because, during the trial, he was able to observe the herd by means of a more sophisticated surveillance system than those commonly commercialised. In fact, the multi-camera video recording system provided the breeder with real-time synchronized panoramic top-view video-recordings that avoided the observation of the herd by using different cameras.

Some commercial multi-camera video-recording systems are composed of very expensive hardware required to synchronize the cameras (Liu, Yang, & You, 2012). However, in literature any description of a multi-camera video recording system that produces the synchronized panoramic top-view image sequence of the framed environment in real-time was not found.

Research based on the analysis of digital images from time-lapse video-recordings (cf. 2.2.1.2) could take advantage of the design of a multi-camera system carried out by following the steps described in this research (cf. 3.1). In fact, the synchronized panoramic top-view image sequence of the breeding environment can facilitate the visual recognition of each animal of the herd.

For further research, the calibrated and synchronized camera images or the panoramic top-view images provided by the multi-camera recording system may be used to acquire the input for other automatic software tools designed with the aim of studying other cow's behavioural activities or analysing cow's posture and/or locomotion.

Finally, the multi-camera video-recording system proposed in this study could be used to validate other animal identification systems. In this context, another research that is still in progress, aims at assessing the accuracy of an individual cow location and tracking system based on the Ultra Wide Band (UWB) radio transmission technology. This UWB system was installed in the same free-stall barn where this research was carried out and is being validated by using as reference data the cow top-view images coming from the installed multi-camera video-recording system. Recent studies carried out in conditions different from those characterizing breeding environments, encourage this application because UWB characteristics, i.e., low power transmission, multi-path fading robustness, ultra-fine time resolution and multiple simultaneous transmissions, are suitable to improve object position estimation in outdoor and indoor environments (Mucchi, Trippi, & Carpini, 2010; Xiong, Song, Lai, Zhang, & Yi, 2010; Yuechun & Ganz, 2005).

5.2 First methodology: the automatic detection of cow behavioural activities in free-stall barns (objective 1)

With regards to the application of the first methodology that was aimed at the automatic detection of cow behavioural activities in free-stall barns, some relevant research findings is discussed in the following of this Section.

For the training of the 8 classifiers a total amount of 3782 positive images and 4308 negative images were used. This result demonstrates that in this research a smaller number of both positive and negative images than those utilized by Viola & Jones for human face detection made it possible to obtain low FPR_{cas} values, about equal to 1.3×10^{-5} on average, and high values of TPR_{cas}, about equal to 0.91 on average. It could be explained, bearing in mind that the classifiers modelled in this study were robust with regards to brightness background variation which were observed in the breeding environment during daytime hours (cf. 4).

The lowest number of positive images, which amounted to about 7% of the overall number of the selected positive images, was required for the training of the classifiers *4a* and *4b* regarding the perching behaviour because of the low frequency of this behavioural activity in comparison to the others examined in this research. The highest number of positive images, which amounted to about 46%, was required for the training of the classifiers *3a*, *3b* and *3c* related to the standing behaviour. This fact occurred because of the high variability of the orientation of cow withers-to-pinbone axis within the feeding alley. As the Viola & Jones algorithm is not invariant to the positive image rotation, a great number of positive images were selected in order to provide the algorithm with as many examples of possible cow positions as possible. A modest number of positive images, which amounted to about 15% of the overall number of selected positive images, was required for the training of the feeding behaviour classifier. Contrary to what found for the standing behaviour classifiers, that occurred because of the very low variability of the position of the cows body when they were standing in front of the feed barrier. The same consideration can be done for the training of lying behaviour classifiers *1a* and *1b* which required about 22% and 8% of positive images, respectively.

At the end of the training phase, the classifiers used for the lying and feeding behaviours achieved the best final values of TPR and FPR as they resulted greater than 0.90 and less than 8×10^{-7}, respectively. Similarly, the classifiers used for the perching behaviour achieved good values of TPR as they resulted greater than 0.90 though values of FPR lower than 8×10^{-6} were obtained. These last values of FPR demonstrated a good reliability of the lying, feeding and perching classifiers in discarding the background since the incidence of false positives resulted equal or less than the order of magnitude 10^{-6}. Concerning the standing behaviour classifiers, though the TPR values were approximately similar to those obtained for the other classifiers, the FPR final values were of the order of magnitude 10^{-5}.

For all the classifiers the execution of the training tool was quite slow since it took approximately 24 hours for each classifier by using an Intel® Core (TM) 2 Quad CPU Q6700.

The test of the classifier produced high values of HR which ranged between 0.86 and 0.89. It demonstrated that all the classifiers have a great ability to detect cows even when noises in the processed images occurred. For the lying, feeding, and perching behaviour classifiers the low value of FPR equal to 0.08 confirms what happened in the training, i.e., these three classifiers have a good capacity to discard the background; whereas for the three standing classifiers FPR values were slightly higher and about equal to 0.10.

With regards to the results obtained from the accuracy assessment procedure, Table 18 reports the incidence, in terms of frequency, of false positives and false negatives for each classification as well as the values obtained for the BF and MF indices which show the omission error, i.e., each classifier does not detect a cow behaviour over a number of true positives have been detected, and in term of commission error, i.e. the classifier produces a false positive over detecting a number of true positives.

For the lying behaviour classifiers (classifiers *1a* and *1b)*, the values obtained for the MF indices showed that for every 11 cows correctly detected, one omission error occurred. The classifiers for feeding behaviour (classifier *2*), standing behaviour (classifiers *3a, 3b,* and *3c*) and perching behaviour (classifiers *4a and 4b*) showed an increase of the MF index. In fact, the one omission error occurred one every 7, 6, and 8 true positives correctly detected for the feeding behaviour classifiers, the standing behaviour classifiers and the perching behaviour classifiers, respectively.

Regarding the BF index, the best results were obtained for the lying behaviour classifiers, feeding behaviour classifier, and perching behaviour classifiers that made one commission error every 13, 12, 10 true positive correctly

Table 18 – Quality indices obtained for the 8 classifiers in the accuracy assessment. Incidence of false negatives and false positives in the classifications.

8-14 August 211											
Group of classifiers	**Real cow presences**	**TP**	**FN**	**FP**	**MF**	**MF**	**BF**	**CDP**	**QP**	**False negative frequency (1/MF)**	**False positive frequency (1/BF)**
1a, 1b	2,281	2,088	193	167	0.09	0.09	0.08	0.92	0.85	1 every 11 TPs	1 every 13 TPs
2	2,217	1,922	295	157	0.15	0.15	0.08	0.87	0.81	1 every 7 TPs	1 every 12 TPs
3a, 3b, 3c	1,286	1,102	184	190	0.17	0.17	0.17	0.86	0.75	1 every 6 TPs	1 every 6 TPs
4a, 4b	553	494	59	49	0.12	0.12	0.10	0.89	0.82	1 every 8 TPs	1 every 10 TPs

detected, respectively. The worst results were obtained for the standing behaviour classifiers that made one commission error every 6 true positives correctly detected.

