Video tracking of dairy cows for assessing mobility scores

Carlos Martinez-Ortiz[1], Richard Everson[1], Toby Mottram[2]
[1]*Department of Computer Science, University of Exeter, Exeter, UK,*
[2]*Royal Agricultural University, Cirencester, UK and eCow, Exeter, UK*
C.A.Martinez-Ortiz@exeter.ac.uk

Abstract

Lameness afflicts a large proportion of dairy herds, but could be considerably reduced by automated monitoring by CCTV. Key to this is reliable, robust detection and tracking of individual cows in crowded video sequences. We introduce a novel detection and tracking method, based on the Viola-Jones detector. We show that animals can be tracked and their overall gait patterns and speed automatically extracted from video sequences. Preliminary work on identification of individual animals through principal component analysis and SIFT feature matching is also described.

Keywords: cow lameness, video, detection, tracking, PCA, SIFT.

Introduction

Lameness in dairy cows is an issue widespread of concern for the dairy industry. At any one time, it is widely estimated that up to a third of dairy cattle in the UK suffers some degree of lameness. Identification and treatment of lameness at an early stage can help prevent lameness from becoming more severe, with concomitant benefits to the animal and cost savings to the farmer.

Current practice for measuring mobility scores and identifying cows at risk of lameness relies on visual inspection of the individuals; an expert observes the cows as they walk and assigns them a grade depending on their mobility. Although this method is currently the norm, there are some drawbacks of concern:

- Lack of robustness – being subject to human perception, it is possible for two experts to give different scores to the same individual.
- Expense of expertise – being dependent on the availability of an expert, there are constraints in the frequency at which each cow can be monitored.

These constraints, coupled with large herds cared for by only limited staff, mean that daily monitoring is infeasible. In this paper, we present an automatic video processing system which can provide information on the mobility of dairy cows without requiring human intervention. The principal obstacle to automatic monitoring of dairy cows is the accurate identification and tracking of individual cows and we therefore focus on this aspect, showing how cows can be accurately located and tracked in video. This provides ready measurement of the speed of each cow, which has been shown to be well-correlated with the cow's mobility score (Bell et al., 2013). It also gives access to measures of the animal's gait which may also be used for mobility assessment. We describe preliminary work on the identification of particular cows from video, with the goal of obviating the need for additional systems such as RFID tags for linking scores to particular cows. We draw attention to another video-based analysis system which, unlike ours, uses back posture to assess lameness (Poursaberi et al, 2011)

Methods

In this section we describe the main elements of our proposed system. To enable widespread use, we aim to use commodity hardware rather than specialised equipment.

Setup

The hardware component of the system consists of video recording equipment used to monitor the exit of the milking parlour. Cows leaving the milking parlour in batches of 24 after milking, walk down an exit race approximately 10 m long before turning into a large barn. This is monitored by a standard home-security surveillance camera system mounted overhead, providing a view, principally of the cow's back (see for example Figures 1 and 2). This view minimises the possibility of occlusion. Three additional cameras, providing additional viewpoints were also installed but were not used in the work reported here. The video recording equipment was scheduled to record for two hours in the morning and two hours in the afternoon. These are the times when milking typically takes place, however there is no guarantee of any exact timing when the cows start walking out. For this reason, the recording schedules span a generous time window ensuring that the moment when the cows walk out will be captured. This also means that there are long periods where there is no activity of interest in the videos. The captured video files were stored on the recording equipment hard drive. Afterwards video files were recovered and processed off line.

Detection and tracking cows

The first step towards developing the cow tracking system is to detect when a cow is visible and when it leaves the scene. It is also important to detect where in the current frame the cow is located. When cows are absent, the recorded video comprises the farmyard concrete floor and neighbouring buildings and it might therefore be expected that a straightforward way of detecting and tracking individual cows would be by simple background subtraction, which is often effective for interior and man-made scenes (e.g., Sonka et al., 2007). However, we find that background subtraction methods are ineffective here due to the changing lighting conditions (particularly as milking is often around dawn) and the varying reflectance of the farmyard floor as animal waste is deposited on it and the floor is washed. A further difficulty arises from the close proximity of the cows as they leave the milking parlour: background subtraction and optical flow methods tend to detect moving objects in the video, but fail to separate individual cows.

Rather than detecting entire cows, we therefore choose to locate and track the heads of cows by constructing a specialised detector for cow heads. By detecting cow heads in individual video frames we avoid false positive detection of other moving objects such as people and shadows. In addition, the heads of cows are generally well separated so that neighbouring cows are easily distinguished.

