Towards Automatic Estimation of the Body Condition Score of Dairy Cattle Using Hand-held Images and Active Shape Models

Rafael Tedín[1], J. A. Becerra[1], and Richard J. Duro[1]
Ismael Martínez Lede[2]

[1] Grupo Integrado de Ingeniería,
Universidade da Coruña
{rafael.tedin,ronin,richard}@udc.es
http://www.gii.udc.es
[2] Feiraco Sociedad Cooperativa Galega
ismael.martinez@feiraco.es
http://www.feiraco.es

Abstract. The *Body Condition Score* (BCS) is considered a critical value for dairy farms, since its observation can be used to optimize milk production. Usually, the BCS is calculated by human experts after visual inspection in a time-consuming and subjective process. There are already some papers where this process is almost automated using image processing on some kinds of pictures and, in this work, the first steps towards a fully automated method based on pictures taken with common photographic cameras are described. Active Shape Models (ASM) are used to obtain a set of features that describe the back shape of cows and those features feed a classifier that computes the BCS. We show that the BCS can be estimated using only a set of angles from the back view with an error similar to that calculated between scores of two experts. To obtain those angles automatically is the hardest step in this process, but we have already achieved reasonable results on that point too.

Keywords: Body Condition Score, Active Shape Models, Computer Vision, Machine Learning

1 Introduction

The dairy sector has a vital importance in the economy of many countries because of its great weight within the food industry. This sector has been forced in recent years to be highly competitive and has started to incorporate all kinds of technical improvements to increase the productivity and maximize milk production per animal, minimizing, at the same time, other costs such as fuel for machinery, purchase of fodder and others.

In some cases, these improvements come from being able to gather detailed information about the condition of the animals over a long period of time in each farm. The information obtained allows for an estimation of the production of milk within the production–reproduction cycle that all cattle undergoes.

One of these indicators is the *Body Condition Score* (BCS), which roughly indicates the energy reserves of an animal by means of a numeric score. While there are different international body scoring systems [1], all of them rely on a numeric scale. In this paper we use a scale that spans from 1 to 5 in 0.25 intervals. The BCS is considered by field experts [1] a critical value for dairy farms, since its observation helps to determine issues such as the maximum peak of milk that a cow should deliver, whether the food supply is adequate in terms of balance, quality and quantity, the correct evolution of the gestation and even the overall health of the animal. Nevertheless, a single observation of the BCS of an animal is meaningless: it is the evolution of the score over the production–reproduction cycle that is relevant. Even more, to assess the overall state of the farm, the progression of a sufficient number of animals has to be tracked.

Despite the general agreement among field experts that periodically calculating the BCS is important, currently this task is not regularly performed and few farms have included it in their schedules [1]. The main reason is that this task is very time consuming as the raters have to be trained, and even trained experts need up to 30 seconds to estimate a score, which for herds of hundreds of cows can be unaffordable [2]. There is also some criticism on expert rating itself, since the score might depend on the subjectivity of the experts. This means that the scores might be inconsistent among raters [3] and this inconsistence can show up even between scores of the same expert for the same cow.

For dairy cattle, one of the first approximations that shows the feasibility of an automatic system for body condition scoring is [4]. The authors used a laser light to paint line patterns across the tailhead area of the cows and images of the animals were taken. After that, those images where manually labeled assigning points over each of the stripes that the laser produced and quadratic curves were fitted to these points, showing a positive correlation and hence the feasibility of an automatic scoring system. The potential of using only images for scoring is shown in [5], where experts rated live animals and then rated the same cows on images. No significant difference was found between the two scores for the same expert. An interesting result is that experts agree that they could make accurate ratings with the rear view but not with the side view. Another approach was demonstrated in [6]. Here, images were taken from the top of the cow and these were manually labeled with 23 anatomical points from the cow's contour. From these 23 points, 15 angles were extracted and a regression model was built upon them showing a good correlation with the expected BCS. In [7], thermal images of the rear of the cow were taken from above. The authors tested the hypothesis that fatter animals should be rounder adjusting the segmented contour to a parabola. They showed that the fitting error was correlated with the BCS. Finally, [8] presents a schematic for building a system for semiautomatic body condition scoring dividing it into two blocks: a training and an execution block. In both blocks images are taken from above with a standard digital network camera and the same anatomical features as in [6] are manually identified. Instead of measuring the angles as in [6], in [8] they construct a model of the cow's shape using statistical shape analysis. The parameters of this model are used for training a regression model in the training block and for estimating the BCS in the execution block. They show results at least as good in [6] and [7].