The ability of the classifiers to recognize the cows is proven by the values obtained for the CDP index, which are directly comparable with the HR values obtained in the test phase. For lying and perching behaviours, the cow detection ability of the modelled classifiers improved compared with the test phase. This improvement was not found in the standing behaviour classifiers. This result is due to the smaller number of postures assumed by the animals while lying in the stalls compared to that found when they are in the feeding alley. In fact, the stalls geometrically identify the area where cow presence is possible and thus facilitate the selection of positive images which contain the body of just one animal.

Lastly, by considering that the operations carried out to obtain the plan view of the area under study produced sequences of panoramic images with a frequency of 0.5 fps and that the time needed to classify the content of a panoramic top-view image using by executing all the 8 classifiers on an Intel® Core (TM) 2 Quad CPU Q670 processor was approximately 900 milliseconds, the automatic detection of cow behavioural activities in free-stall barns can be obtained within 2 seconds. Since shape of the body of the animal did not vary significantly within time intervals of the order of a few seconds, during the behavioural activities considered in this study, the cow detection by means of the CVBS can be considered in real-time.

5.3 Second methodology: the identification of dairy cows in free-stall barns (objective 2)

During the building of the contour database only 105 positive images regarding the contours of the symbols used to mark the cows were used. This low number of positive images in comparison to the whole number of camera images extracted for the building of the database, which amounted to 588 images, made it possible to obtain a comprehensive set of contours that was representative of the different shapes assumed by the symbol during the different cow behavioural activities. This result was due to the operations carried out to enhance the quality of the input positive images that increased the number of the contours extracted from the 105 positive images which amounted to 690 contours subdivided into the sub-class *1a* and the sub-class *1b*. A low number of contours stored in the database made it possible the execution of the contour matching in real-time.

The building of the database required about three hours for positive image extraction and about three hours for contour identification.

The high values of HR, which was equal to 0.91 for the sub-class *1a* and equal to 0.87 for the sub-class *1b*, obtained during the test of the software tool for the automatic extraction and matching of contours, demonstrated that the CVBS had a great ability to identify the symbol marked in the coat of the cow also when image noise in the processed images occurred. Furthermore, the very low value of FPR, equal to 0.04 for the sub-class *1a* and equal to 0.07 for the sub-class *1b*, highlighted the suitability of the system to discard the contours from the background. By comparing the results obtained from the two considered sub-classes, it is possible to observe that the contour classifier had a great ability to detect the contours belonging to the sub-class *1a* than those of the sub-class *2a*. Also the incidence of FPs for the sub-class *2a* was twice as much as the incidence of FPs for sub-class *1a*. This worse performance of the contour classifier in recognizing the contours belonging to the sub-class *2a* was mainly due to the deformation of the symbol that marked the coat of the cow when it was lying in the stall as well as to the scarce visibility of the symbol itself.

With regard to the results of the validation phase, they are summarized in Table 19 which reports the incidence of false positive and false negative for each classification and the values obtained for BF and MF indices which show the omission error, i.e., the contour classifier does not detect the symbol marked on the coat of the cow over a number of true positive detected, and in the commission error, i.e. contour classifier produces a false positive over a number of true positive detected.

For the contour classifier of sub-class *1a,* the value obtained for the MF index showed that for every 11 cows correctly identified, one omission error occurred; whereas for the contour classifier of sub-class *1b* there was an increase of the MF index. In fact, one omission error occurred every 9 true positives correctly detected.

Regarding the BF index, the best results were obtained for the identification of the contours of the sub-class *1a*, where the contour classifier made one commission error every 16 true positive correctly detected. The worst results were obtained for the identification of the contours of the sub-class *1b*, where the contour classifier made one commission error every 7 true positive correctly detected.

The ability of the CVBS to identify the marked cow is proven by the values obtained for the two CDP indices, which are directly comparable with the HR values obtained in the test phase. The values obtained for the two indices

Table 19 - Quality indices obtained in the accuracy assessment carried out on camera images selected at 10-minute scan sampling for the identification of the contours belonging to the two sub-classes. Incidence of false negatives and false positives in the identification.

15 June 2012										
Sub-class	**Real marked cow presences**	**TP**	**FN**	**FP**	**MF**	**BF**	**CDP**	**QP**	**False negative frequency (1/MF)**	**False positive frequenc y (1/BF)**
1a	231	212	19	13	0.09	0.06	0.92	0.87	1 every 11 TPs	1 every 16 TPs
1b	122	120	13	17	0.11	0.14	0.90	0.80	1 every 9 TPs	1 every 7 TPs

demonstrated a high level, approximately equal to 90%, of correct identifications of the marked cow. Moreover, by comparing the results obtained during the test phase with those achieved during the validation phase, an improvement of the identification accuracy was recorder for both the contours belonging to sub-class *1a* (from 0.91 to 0.92) and sub-class *1b* (from 0.87 to 0.90). These results were due to the high level of alterations applied to the test images that, therefore, resulted in a worse quality in comparison to the real images used in the validation phase.

The time required for the identification of the marked cow in each camera of the multi-camera system by executing the software tool for the automatic identification on an Intel® Core (TM) 2 Quad CPU Q670 processor, was approximately 100 milliseconds. On the basis of this result and the availability of a sequence of ten synchronized and calibrated camera images having a frequency of 0.5 fps achieved by the CVBS, it was proved that the CVBS proposed in this study could allow the automatic cow identification of herds constituted by a maximum number of 20 marked cows within 2 seconds. Similarly to that observed for the first methodology, the cow identification by means of the CVBS can be considered a real-time application because of the characteristics of cows' locomotion.

Finally, the proposed CVBS made it possible the identification of each animal of the herd by avoiding the use of sensors such as active and passive RFID tags and position markers, (Barbari et al., 2008; Porto et al., 2012) and other wireless technology-based instruments (Huhtala et al., 2007). This represents a great advantage since the integration of other systems for cow identification within the proposed CVBS is not recommended because of the increasing of the overall cost of the system.

5.4 Potential applications of the CVBS and further improvements

5.4.1 First methodology: the automatic detection of cow behavioural activities in free-stall barns (objective 1)

The CVBS proposed in this study allows for the detection of some behavioural activities of dairy cows housed in free-stall barn. It can be used for the computation of a number of indices related to the investigated behaviours and the barn utilization (Carreira et al., 2009; Cook, Bennett, & Nordlund, 2005; Mattachini et al., 2011; Overton et al., 2002), such as the Cow Lying Index (CLI), the Cow Feeding Index (CFI), and the Cow Standing Index (CSI).