The Viola-Jones object detection algorithm (Viola and Jones, 2004), commonly used for detecting human faces, was adapted to locate the heads of cows in individual video frames. The Viola-Jones detector uses a set of simple image features and combines them to determine whether a face (or head) has been detected. The features used are very similar to the well-known Haar wavelet basis functions and are very simple in their nature; one feature could detect, for example, a horizontal edge of a shape. These features can be very easily and quickly computed. The Viola-Jones detector computes a large number of these features, each one of which on its own is a *weak* classifier, able to detect the presence of a cow head little better than random. The many weak classifiers are combined during training to form a strong classifier using boosting to select the most useful (Freund and Schapire, 1997). In order to achieve very high true positive detections and a low false positive rate, classifiers are arranged in a cascade. At the top level of the cascade a sub-window of the image is checked to discover whether (on the basis of a few features) it may be rejected as containing a cow head; if not the sub-window is processed by further stages of the cascade. Early rejection of a sub-window means that the detector is computationally very efficient and the whole image may be scanned, one sub-window at a time, for the sought object.

In order to use the Viola-Jones detection algorithm to suit our application, it was necessary to construct a cascade which was trained for detecting cow heads. A training set consisting of 1000 heads was manually selected from our video recordings, from different cows and under different lighting conditions. Each training image was 60 pixels square as shown in Figure 1(a). These samples where used to train the weak classifiers forming the head detector cascade. Note that during training each of the training heads is used multiple times after application of various randomly chosen affine transformations (rotation, scaling, shearing), which confers robustness to changes in pose and precise detail of the head. The training heads are used in conjunction with a range of backgrounds, not just from the farmyard which means that cow heads are effectively detected in a wide range of scenes.

Having trained the head detector, it was applied to every frame of each video being inspected. Figure 1(b) shows an example of a detection. For clarity, this image has been cropped to the region surrounding the cow, but detection takes place across the whole video frame without any additional preprocessing and several cow (heads) may be detected in a single frame, see for example Figure 2.

<u>Track extraction</u>

The head detector described above provides a very high detection rate; we estimate the true positive rate to be in excess of 95% with a false positive rate of less than 1%. However, it is still possible for the detector to occasionally miss a head or detect a head where there is none. Therefore, simply joining detections from one frame to the next would yield erroneous tracks. Detections on a series of frames were joined together and smoothed using the Kalman filter (Kalman, 1960; Roweis & Ghahramani, 1999).

We regard the true location of the cow's head as a hidden state, which is related to the observed location of the centres of the detection squares. The Kalman filter can be thought of as a two-stage process in which the location of the hidden location in the next frame is first predicted and then, on observing the next frame, corrected using the new observation. We model the probability of making a transition from one location to another as a simple Gaussian diffusive process and the observed head location is modelled as the true location plus Gaussian distributed observational noise. Given the location of a head in a frame at time t, the predictive step of the Kalman filter is used to estimate the region where the head is likely to be located at time $t+1$. If a head is located within the predicted region, then the true location is updated with the detection

(a) (b)

Figure 1: (a) Sample cow heads used for training and (b) detected cow head

at time *t+1* and the new location added to the track. If no head is located at *t+1*, a predicted location is calculated at *t+2*, with an increased uncertainty, and so on. Notice that for every missed frame, the uncertainty increases until it reaches a maximum uncertainty in which case the track has been lost. In this way detected locations are joined together to form smooth tracks and the location of the cow’s head is interpolated in frames where no detection was made. The smoothness of the track and the prediction window in which detections are sought depend on the values of the state noise uncertainty and the observational noise; however, the resulting tracks are insensitive to their precise values. The Kalman filter updates are all accomplished with linear algebra and so are computationally fast. Figure 2 shows a cow head detected as the cow leaves the milking parlour at the lower left of the image, together with a cow that has been tracked through the exit race. The green squares mark the location of the detected head and the radius of the circle is proportional to the uncertainty in the Kalman filter’s estimated true location of the head. The uncertainty in the right-hand cow’s location is due to it having just passed under a wire which inhibited head detection for a few frames. Figure 3 shows the tracks taken by several cows.

Once this process is completed, the individual head detections have been merged into tracks which describe the movement of the cow's head over time. From these tracks it is possible to calculate the time it takes a cow to cross the corridor where they have been recorded.