We are not aware of any method that fully automates the process of BCS assessment. In particular, each method requires, at least, a manual image labeling step. Additionally, some methods might require expensive equipment such as thermal cameras. Note also that the last methods all follow a similar setup were it is assumed that images can be taken from above in a fixed position. As we will see in Section 3, this is not always possible or preferable.

Our goal is to build a system for automatic body condition scoring for dairy cattle that overcomes the flaws of expert scoring, so that the farms can rely on repeatable and objective results. We try to achieve this objective through the use of machine learning algorithms and, since the experts' judgments are based on visual information, by computer vision techniques such as Active Shape Models (ASM) [9]. We count for this purpose with the collaboration of Feiraco Sociedad Cooperativa Galega and some of their experts.

It's important to know that most farms that are partners of the cooperative Feiraco are family businesses. Those farms are quite isolated from each other and have a relative small size (in terms of revenues and space). In addition to the stated flaws of expert scoring, isolation lends the access to scoring experts difficult. On the other hand, the small size of the farms means that no expensive or complicated setup is affordable for a system for automatic condition scoring. This discards a fixed universal setup and, in turn, suggests that a mobile system is a better solution since it could be fixed if a farm can afford it and can still be used by farms that can't. Furthermore, a mobile system would also be useful for veterinarians and other experts who visit the farms regularly and even for expert training.

In this paper we show the current state of our work towards the mentioned objective. Section 2 introduces Active Shape Models. In section 3 we describe our approach to this problem and the current state of the system, the experiments that we have executed and the results that have been achieved so far. Finally, section 4 summarizes the conclusions of our current work and enumerates some of our future lines of work.

2 Active Shape Models

Active Shape Models [9], also known as smart snakes, are deformable shape models that use statistical shape analysis to build a model that restricts the allowed variations of a shape to those seen on a given labeled training set of s images. A shape is a set of n points, called landmarks, that can be represented by a vector $(x_1, \ldots, x_n, y_1, \ldots, y_n)^T$ where x_i and y_i are, respectively, the x and y coordinate of point i, $i = 1 \ldots n$. After (optionally) aligning and normalizing all the shapes, the mean shape $\bar{\mathbf{x}}$ is calculated and a principal component analysis of the shapes is performed. Only enough eigenvectors (i.e. principal components) are retained to ensure that the model covers an acceptable amount of variability. The model itself has the form

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \ . \tag{1}$$

where $\mathbf{P}$ is a matrix containing the selected principal components and $\mathbf{b}$ are the so-called shape parameters.

A local appearance model for each point of the shapes is built by sampling normal profiles to the shape boundary. Those profiles are constructed by sampling k pixels on each side of a point, giving profiles of length $2k+1$. The first derivatives of the profiles are computed using finite differences of their components and these derivatives are then normalized so that the absolute sum of its values equals 1. If we denote, for a particular landmark, those normalized first derivatives as $\mathbf{g}_i \ldots \mathbf{g}_s$, a local appearance model for the gray values around that landmark can be constructed using the mean $\bar{\mathbf{g}}$ and the covariance $\mathbf{S}_g$, assuming that $\mathbf{g}_i \ldots \mathbf{g}_s$ is drawn from a multidimensional Gaussian distribution, as follows

$$f(\mathbf{g}_{new}) = (\mathbf{g}_{new} - \bar{\mathbf{g}})\, \mathbf{S}_g^{-1}\, (\mathbf{g}_{new} - \bar{\mathbf{g}}) \ . \tag{2}$$

which is the Mahalanobis distance between a new profile $\mathbf{g}_{new}$ and the observed profiles in the training set. Minimizing this distance is equivalent to maximizing the probability that $\mathbf{g}_{new}$ is drawn from the multidimensional Gaussian distribution.