Figure 37 shows the comparison between the values of the CLI index computed by means of the visual recognition of cow activities from the panoramic top-view images and those obtained by means of the classification results of the CVBS. The Pearson's correlation coefficient between the two datasets was equal to 0.98 (p=0.000). This proved that the CVBS is suitable for the computation of the CLI index. A slight difference between the two curves was recorded in the time interval before the second milking, i.e., before the 7:30 p.m., because of the restlessness of the herd, mainly due to the need of the cows to be milked.

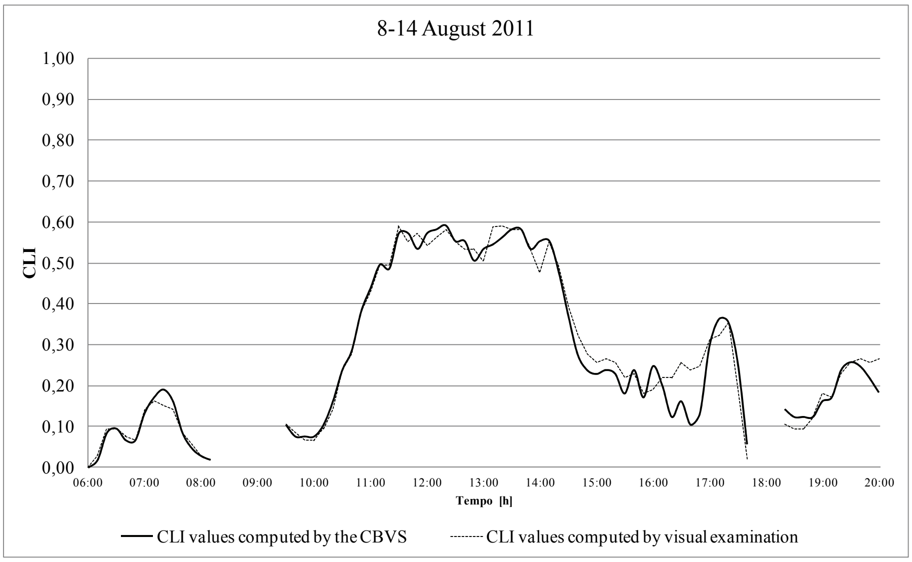


Figure 37 - Comparison between the CLI values obtained by the CVBS and those achieved by the visual recognition of cow activities from the panoramic top-view images of the resting area.

The classification results obtained by the CVBS were used also for the computation of the CFI index. Figure 38 illustrates the comparison between the values of the CFI index computed by means of the visual recognition of cow activities from the panoramic top-view images and those obtained by means of the CVBS. The Pearson's correlation coefficient between the two datasets was equal to 0.96 (p=0.000). Also for CFI index, this result proves the adequacy of the

CVBS for the computation of this index. A slight difference between the two curves was recorded in the early morning, i.e., around 7:00 a.m. when a high number of cows was standing in front of the feed barrier. In this situation a higher number of false negatives occurred in comparison to that obtained in the other periods of the analysed daytime interval. This was due to the fact that sometimes the classifier missed the cows when their bodies were in close contact with each other during the feeding (Porto, Arcidiacono, Guarnera, & Cascone, 2011).

Finally, also a comparison between the CSI index values computed by means of the visual recognition of cow activities from the panoramic top-view images and those obtained by means of the CVBS was carried out (Figure 39). The Pearson's correlation coefficient between the two datasets, though still high and equal to 0.91 (p=0.000), was slightly worse than those obtained for CLI index and CFI index. Even though the classification errors are uniformly distributed during the whole considered daytime interval, the high value of the Pearson's correlation coefficient proves, also in this case, the adequacy of the CVBS for the computation of the CSI index.

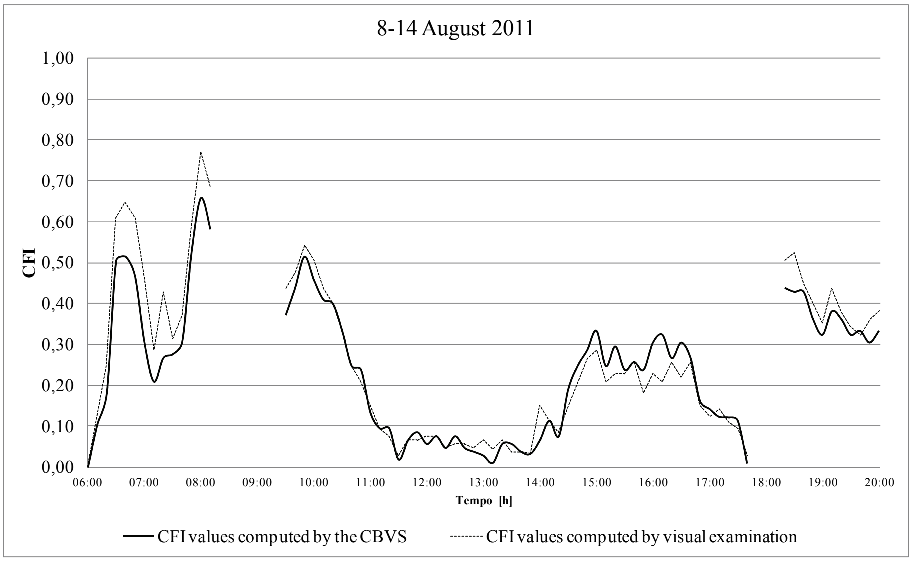


Figure 38 - Comparison between the CFI index values obtained by the CVBS and those achieved by the visual recognition of cow activities from the panoramic top-view images of the resting area.

The results previously described are of relevant importance when the analysis of cow behavioural activity is demanded to the computation of the indices reported above.

In general, by using traditional methods for the analysis of digital images coming from time-lapse video recordings the computation of such indices is highly dependent to the choice of a suitable 'scan sampling interval', its amplitude is strictly related to the type of behaviour to be analysed (Mattachini et al., 2011). The selection of the most adequate scan sampling interval is a crucial issue in order to reduce the time-consuming operations needed for visual image interpretation of video-recordings. As concerns the obtained estimate of the considered indices, the higher the number of the frames extracted from the video sequences, the closer is the estimate to the real value.

In this context, the application of the CVBS proposed in this study made it possible to obtain indices values very close to the real ones because of the capability of the CVBS to work in real time. This last characteristic also avoids the onerous activity of video-recording storage because the results of the detections could be stored in text files.