Results and discussion

<u>Analysing individual tracks</u>

After the heads detected over a number of video frames have been merged into a single track, it is possible to analyse different aspects of the track. For example it is possible to analyse the path a cow has followed. In this way, the gait asymmetry can be measured. Previous studies (Chapinal et al., 2011) indicate that gait asymmetry is an indicator of mobility scores. Figure 2 illustrates the extracted path which could be used for assessing gait asymmetry.

<u>Timing tracks</u>

Other information that can be obtained from the tracks is the speed of the cow. Position is known at every frame and the total time elapsed is also known, therefore calculating

Figure 2: Detected and tracked cows. Green squares show the location of detected heads; green circles show the Kalman filter uncertainty in the true location and red lines indicate the true path followed and can be used to measure gait asymmetry.

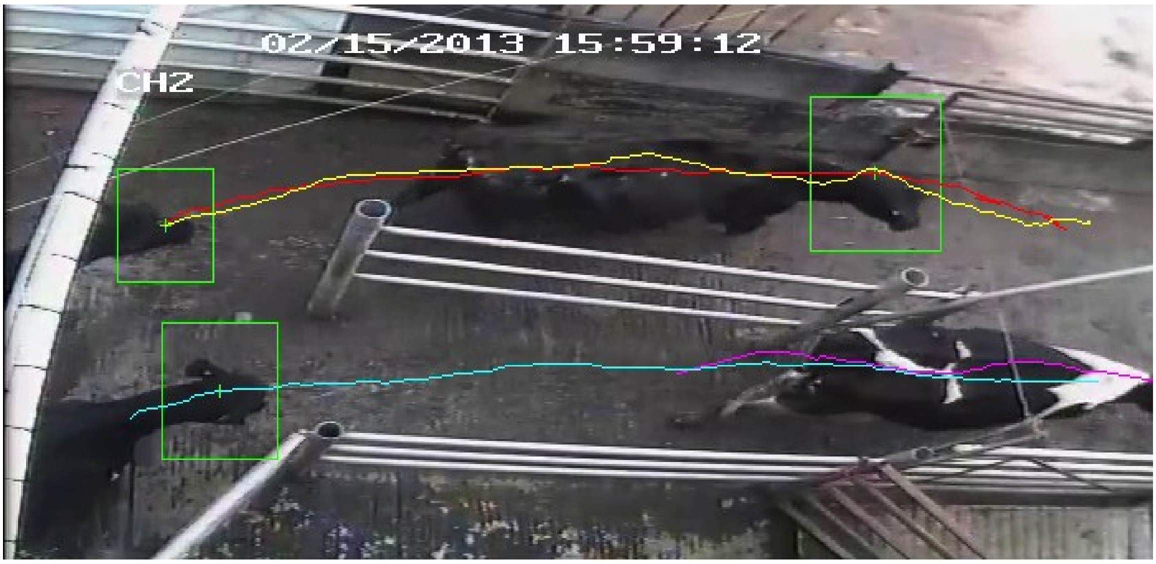

Figure 3: Detected and tracked cows showing the simultaneous detection and tracking of several cows.

speeds is straight forward. Notice that speeds may vary through time (i.e. the cows may slow down or move faster); for this study we use the average speeds over whole tracks.

Bell et al (2012) have established that deterioration of walking speed is one of the characteristic symptoms of lameness. The video processing system presented here can exploit this fact and help in the early identification of lameness. Figure 4 shows a

