Automatic classification of sheep behaviour using 3-axis accelerometer data

Jacques Marais*, Solomon Petrus Le Roux†, Riaan Wolhuter‡ and Thomas Niesler§
Department of Electrical and Electronic Engineering
Stellenbosch University, South Africa
Email: {16037103*, 16084454†, wolhuter‡, trn§}@sun.ac.za

***Abstract*—Monitoring animal behaviour can prove challenging when working in inaccessible environments. This problem can be addressed by using animal attached accelerometers and automatic classifiers. This study considers the feasibility of using specially designed hardware to capture three-dimensional accelerometer data from sheep and to subsequently automatically classify their behaviour on the basis of these measurements. Five common behaviours have been identified: Lying, standing, walking, running and grazing. Linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) classifiers were trained based on 10 features. A greedy selection procedure was used to determine which features provide the highest classification accuracy. It is shown that both classifiers can automatically identify the five behaviours with high accuracy when all the features are used for training. The LDA and QDA classifiers achieved an overall accuracy of 87.1% and 89.7% respectively. Grazing was misclassified the most in both classifiers, because it was confused with lying. This result was expected considering the high similarity between the raw accelerometer data associated with grazing and lying. The QDA classifier showed larger improvements when using a smaller number of features.**

I. Introduction

The study of animal behaviour provides us with a better understanding and insight to their health, which is a priority in conservation attempts. While monitoring animals in a supervised environment is relatively simple, this is not true for unsupervised animals in their natural environment. In this latter case, monitoring can be achieved by attaching sensors to the animal. Depending on the type of sensor, information concerning the animals energy expenditure, location, speed, heart rate, temperature and behaviour can be monitored. This provides researchers with information on the behaviour of animals in environments that are inaccessible for human observation.

An accelerometer is often used in animal monitoring systems. It can be embedded in a lightweight tag which can be attached to the animal with minimal obstruction or interfering with the natural movement of an animal. Mono- and bi-axial accelerometers are available, but have a few disadvantages compared to a tri-axial accelerometer [1]. Measuring the acceleration in all three axes provides a more complete picture of animal movement patterns and could reveal information that would have otherwise been missed.

Several studies have considered the classification of animal behaviour with the aid of accelerometers. Goat grazing behaviour has been automatically classified with high success rate using tri-axial accelerometers [2]. In this work moving averages were used in conjunction with selected threshold values to distinguish between resting, eating and walking. A recent study extended this work by using the k-nearest neighbour classifier [3]. It found that the classification accuracy was comparable to what was achievable using more complex techniques. In different work, the recognition of cow behaviour was investigated by applying support vector machines to 3-axis accelerometer data [4]. This study attempted to distinguish between eight behaviours and achieved an accuracy of over 80% for all the classified behaviours. Another study combined support vector machine and hidden-Markov models to classify the behaviour of cheetahs [5]. In this way it was possible to classify five minute segments intro three behaviour classes (feeding, mobile and stationary). Finally, the behaviour of vultures was classified using classification and regression trees, random forests, support vector machines and artificial neural networks with linear discriminant analysis (LDA) [6]. Although the LDA classifier was intended as a baseline, it was found that all the algorithms had an accuracy of 80% or higher, with the non-linear algorithms outperforming LDA.

This study considers the feasibility of using specially designed hardware to capture three-dimensional accelerometer data and subsequently automatically classify the behaviour of sheep using LDA as well as quadratic discriminant analysis (QDA). LDA was chosen for this preliminary study because it was used as a baseline classifier in a previous study. The QDA classifier was subsequently investigated and compared to LDA, since it is a simple extension of the LDA classifier. The classifiers were used to distinguish between five common behaviours types.

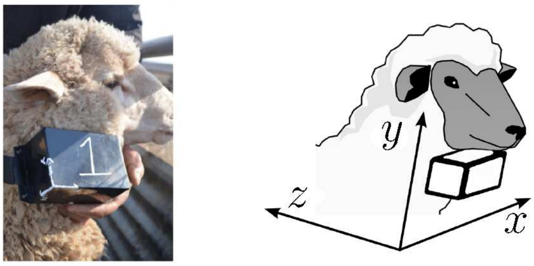
II. Material and methods

A. Site and animals

Data collection was performed on a farm (Rooivlei) in Carnarvon, Northen Cape, in July 2014. The owner of the farm provided access to the livestock for the experiment. Five sheep were randomly chosen on each of 3 separate days. The sample size was constrained by what was practically feasible at the time of the study. The sheep were held in a camp large enough to allow for unrestricted movement throughout the day.