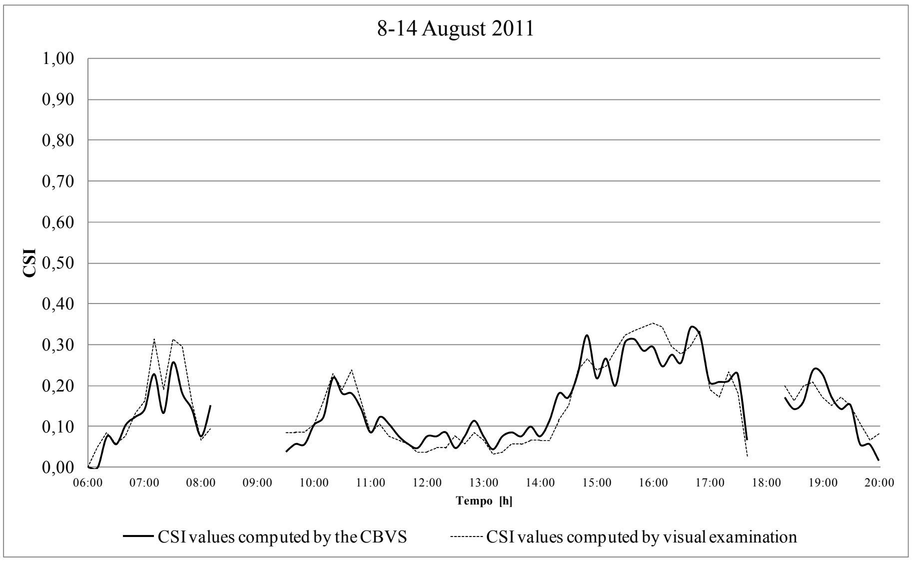


Figure 39 - Comparison between the CSI index values obtained by using the CVBS and those achieved by the visual recognition of the cow activities from the panoramic top-view images of the resting area.

5.4.2 Second methodology: the identification of dairy cows in free-stall barns (objective 2)

The automatic identification of the animals obtained by means of the proposed CVBS can provide also the position of each cow of the herd within the barn. The information on cow position can be used to track each cow marked by a symbol. The possibility of tracking each cow of the herd is an important aspect in order to examine the behavioural patterns and activities, for health inspection, for an estimation of missed operations (for example milking) and for automatic isolation of individual animals at risk (Huhtala et al., 2007).

In the following of this section a potential application of the CVBS for tracking the cow marked during the trial is demonstrated.

The availability of a sequence of synchronized and calibrated camera images having a frequency of 0.5 fps as well as the characteristics of cows' locomotion allowed the definition of a positioning constraint of the marked cow within consecutive camera images. The information on the position of the cow and the defined constraint made it possible to track the marked cow within the study area as well as to improve the quality of the identification process by removing those positions that were incompatible with the defined constraint.

By a visual analysis of the video recordings acquired by the multi-camera system, it was possible to observe that when the marked cow was framed by a camera at the instant *t*, it was still framed by the same camera or by an adjacent one after 2 seconds. This characteristic was considered a position constraint.

To demonstrate the potential application of the CVBS for tracking the cow marked during the trial a probabilistic tracking model was implemented (Burghardt & Càlic, 2006). This model estimates the probability of the presence of the marked cow in each camera of the multi-camera system by using the temporal accumulation of the presences of the marked cow in the whole cameras. A threshold based on the duration of the cow presence in the images obtained from one camera must be defined in order to validate the position of the marked cow.

The model was applied to video-recordings acquired on the 15th of June, 2012, from 7:00 a.m. to 7:00 p.m. The camera images were selected at 2-second sampling intervals. A large number of calibrated camera images, equal to 23253, was obtained for each set constituted by the 10 cameras. Therefore, an amount of 232530 images were extracted and analyzed by the automatic identification tool and by operator. The operator found the cow in 17553 images The marked cow was in feeding, standing, and perching in 11952 images, whereas in lying in the remaining images, i.e., 5601 images.

The software tool for the automatic identification found 10956 TPs and 996 FNs among the symbol appertaining to the contour of sub-class *1a;* whereas found 5030 TPs and 571 FNs among the symbol appertaining to the contour of sub-class *1b.*

The number of FPs given by the software tool was 558 for the sub-class *1a* and 782 for the sub-class *1b*. The accuracy indices computed on the basis of these values were reported in Table 20 and confirmed what found in the case of adopting a 10-minute scan sampling interval (cf. Table 19). It is possible to affirm that since the 2-second scan sampling interval does not provide a better identification of the marked cow, a reduction of the scan sampling interval is not required in the studies concerning behavioural activities that can be analyzed by means of 10-minute scan sampling interval.

Table 21 – Quality indices obtained from the accuracy assessment carried out on camera images selected at 2-second scan sampling by applying the probabilistic tracking model to the outputs of the identification of the contours belonging to the two sub-classes. Incidence of false negatives and false positives in the identification process.

15 June 2012										
Sub-class	**Real marked cow presences**	**TP**	**FN**	**FP**	**MF**	**BF**	**CDP**	**QP**	**False negative frequency (1/MF)**	**False positive frequency (1/BF)**
1a	11952	10910	1042	228	0.10	0.02	0.91	0.90	1 every 10 TPs	1 every 48 TPs
1b	5601	4984	617	276	0.12	0.06	0.89	0.85	1 every 8 TPs	1 every 18 TPs

The probabilistic tracking model was applied to the 23253 outputs of the identification process, i.e., position of the contour, obtained for each of the ten camera images.

The threshold used to validate the presence of the marked cow within a specific camera was set equal to 2, i.e., the cow must be present in the same camera image for at least 4 seconds. The value of the threshold was obtained empirically taking into account the reduction of FPs and only a slight decrease of TPs.

The values of quality indices obtained in the accuracy assessment carried out on camera images selected at 2-second scan sampling by applying the probabilistic tracking model to the outputs of the identification of the contours belonging to the two sub-classes were reported in Table 21. In this case, a high decrease of the false positives of about 59% for the sub-class *1a* and of about 65% for the sub-class *1b* was recorded. However, a slight reduction of the TPs was observed for both the two contours sub-classes (about 0.5% for the sub-class *1a* and 1% for the sub-class *1b*).

This result demonstrated that the CVBS was very suitable for tracking the cow marked during the trial.

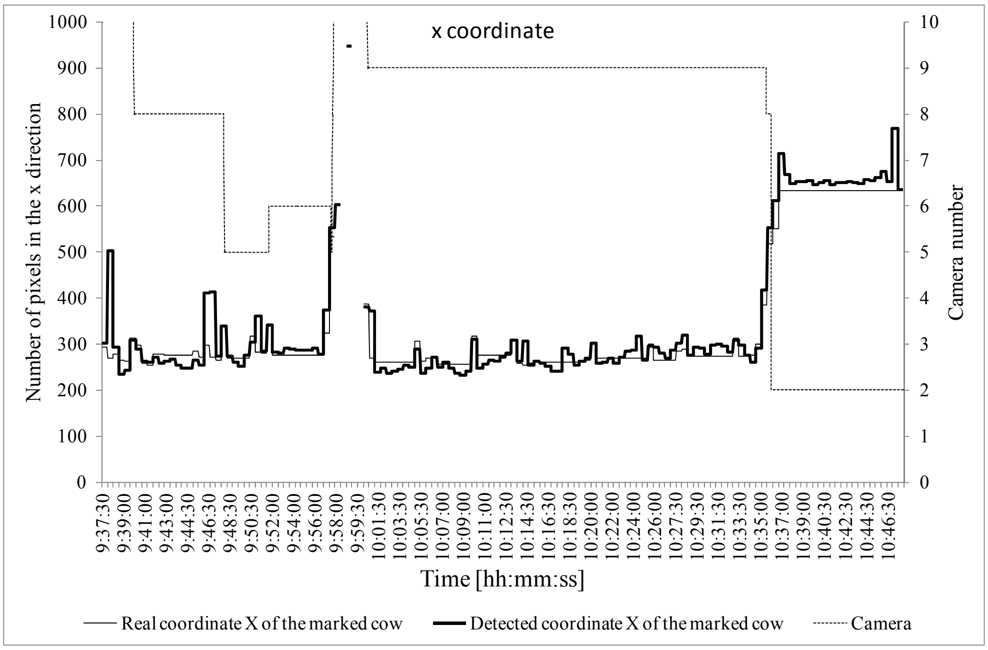
The path observed by the operator (dataset A) described by means of x and y coordinates and that detected by the identification system (dataset B) were reported in Figure 40.