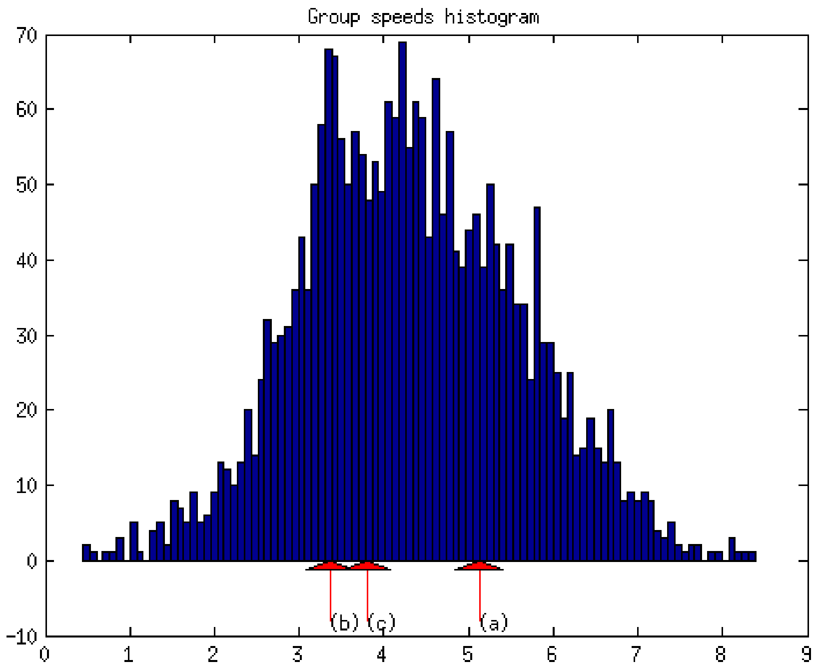

Figure 4: Distribution of average track speeds. Arrows indicate the average speed of three individuals (a), (b) and (c).

histogram of the speeds of approximately 190 dairy cows inspected by this system over a number of weeks.

The relative speed of a cow with respect to the group on its own is not sufficient to detect lameness; a cow may be consistently slower than the group due to old age or simply due to its own preferred pace of walking. The arrows on Figure 4 indicate the mean speed of a number of the observed individuals. We therefore propose to detect lameness by monitoring each individual cow's speed over a number of days to look for consistent changes in mean speed, excluding those caused by bunching of cows as they leave the milking parlour.

Identifying individuals

Key to monitoring an individual's speed is identifying each individual. While a number of technological solutions to this, such as RFID tagging, are possible, here we report on preliminary work on identify cows from CCTV which if reliable would be a cheaper, more robust alternative.

In computer vision, the problem of individual identification has been addressed repeatedly. Particularly promising methods are eigenfaces (Kirby & Sirovich, 1990; Turk, 1991) and the use of SIFT features (Lowe, 2004).

We extract an image representing each cow's body by capturing the region immediately behind the cow's head when she is walking in an approximately straight line (e.g., Figure 4). Principal components analysis (PCA) is used to find the subspace of these images which best approximates the full space. The principal components (eigencows)

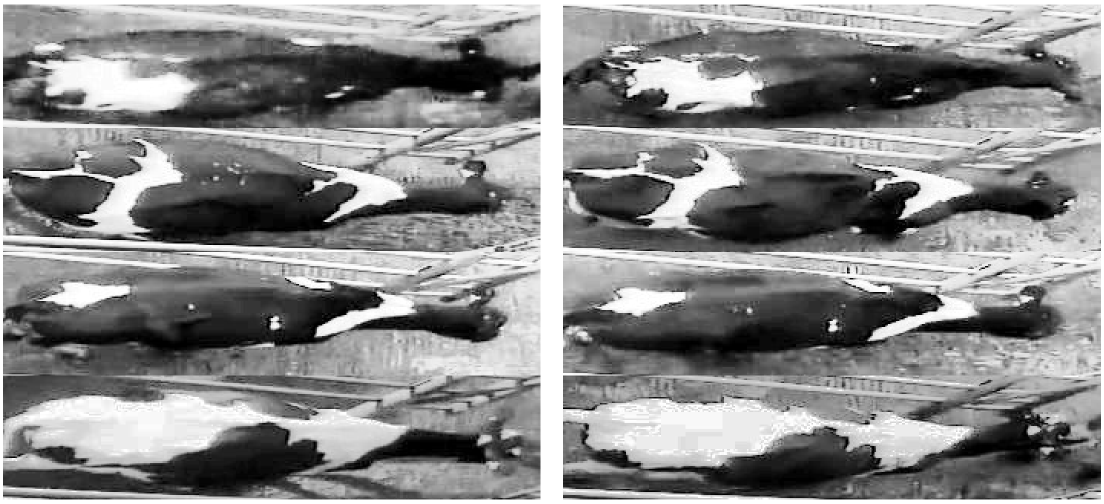

Figure 5: Left column: unidentified images from testing set. Right column: identified images from training set.

capture the main variation in the data set and discarding those representing small variations helps remove noise. Here 500 sample cow images were used to create a subspace of 150 dimensions. A cow is now identified by projecting her image onto the 150-dimensional space and finding the nearest neighbour to the projections of cow images in the training set, whose identities are known.