Starting from a initial estimate of the shape, for example the mean shape, each landmark is moved on each iteration of the search along the direction normal to the shape boundary $n_s - k$ positions on each side of the landmark. For each displacement a profile is sampled, its normalized first derivate is calculated and the corresponding Mahalanobis distance is evaluated. The landmark is finally moved to the position with the least distance. After moving all the landmarks, the shape model (1) is fitted to the displaced points ensuring that the shape is consistent with the model. In particular, the shape parameters $\mathbf{b}$ are computed and, normally, it is ensured that $|b_i| \leq 3\sqrt{\lambda_i}$, where λ_i are the eigenvalues of the selected components computed during the principal component analysis. This process is repeated a fixed number of iterations N_{max} or until a convergence criterion is met.

The search can also be run in a multiresolution framework of L_{max} levels. In such a framework, level $i+1$ has a coarser resolution than level i, being the level with the finest resolution the original image. This requires calculating independent $\bar{\mathbf{g}}$ and $\mathbf{S}_g$ for each resolution level. The search starts at the coarsest resolution level and moves to the next level after N_{max} iterations or after the convergence criterion is met.

3 Approach, Experiments and Results

We have decomposed the problem into three subproblems and we address the first two in this paper. The first subproblem is to find a method that, given an image of a cow, gives us a representation of the anatomy of the cow that can be handled computationally. The second subproblem is to train a model that, given a representation of the cow, estimates its BCS. Finally, the last subproblem is combining the solutions of the other subproblems to build the whole system. As a suitable representation of the cow we chose the shape of the animal given by a set of interconnected anatomical points and we use Active Shape Models (see Section 2) to find this shape and machine learning classifiers to estimate the BCS. The following three sections provide the details of the experiments we have performed.

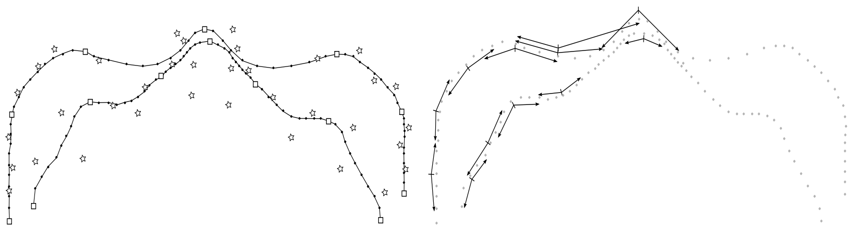
3.1 Labeled Set of Images and Inter-rater Error Assessment

We used a Canon 300D consumer camera and its standard 18-55mm kit lens for taking hand-held images of cows in several farms. We restrict the images to the rear view of the cows since, as seen in [5], it should be enough. The focal length chosen was around 28mm, depending on the space behind the cow, so as to have more or less the same size for the cows in the images. All cows were centered in the images while shooting. We took images for each cow and at least one of the two experts estimated the BCS following their established procedure.

Out of all the images, we selected a set of 125 images having scores between 2 and 5 from a scale that ranges from 1 to 5. No cow with a BCS of less than 2 was observed and most of them have scores between 3 and 4. The images were cropped and resized to a resolution of 255x170 pixels. With this resolution we find that most features of the rear of the cow are distinguishable and it represents a good compromise for less powerful devices such as smartphones, as an objective for future work is that such devices can perform the BCS computation.

We need a computationally tractable representation of the anatomy of the cow for estimating the BCS from measures taken from it. For this purpose, we choose the shape of the rear of the cow. In order to facilitate image processing, we developed an ad-hoc Java tool to annotate each of the 125 images with 14 good recognizable anatomical points, which represent the landmarks of the shape. Annotating images is a very time consuming process but, on the other hand, the more points we introduce, the better the shape is described. For better capturing the curvatures of the shape and, at the same time, making the process more agile, we added control points between the landmarks, which enabled us to treat the shape as two splines of Bézier curves of degree four. With this representation, it is straightforward to automatically generate new points between landmarks, even if the curvature is not a clear edge as it occurs when dealing with fatter cows. Note that those automatically generated points are not true landmarks, but an approximation to them. A schematic of the representation of the rear of the cow can be seen in Fig. 1a.