Table 20 – Quality indices obtained from the accuracy assessment carried out on camera images selected at 2-second scan sampling for the identification of the contours belonging to the two sub-classes. Incidence of false negatives and false positives in the identification process.

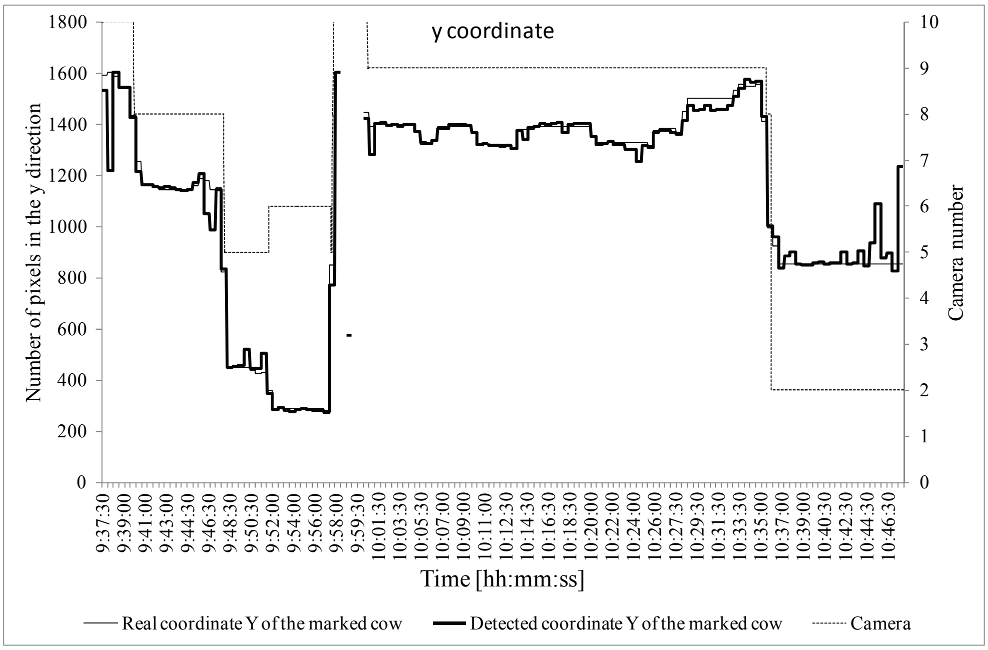
15 June 2012										
Sub-class	**Real marked cow presences**	**TP**	**FN**	**FP**	**MF**	**BF**	**CDP**	**QP**	**False negative frequency (1/MF)**	**False positive frequency (1/BF)**
1a	11952	10956	996	558	0.09	0.05	0.92	0.87	1 every 11 TPs	1 every 20 TPs
1b	5601	5030	571	782	0.11	0.16	0.90	0.79	1 every 9 TPs	1 every 6 TPs

Whereas, the Figure 41 shows the comparison between the dataset A and that detected by the identification system which was improved by the probabilistic tracking model (dataset C). For both the figures the observation time interval was from 09:37 a.m. to 10:47 p.m. of the 14th of June 2012. Moreover, to improve the readability of the figures the values of the coordinates x and y were averaged every 30 seconds.

The dataset B had a high correlation (r=0.975 for the x coordinate and r=0,988 for the y coordinate) with the dataset A. However, though the fitting of the two dataset is high, the presence of false positives can be observed at 09:38 a.m., 09:47 a.m., 09:58 a.m., and 10:47 a.m. On the oher hand, the dataset C had shown a slight higher correlation coefficient (r=0.994 for the x coordinate and r=0.999 for the y coordinate) and a substantial reduction of the presence of false positives than those recorded by the dataset B. The improvement of the quality of the tracking is also demonstrated by comparing the maximum error and the mean error that were of 37 pixels and 452 pixels, respectively, for the dataset B and 23 and 79 pixels, respectively, for the dataset C.

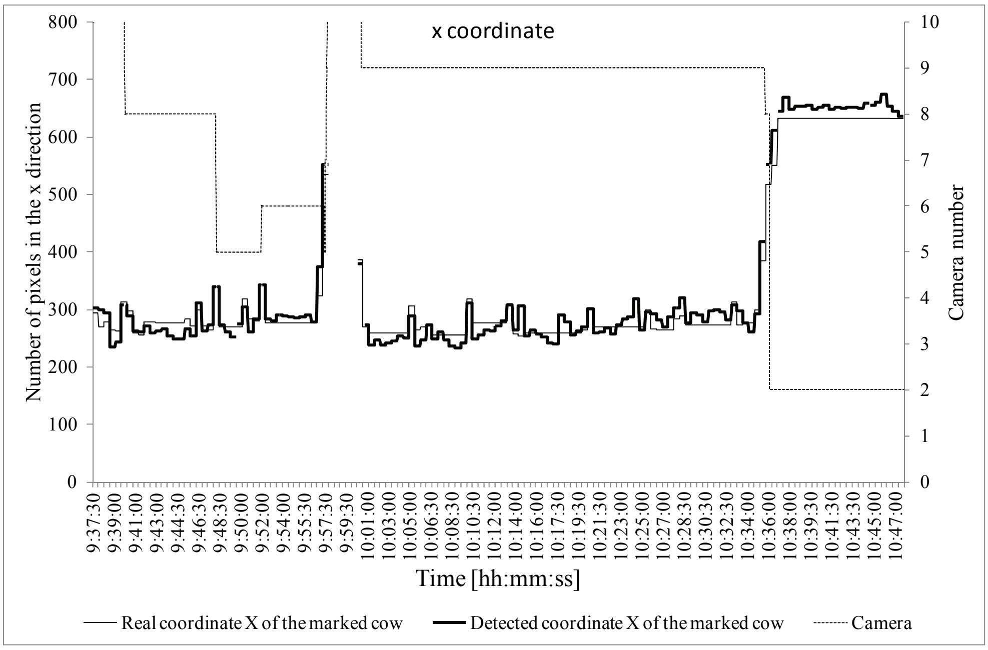


(a)