Figure 5 shows example images from the training set (left-hand column) which were identified as the closest matches to test images shown in the right-hand column. All test images were taken from videos recorded at different milking sessions to the training images. As the figure illustrates the use of principal components allows matching of images in the presence of noise, focus and lighting.

While PCA provides matching of global image information, scale invariant feature transform (SIFT; Lowe, 1999, 2004), features characterise the local structure of an image such as elements of the patterns on a cow's back). SIFT features were extracted for *keypoints* in each cow image. As Figure 6 illustrates, a large proportion of SIFT features correspond in images of the same cow, whereas the proportion is low for different cows. Our initial work indicates that identification using SIFT features will be more robust than global features such as PCA.

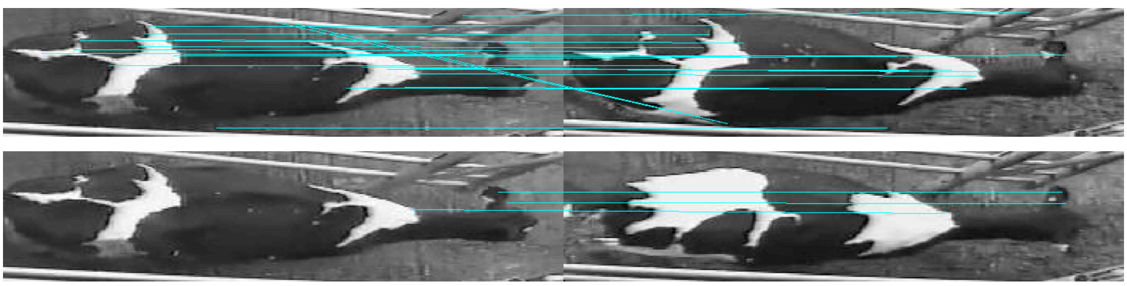

Figure 6: Matching local SIFT features. Lines are drawn between points with matching SIFT features for (top) different images of the same cow and (bottom) images of different cows.

Conclusions & further work

The principal contribution of this work is the introduction of a method for reliably detecting and tracking cows in video. This permits the easy measurement of their speeds which are well correlated with mobility scores and opens the way to characterisation of their gait and body condition monitoring.

We have also highlighted the need for individual identification and proposed methods for machine identification in video them based on the patterns on their back. Current work is on developing PCA and SIFT identification methods to allow lameness monitoring solely from video.

Acknowledgements

We gratefully acknowledge the assistance of Bicton Home Farm, particularly Mr Paul Voysey, Mr Paul Redmore and Mr Philip Squires.

References

Bell, N.J., Miedema H., Blajan I. & Mottram T., 2013 *Automatic measurement of mobility score in dairy cows using walking speed*. Computers and Electronics in Agriculture, in press.

Chapinal, N., de Passille, A. M., Pastell, M., Hanninen, L., Munksgaard, L. & Rushen, J., 2011. *Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle.* J. Dairy Sci., **94**(6), pp 2895-2901

Freund, Y. and Schapire, R.E., 1997, *A decision-theoretic generalization of on-line learning and an application to boosting*. J. Systems Sci., **55**, pp 119-139.

Kalman, R.E., *New results in linear filtering and prediction theory*. 1960. Journal of Basic Engineering, **82**(1) pp 35-45.

Kirby, M. and Sirovich, L., 1990, *Application of the Karhunen-Loeve procedure for the the characterization of human faces.* IEEE Transactions on Pattern Analysis and Machine Intelligence. **12**, pp 103-108.

Lowe, D., 2004, *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision, 60(4), pp 91-110.

Poursaberi, A., Bahr, C., Pluk, A., Berckmans, D., Veermae, I., Kokin, E. and Pokalainen, V., 2011, *Online lameness detection in dairy cattle using Body Movement Pattern (BMP).* International Conference on Intelligent Systems Design and Applications , pp 732 – 736.

Roweis, S., Ghahramani, Z., 1999. *A Unifying Review of Linear Gaussian Models*. Neural Computation **11**(2) pp 305–345.

Sonka, M., Hlavac, V., and Boyle, R. 2007, Image Processing, Analysis and Machine Vision, Chapman & Hall, London.

Turk, M.A., 1991, *Face recognition using eigenfaces*. Computer Vision and Pattern Recognition Proceedings, pp 586 – 591.

Viola, P. and Jones. M., 2004. *Robust Real-time Object Detection.* International Journal of Computer Vision, **57**(2) pp 137 – 154.