For consistency, the labeled image set incorporates only scores from the same expert. But, as we had two experts, we estimated the error that could be assumed by a classifier or the future whole system by calculating the discrepancy between those two experts. Both scored a set of the same 69 cows at the same time we



(a) (b)

Fig. 1: Schematic of landmarks (squares), control points (stars) and automatically generated points (knots) of the rear of a cow (a) and schematic of the 22 angles used for estimating the BCS (only the angles on the left side are shown) (b).

took images of them. We compared the results and calculated the root mean squared error (0.32) and its 95% confidence interval using boosting (0.27, 0.37).

3.2 Experiments with Classifiers

Once we have a representation that can be handled computationally and a set of images labeled with this representation and the corresponding BCS, it has to be shown that we can estimate the score using measures taken from it. As in previous works, such as [6], the measurement of angles gave good results, we decided to start with this approach. A schematic of the set of 22 angles chosen can be seen in Fig. 1b. An angle is measured given three points: a central point and two edge points. Angles are normalized between 0 and 1. But instead of trying to fit regression models to the data, we wanted to test the feasibility of other machine learning algorithms. Initially, we focused on a set of classifiers that can handle numeric class attributes since the BCS has this form.

We selected WEKA [10] as our basis test tool, since it includes many classifiers and comes in the form of a Java library easy to use within other developments. We have respected the WEKA class names in Table 1 so that the WEKA documentation of the classifiers can be consulted. All classifiers were used with the standard parameters from the WEKA distribution (version 3.6.6) unless stated otherwise. A special classifier is ZeroR, since it always returns the mean value of the BCS. It is used as a control classifier for assessing a maximum error value.

Table 1 summarizes the errors obtained by the classifiers. The errors were calculated using a 10-fold cross-validation process.

Comparing the confidence intervals of the errors between experts in Section 3.1 with those between the classifiers and an expert in Table 1, we see that the BCS can be reliably estimated using only the rear view of the cow by the M5P classifier.

Table 1: Root mean squared error (RMSE) and its confidence interval of the results of the classifiers compared to the expert's scores for the labeled image set.

Classifier	**RMSE**	**95% C.I.**
ZeroR	0.79	(0.71, 0.87)
M5P	0.43	(0.36, 0.50)
KStar	0.58	(0.50, 0.65)
IBk ($k = 1$)	0.52	(0.44, 0.59)
IBk ($k = 2$)	0.47	(0.41, 0.53)
GaussianProcesses	0.46	(0.40, 0.53)

3.3 Experiments with ASM

In this section we compare several ASM setups in order to have information about which one could be more suitable for our problem and to identify where the ASMs have to be improved. Since the ASMs have quite a few parameters to tune, we performed preliminary tests and determined that the following parameters are a good starting point: $k = 5$, $n_s = 10$, $N_{max} = 10$ and a covered variability of 98% for the PCA. We align then the shapes and generate 9 points automatically between each landmark. In the future, more exhaustive tests involving these parameters will be carried out.