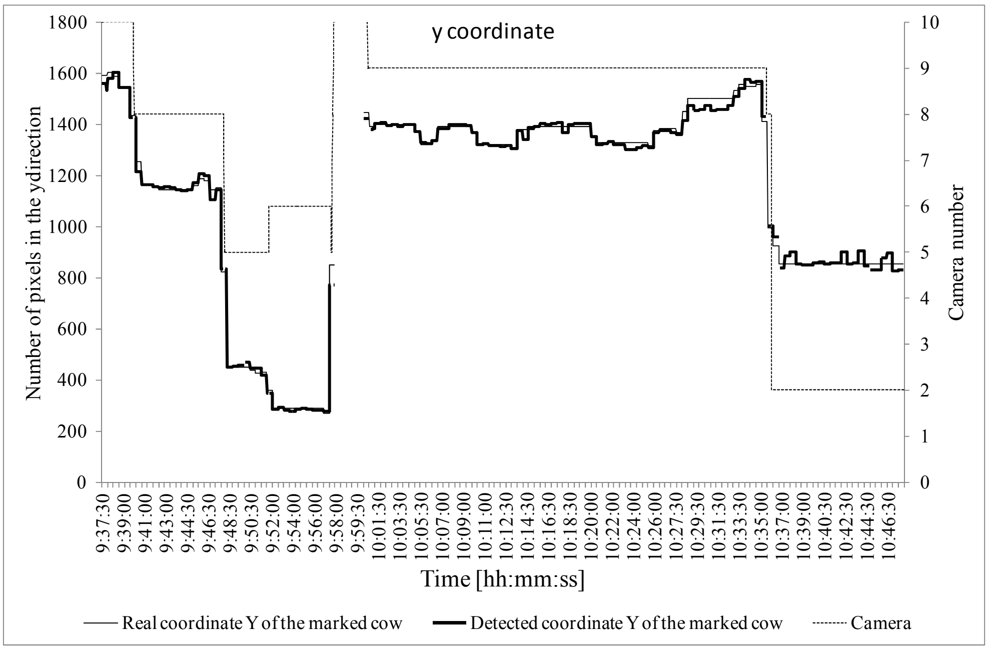


(b)

Figure 40 – a) x coordinates of the positions of the marked cow observed by the operator (dataset A) and detected by the identification system (dataset B); b) y coordinates of the positions of the marked cow observed by the operator (dataset A) and detected by the identification system (dataset B).



(a)



(b)

Figure 41 – a) x coordinates of the positions of the marked cow observed by the operator (dataset A) and detected by the identification system (dataset C) improved by the probabilistic tracking model; b) y coordinates of the positions of the marked cow observed by the operator (dataset A) and detected by the identification system improved by the probabilistic tracking model (dataset C).

6 CONCLUSIONS

The development and application of the two methodologies for the automatic detection of a number of cow behavioural activities and for the automatic identification of dairy cows in free stall barns required the design of a multi-camera video recording system and its installation in a free-stall barn. For the farmer the multi-camera video recording system represented a benefit because he was able to observe the herd by means of a sophisticated surveillance system which provided him with real-time synchronized panoramic top-view video-recordings that avoided a direct observation of the herd or the observation of the herd by using different cameras.

The method proposed in order to design the multi-camera video recording system can also be utilized in other research based on the analysis of digital images to facilitate the visual recognition of each animal of the herd within breeding environment of large dimensions.

The proposed CVBS made it possible to detect of feeding, lying, standing, and perching dairy cow behavioural activities in free-stall barns and cow identification. The identification of the cow was obtained without increasing the overall cost of the system by avoiding the use of sensors such as active and passive RFID tags and other wireless technology-based instruments.

The classifiers modelled in this study to develop the two proposed methodologies were robust with regards to brightness background variations which were observed in the breeding environment during daytime hours. This represents an advance in the state of the art because in literature the use of automatic systems for behaviour detection was tested only in laboratory trials. In fact, it is widely recognized that in commercial livestock houses the application of image analysis for behaviour classification becomes more complicated due to the higher noises deriving from the characteristics of the breeding environment.

The CVBS made it possible to obtain high quality indices Cow detection percentage (CDP) and Quality percentage (QP) computed in the real context related to the case study provided the following results: the detection of the four dairy cow behavioural activities in the free-stall barn provided a CDP > 86% and a QP > 75%; marked cow identification had a CDP > 90% and a QP > 79% in the case when the probabilistic tracking model was not utilized, and it obtained a CDP > 90% with a QP > 85% when using also the probabilistic tracking model which required, however, the availability of a 2-second set of camera images.

The tasks executed by the CVBS to both produce the sequences of panoramic top-view images of the area under study and perform the detection of cow behavioural activities and the cow identification, took 2 seconds on an Intel® Core (TM) 2 Quad CPU Q670. With reference to the behavioural activities considered in this study, since the shape of the animal body did not vary significantly within 2-seconds time interval, the cow detection by means of the CVBS can be considered in real-time.

From a scientific point of view, researchers may be interested in making use of the methodologies proposed in this thesis work in order to perform automatic analyses of animal behaviour, as well as to validate other systems used for studying animal behaviours in different environmental conditions.

REFERENCES

Albertini, M., Canali, E., Cannas, S., Ferrante, V., Mattiello, S., Panzera, M., & Verga, M. (2008). *Etologia applicata e benessere animale* (Vol. 1).

Altmann, J. (1974). Observational study of behaviour: sampling methods. *Behavior, 4*, 1-41.

Banhazi, T. M., Lehr, H., Black, J. L., Crabtree, H., Schofield, P., Tscharke, M., & Berckmans, D. (2012). Precision Livestock Farming: An international review of scientific and commercial aspects. *International Journal of Agricultural and Biological Engineering, 5*(3).

Barbari, M., Conti, L., Koostra, B. K., Masi, G., Guerri, F. S., & Workman, S. R. (2006). The Use of Global Positioning and Geographical Information Systems in the Management of Extensive Cattle Grazing. *Biosystems Engineering, 95*(2), 271-280.

Barbari, M., Conti, L., & Simonini, S. (2008). *Spatial identification of animals in different breeding systems to monitor behavior.* Paper presented at the Proceedings of the VIII Congress on Livestock Environment, Iguassu Falls, Brazil, 31 August - 4 September.

Bicalho, R. C., Vokey, F., & Guard, C. L. (2007). Visual Locomotion Scoring in the First Seventy Days in Milk: Impact on Pregnancy and Survival. *Journal of Dairy Science*(90), 4586-4591.

Blumstein, D. T., & Janice, C. D. a. (2007). *Quantifying Behavior the JWatcher Way*.

Bradsky, G., & Kaehler, A. (2008). *Learning OpenCV.* 1005 Gravenstein Highway North, Sebastopol, CA 95472.: O'Reilly Media, Inc., .

Burghardt, T. (2004). *Automated Visual Recognition of Individual African Penguins.* Paper presented at the Fifth International Penguin Conference, Ushuaia, Tierra del Fuega, Argentina, September, 2004.

Burghardt, T., & Càlìc, J. (2006). *Real-time face detection and tracking of animals.* Paper presented at the Proccedings of the 8th Seminar on Neural Network Application in Electrical Engineering.

Burley, N. (1988). Wild zebra finches have band-colour preferences. *Animal Behaviour, 36*, 1235–1237.

Cangar, O., Leroy, T., Guarino, M., Vranken, E., Fallon, R., Lenehan, J., . . . Berckmans, D. (2008). Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis. *Computers and Electronics in Agriculture, 64*(1), 53-60.

Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on pattern analysis and machine intelligence, VOL. PAMI-8, N. 6*, 279-289.

Carreira, X. C., Fernández, M. E., & Mariño, R. A. (2009). Indices for estimation of dairy free-stall occupancy. *Applied Animal Behaviour Science, 119*(1-2), 23-31.

Cook, N. B., Bennett, T. B., & Nordlund, K. V. (2005). Monitoring indices of cow comfort in free-stall-housed dairy herds. *Journal of Dairy Science, 88*(11), 3876-3885.

Cootes, T. F., Taylor, C., J.,, Cooper, D. H., & Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding, 61*(1), 38-59.

Dawkins, M. (2007). *Observing Animal Behaviour. Design and analysis of quantitative data*.

DeVries, T. J., & Von Keyserlingk, M. A. G. (2006). Feed stalls affect the social and feeding behavior of lactating dairy cows. *Journal of Dairy Science, 89*(9), 3522-3531.

DeVries, T. J., Von Keyserlingk, M. A. G., & Weary, D. M. (2004). Effect of feeding space on the inter-cow distance, aggression, and feeding behavior of free-stall housed lactating dairy cows. *Journal of Dairy Science, 87*(5), 1432-1438.

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

DeVries, T. J., Von Keyserlingk, M. A. G., Weary, D. M., & Beauchemin, K. A. (2003b). Technical note: Validation of a system for monitoring feeding behavior of dairy cows. *Journal of Dairy Science, 86*(11), 3571-3574.

Endres, M. I., DeVries, T. J., Von Keyserlingk, M. A. G., & Weary, D. M. (2005). Short communication: Effect of feed barrier design on the behavior of loose-housed lactating dairy cows. *Journal of Dairy Science, 88*(7), 2377-2380.

Firk, R., Stamer, E., Junge, W., & Krieter, J. (2002). Automation of oestrus detection in dairy cows: A review. *Livestock Production Science, 75*(3), 219-232.

Fregonesi, J. A., & Leaver, J. D. (2001). Behaviour, performance and health indicators of welfare for dairy cows housed in strawyard or cubicle systems. *Livestock Production Science, 68*(2-3), 205-216.

Fregonesi, J. A., Tucker, C. B., Weary, D. M., Flower, F. C., & Vittie, T. (2004). Effect of rubber flooring in front of the feed bunk on the time budgets of dairy cattle. *Journal of Dairy Science, 87*(5), 1203-1207.

Fregonesi, J. A., Veira, D. M., Von Keyserlingk, M. A. G., & Weary, D. M. (2007). Effects of bedding quality on lying behavior of dairy cows. *Journal of Dairy Science, 90*(12), 5468-5472.

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences, 55*(1), 119-139.

Frost, A. R., Schofield, C. P., Beaulah, S. A., Mottram, T. T., Lines, J. A., & Wathes, C. M. (1997). A review of livestock monitoring and the need for integrated systems. *Computers and Electronics in Agriculture, 17*(2), 139-159.

Galindo, F., Broom, D. M., & Jackson, P. G. G. (2000). A note on possible link between behaviour and the occurrence of lameness in dairy cows. *Applied Animal Behaviour Science, 67*(4), 335-341.

Gledhill, D., Tian, G. Y., Taylor, D., & Clarke, D. (2003). Panoramic imaging - A review. *Computers and Graphics (Pergamon), 27*(3), 435-445.

Gonzales, R. C., & Woods, R., E. (2002). *Digital image processing second edition*. Upper Saddle River, New Jersey 07458: Prentice Hall.

Gygaxa, L., Neisen, G., & Bollhalder, H. (2007). Accuracy and validation of a radar-based automatic local position measurement system for tracking dairy cows in free-stall barns. *Computers and Electronics in Agriculture, 56*, 23-33.

Halachmi, I., Edan, Y., Maltz, E., Peiper, U. M., Moallem, U., & Brukental, I. (1998). A real-time control system for individual dairy cow food intake. *Computers and Electronics in Agriculture, 20*(2), 131-144.

Huhtala, A., Suhonen, K., Mäkelä, P., Hakojärvi, M., & Ahokas, J. (2007). Evaluation of Instrumentation for Cow Positioning and Tracking Indoors. *Biosystems Engineering, 96*(3), 399-405.

Kaihilahti, J., Suokannas, A., & Raussi, S. (2007). Observation of Cow Behaviour in an Automatic Milking System using Web-based Video Recording Technology. *Biosystems Engineering, 96*(1), 91-97.

Lefkovits, S. (2009). Assessment of Building Classifiers for Face Detection. *Acta Universitatis Sapientiae Electrical and Mechanical Engineering,*(1), 175-186.

Li, S., & Lee, M. C. (2005). Effective Invariant Features for Shape-Based Image Retrieval. *Journal of the American Society for Information Science and Technology,, 56*(7), 729-740.

Lienhart, R., & Maydt, J. (2002). *An extended set of Haar-like features for rapid object detection.* Paper presented at the Proceeding IEEE International Conference on Image Processing (ICIP '02).

Liu, Y., Yang, M., & You, Z. (2012). Video synchronization based on events alignment. *Pattern Recognition Letters 33*, 1338–1348.

Maertens, W., Vangeyte, J., Baert, J., Jantuan, A., Mertens, K. C., De Campeneere, S., . . . Van Nuffel, A. (2011). Development of a real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: The GAITWISE system. *Biosystems Engineering, 110*(1), 29-39.

Malik, J., Belongie, S., Leung, T., & Shi, J. (2001). Contour and Texture Analysis for Image Segmentation. *International Journal of Computer Vision, 43*(1), 7-27.

Martin, P., & Bateson, P. (2007). Measuring behaviour: An Introductory Guide. *Cambridge University Press*.

Mattachini, G., Riva, E., & Provolo, G. (2011). The lying and standing activity indices of dairy cows in free-stall housing. *Applied Animal Behaviour Science, 129*(1), 18-27.