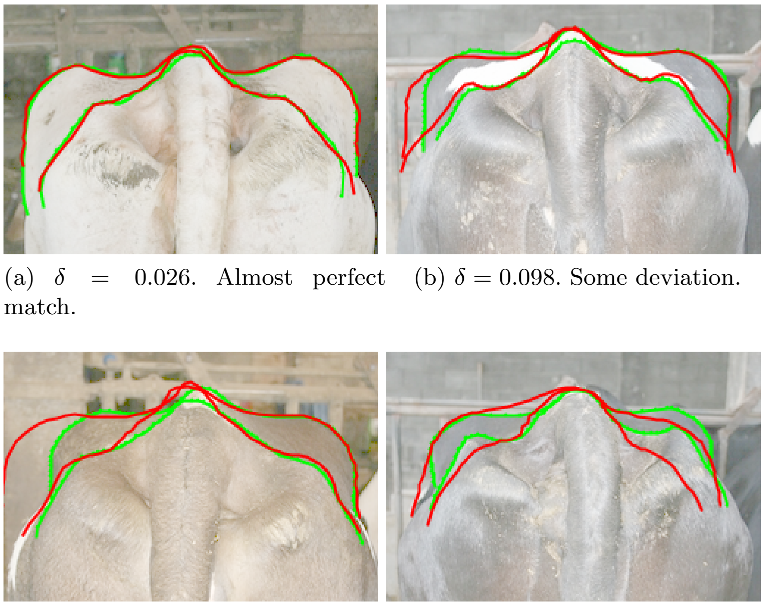
We ran a set of different setups, using an ASM library written in Java in our research group, the results of which can be seen in Table 2a. The first column indicates the setup. "Canonical" refers to the standard setup described in Section 2. "HSI" means that instead of using just gray level values of the image, the components of this color space are employed, creating a separate model for each component and averaging their outputs. Since the ASM uses a local search algorithm, it can easily get trapped in local optima depending on the first estimation of the shape. With "MS", which means multi start, we run the search several times displacing each time the first estimate vertically and scaling it. The best shape is assumed to be the one with less average output of its appearance models, since that should mean that on average each landmark is near its optimal position. Finally, "M5P" means that instead of using (2) to build local appearance models, a trained M5P classifier is used, where M5P is the same classifier seen in Section 3.2. The model for a point is trained by sampling normalized first derivative profiles along the normal to the shape boundary on each image and on each side of the point. Each profile is assigned a fitness which is a function (3) of the displacement from the central pixel of the profile to the point.

$$fitness\,(profile) = 1 - e^{-\left(\frac{displacement}{n_s - k}\right)^2} \quad . \tag{3}$$

To estimate the quality of the setups, we computed what we call the relative area difference:

$$\delta = \frac{\sum_i^n d\,(e_i, o_i)}{\sum_i^n A\,(e_i)} \quad . \tag{4}$$

Where e_i is the i-th expected subshape of the shape out of n subshapes, o_i is the i-th subshape found after searching and $A(s)$ returns the area of subshape s taken as a closed path. On the other hand, $d(s_1, s_2)$ computes the difference of areas of subshapes s_1 and s_2 as follows. Since the subshapes are open paths, we append s_2 to s_1 and close the path. Depending on the relative orientation of the two subshapes, a set of polygons is hereby created as the path crosses itself and the result of $d(s_1, s_2)$ is defined to be the sum of all these polygons' areas. This result is divided by the sum of all expected subshape areas in order to make this measure independent of the resolution and shape size. In a perfect match $\delta = 0$. Since we want to measure angles, we claim that this measure is better than others that consider distances between points because it is less dependent on the automatically generated points, which are less reliable than true landmarks. In Fig. 2 we show several examples of executions of our library and the associated δ values.



(a) $\delta = 0.026$. Almost perfect match.

(b) $\delta = 0.098$. Some deviation.

(c) $\delta = 0.125$. Evident deviation.

(d) $\delta = 0.179$. Shape totally missed.

Fig. 2: Examples of different values of δ. The expected shape is depicted in green, the obtained one in red.

As shown in Table 2a, the best results are obtained using a M5P classifier as the local appearance model (instead of the Mahalanobis distance) with a multistart search and the HSI color space. The combinations tried are, of course, not exhaustive, so there may be some combination of components of different color spaces and search algorithm that produces better results. This will be the scope of some of our future work to find such a combination, perhaps by applying an evolutionary search. In Table 2b we see two examples of δ values for the top and bottom subshapes separately. We have observed that, consistently, the values for the top subshape are better than those for the bottom subshape. The top subshape lies on a hard boundary and, since we are sampling the gradient, it is easier to find. On the other hand, the boundary of the bottom shape might vanish for fatter cows and becomes difficult to find. We also realized that some images are, perhaps, not valid. This becomes evident in the many outliers we have observed in our results that eventually pushed the δ values up. We need some sort of mechanism to get rid of them.

Table 2: Results of ASM tests for the whole shape (a) and examples for the top and bottom subshapes separately (b) . The median of the relative area difference (δ) and 90% confidence intervals (90% C.I.) are shown. See the text for details.