Mills, A., & Dudek, G. (2009). Image stitching with dynamic elements. *Image and Vision Computing, 27*(10), 1593-1602.

Mitlohner, F. M., Morrow-Tesch, J. L., Wilson, S. C., Dailey, J. W., & Mcglone, J. J. (2001). Behavioural sampling techniques for feedlot cattle. *Journal of Animal Science, 79*(5), 1189-1193.

Mucchi, L., Trippi, F., & Carpini, A. (2010). *Ultra wide band real-time location system for cinematic survey in sports.* Paper presented at the In 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL). Roma.

Müller, R., & Schrader, L. (2003). A new method to measure behavioural activity levels in dairy cows. *Applied Animal Behaviour Science, 83*(4), 247-258.

Munksgaard, L., Jensen, M. B., Pedersen, L. J., Hansen, S. W., & Matthews, L. (2005). Quantifying behavioural priorities - Effects of time constraints on behaviour of dairy cows, Bos taurus. *Applied Animal Behaviour Science, 92*(1-2), 3-14.

Nishida, T., Hosoda, K., Matsuyama, H., & Ishida, M. (2004). Effect of lying behavior on uterine blood flow in cows during the third trimester of gestation. *Journal of Dairy Science, 87*(8), 2388-2392.

O'Driscoll, K., Boyle, L., & Hanlon, A. (2009). The effect of breed and housing system on dairy cow feeding and lying behaviour. *Applied Animal Behaviour Science, 116*(2-4), 156-162.

Overton, M. W., Sischo, W. M., Temple, G. D., & Moore, D. A. (2002). Using time-lapse video photography to assess dairy cattle lying behavior in a free-stall barn. *Journal of Dairy Science, 85*(9), 2407-2413.

Papageorgiou, C. P., Oren, M., & Poggio, T. (1998). *A general framework for object detection* Paper presented at the Sixth International Conference onComputer Vision.

Pastell, M., Tiusanen, J., Hakojärvi, M., & Hänninen, L. (2009). A wireless accelerometer system with wavelet analysis for assessing lameness in cattle. *Biosystems Engineering, 104*(4), 545-551.

Pluk, A., Bahr, C., Poursaberi, A., Maertens, W., van Nuffel, A., & Berckmans, D. (2012). Automatic measurement of touch and release angles of the fetlock joint for lameness detection in dairy cattle using vision techniques. *Journal of Dairy Science, 95*(4), 1738-1748.

Porto, S. M. C., Arcidiacono, C., Cascone, G., Anguzza, U., Barbari, M., & Simonini, S. (2012). Validation of an active RFID-based system to detect pigs housed in pens. *Journal of Food, Agriculture and Environment, 10*(2), 468-472.

Porto, S. M. C., Arcidiacono, C., Guarnera, G. C., & Cascone, G. (2011). Preliminary study for the implementation of an image analysis algorithm to detect dairy cow presence at the feed barrier. *Journal of Agricultural Engineering*(4), 17-23.

Poursaberi, A., Bahr, C., Pluk, A., Van Nuffel, A., & Berckmans, D. (2010). Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques. *Computers and Electronics in Agriculture, 74*(1), 110-119.

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

Quinlan, J. (1986). Induction of decision trees. *Machine Learning*(1), 81-106.

Rousing, T., & Wemelsfelder, F. (2006). Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Applied Animal Behaviour Science, 101*(1-2), 40-53.

Rulquin, J., & Caudal, J. P. (1992). Effects of laying or standing on mammary bllod flow and heart rate of dairy cows. *Annales de zootechnie, 41*(1), 101.

Rushen, J. (1991). Problems associated with the interpretation of physiological data in the assessment of animal welfare. *Applied Animal Behaviour Science, 28*(4), 381-386.

Schmied, C., Waiblinger, S., Scharl, T., Leisch, F., & Boivin, X. (2008). Stroking of different body regions by a human: Effects on behaviour and heart rate of dairy cows. *Applied Animal Behaviour Science, 109*(1).

Siegel, S., & Castellan Jr., N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences*.

Simonini, S. (2009). *Farm Animal Identification and Spotting Case Study: RFID Active Tags with Marked Area.* University of Firenze.

Sivic, J., & Zisserman, A. (2009). Efficient Visual Search of Videos Cast as Text Retrieval. *IEEE Transactions on pattern analysis and machine intelligence, 31*(4), 591-606.

Smidt, D. (1983). Advantages and problems of using integrated systems of indicators as compared to single traits. *Current topics in veterinary medicine and animal science, 23*, 201-207.

Song, X., Leroy, T., Vranken, E., Maertens, W., Sonck, B., & Berckmans, D. (2008). Automatic detection of lameness in dairy cattle-Vision-based trackway analysis in cow's locomotion. *Computers and Electronics in Agriculture, 64*(1), 39-44.

Sumpter, D. J. T. (2006). The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B, 361*, 5–22.

Suzuki, S. (1985). Topological Structural Analysis of Digitized Binary Images by Border Following. *Computer vision, graphics, and image processing 30*, 32-46.

Viola, P., & Jones, M. (2001). *Rapid object detection using a boosted cascade of simple features.* Paper presented at the In Computer Vision and Pattern Recognition, IEEE Computer Society.

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

Wertheimer, M. (1938). Laws of organization in perceptual forms (partial

translation). *A Sourcebook of Gestalt Psychology,W. Ellis* (pp. 71–88): Harcourt Brace and Company.

Wilson, S. C. (2005). Spectral Analysis of Feeding and Lying Behavior of Cattle Kept Under Different Feedlot Conditions. *Journal of applied animal welfare science, 8*(1), 13-24.

Xiong, H., Song, H., Lai, Z., Zhang, J., & Yi, K. (2010). *A novel indoor localization scheme.* Paper presented at the 12th IEEE International Conference, Nanjing, 338-341.

Yadav, R., B., Nishchal, N., K., Gupta, A., K., & Rastogi, V., K. (2007). Retrieval and classification of shape-based objects using Fourier, generic Fourier, and wavelet-Fourier descriptors technique: A comparative study. *Optics and Lasers in Engineering, 45*, 695–708.

Yang, M., Kidiyo, K., & Rosin, J. (2008). *A survey of shape feature extraction techniques* (Vol. Pattern Recognition).

Yuechun, C., & Ganz, A. (2005). *A UWB-based 3D location system for indoor environments.* Paper presented at the 2nd International Conference on Broadband Networks, BROADNETS 2005, Boston, MA, 224-232.

Zahn, C., & Roskies, R. (1972). Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers, c-21, n.3*, 269-281.

Zhang, D., & Lu, G. (2004). Review of shape representation andd escription techniques. *Pattern Recognition, 37* 1 – 19.

Zhang, Z. (2000). A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(11), 1330-1334.