(a)

Setup	δ	**90% C.I.**
Canonical	0.107	(0.096, 0.118)
HSI	0.113	(0.105, 0.124)
HSI+MS	0.128	(0.113, 0.148)
M5P	0.099	(0.092, 0.114)
M5P+MS	0.107	(0.092, 0.113)
HSI+M5P	0.098	(0.090, 0.107)
HSI+M5P+MS	0.081	(0.073, 0.094)

(b)

Setup	δ	**90% C.I.**
Canonical (top)	0.080	(0.069, 0.093)
HSI+MS+M5P (top)	0.054	(0.042, 0.069)
Canonical (bottom)	0.122	(0.114, 0.135)
HSI+MS+M5P (bottom)	0.102	(0.087, 0.115)

4 Conclusions

In this paper we have shown the current state of our work towards a fully automatic system for assessing the *Body Condition Score* (BCS) for dairy cattle. We gave an outline of the problem we want to solve and decomposed it into three subproblems: to find the back shape of cows given its pictures, to classify those shapes assigning a BCS and, finally, to build up the complete system. In the current work, we have addressed the first two subproblems. Using Active Shape Models, our results show that the Mahalanobis distance is not very suitable for building local appearance models in our problem and instead that a machine learning classifier performs better. We have also shown that using more components than just gray values, which is the common choice, improves the search with ASM. The use of a multistart strategy for the search also improves it. We have demonstrated that the automatic estimation of the BCS is feasible with the use of machine learning classifiers employing a suitable representation of the back view of the animals (i.e. its shape). It has been also shown that the error of such classifiers is comparable with that of two human experts, which means that a reliable system can be built. There remain though open questions about which components of color spaces are the best suited to our problem and how to automatically detect what images are not valid for our system in order to obtain a correct BCS. These questions will be addressed in future work.

Acknowledgments. This work was partially funded by the Spanish MICINN through project TIN2011-28753-C02-01 and the Xunta de Galicia and European Regional Development Funds through project 09MRU024E.

References

1. Bewley, J.M., Schutz, M.M.: An Interdisciplinary Review of Body Condition Scoring for Dairy Cattle. The Professional Animal Scientist **24** (2008) 507–529
2. Upham, G.: The Use of Body Condition Scores in Grouping of Lactating Cows. Compendium on Continuing Education for the Practicing Veterinarian **12** (1990) 581
3. Leroy, T., Aerts, J.M., Eeman, J., Maltz, E., Stojanovski, G., Berckmans, D.: Automatic Determination of Body Condition Score of Cows Based on 2D Images. Wageningen Acad. Publishers (2005)
4. Wu, J.R., McFarlane, N., Siebert, X.J.L.P., Schofield, P.: Genetic Evaluations of Dairy Bulls for Daughter Energy Balance Profiles Using Linear Type Scores and Body Condition Score Analyzed Using Random Regression. Journal of Dairy Science **86** (2003) 2205
5. Ferguson, J.D., Licitra, G.A.G.: Body Condition Assessment Using Digital Images. Journal of Dairy Science **89** (2006) 3833–3841
6. Bewley, J.M., Peacock, A.M., Lewis, O., Boyce, R.E., Roberts, D.J., Coffey, M.P., Kenyon, S.J., Schutz, M.M.: Potential for Estimation of Body Condition Scores in Dairy Cattle from Digital Images. Journal of Dairy Science **91** (2008) 3439–3453
7. Halachmi, I., Polak, P., Roberts, D.J., Klopcic, M.: Cow Body Shape and Automation of Condition Scoring. Journal of Dairy Science (2008) 4444–4451
8. Azzaro, G., Caccamo, M., Ferguson, J., Battiato, S., Farinella, G., Guarnera, G., Puglisi, G., Petriglieri, R., Licitra, G.: Objective Estimation of Body Condition Score by Modeling Cow Body Shape from Digital Images. Journal of Dairy Science **94**(4) (2011) 2126 – 2137
9. Cootes, T.F., Taylor, C.J.: Active Shape Models – 'Smart Snakes'. In: Proceedings of the British Machine Vision Conference, Springer-Verlag (1992) 266–275
10. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: an Update. SIGKDD Explor. Newsl. **11**(1) (2009) 10–18