B. Hardware

The hardware (referred to as the tag hereafter) consisted of a 3-axis accelerometer, SD card, microcontroller and battery. The tag was designed to allow the accelerometer to be constantly sampled at a frequency of 100Hz and to subsequently store the measurements on the SD card. The tag was enclosed in a durable casing and fitted around the necks of the sheep



(a) (b)

Fig. 1. (a) A photograph showing how the tag was attached around the sheep's neck. (b) A sketch of the accelerometer axes relative to the sheep

TABLE I. FIVE OBSERVABLE BEHAVIOURS OF THE SHEEP

Behaviour	Description
Lying	The sheep is lying down with its head on the ground.
Standing	The sheep is standing with its head up.
Walking	The sheep is walking at a slow pace.
Running	The sheep running at a fast pace.
Grazing	The sheep is grazing with its head down while standing still or walking slowly.

using collars as seen in Figure 1a. The hardware was placed in approximately the same orientation on all the tested animals.

Figure 1b shows the orientation of the accelerometer's axes when the tag is attached around the sheep's necks. In the same figure it can be seen that movement in the x-axis is associated with the left-right movement of the sheep. The y-axis is associated with the up-down movement of the sheep and the z-axis is associated with the forward-backwards movement of the sheep.

C. Data collection

Data was gathered on three separate days. The sheep were collected in the mornings and the tags were fitted around their necks. The sheep were lead to the camp and left undisturbed for the duration of the day. The hardware logged the acceleration of the sheep at a rate of 100 samples per second throughout the day. Five common behaviours (Table I) were identified: lying, standing, walking, running and grazing. The sheep were monitored from a distance on the three testing days and the behaviour of the herd and corresponding time-stamps were manually logged. The sheep were also chased on various occasions to collect "run" data, because there was little reason for them to run without provocation.

At the end of the day the sheep were collected, the collars were removed, the data was downloaded from the SD cards and saved to a file. Each data file included the following information of interest: an ID of the current logging session, timestamps in milliseconds and the acceleration along the three axes.

Each data file was manually segmented and labelled with the corresponding behaviour by reconciling the logged data with the manual annotations. A summary of the data collected for each behaviour and sheep is displayed in Table II. The sheep ID consists of the testing session (day 1, 2 or 3) and the tag's number. "Lying" behaviour was not observed during the first testing session.

TABLE II. DISTRIBUTION OF DATA COLLECTED FOR EACH BEHAVIOUR CLASS AND SHEEP (HH:MM:SS).

Sheep	Lying	Standing	Walking	Running	Grazing	Total
1_tag1	0:00:00	0:03:43	0:09:27	0:04:14	0:13:37	0:31:01
1_tag2	0:00:00	0:02:46	0:08:55	0:04:16	0:17:32	0:33:29
1_tag3	0:00:00	0:01:32	0:10:45	0:05:24	0:19:26	0:37:07
1_tag4	0:00:00	0:04:13	0:07:36	0:05:02	0:09:20	0:26:11
1_tag5	0:00:00	0:04:03	0:08:04	0:05:31	0:08:01	0:25:39
2_tag1	5:08:43	0:13:28	0:55:27	0:05:45	1:11:45	7:35:08
2_tag2	5:52:14	0:04:37	1:07:50	0:05:45	1:43:18	8:53:44
2_tag3	1:17:24	0:01:09	0:28:48	0:04:20	0:21:28	2:13:09
2_tag4	1:23:32	0:04:06	0:20:51	0:04:52	0:11:17	2:04:38
2_tag5	0:21:12	0:00:34	0:04:58	0:04:22	0:03:09	0:34:15
3_tag1	0:29:59	0:03:44	0:24:16	0:04:18	0:15:50	1:18:07
3_tag2	0:39:33	0:00:44	0:25:29	0:04:19	0:23:00	1:33:05
3_tag3	0:30:24	0:06:49	0:31:22	0:04:16	0:27:02	1:39:53
3_tag4	0:54:56	0:02:33	0:33:13	0:04:16	0:22:15	1:57:13
3_tag5	0:40:05	0:00:53	0:31:14	0:04:50	0:20:17	1:37:19
Total	17:18:0[illegible]	0:54:54	6:08:15	1:11:30	6:27:17	1d 7:59:58

D. Linear discriminant analysis

Linear discriminate analysis (LDA) is a simple linear model used for classification. This classification technique can assign an input vector x (features) to one of K classes (C_k) using a linear decision boundary. An underlying assumption of LDA is that all classes share the same covariance matrix. LDA can be extended to quadratic discriminant analysis (QDA) if a covariance matrix for each class is employed. This changes the decision boundary between classes from linear to quadratic [7].

1) The two class case: Fisher's linear discriminant is used to solve this classification problem for higher dimensional data by using the following criterion [8]:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \tag{1}$$

where $\mathbf{S}_B$ is the between-class covariance matrix:

$$\mathbf{S}_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T \tag{2}$$

and $\mathbf{S}_W$ is the within-class covariance matrix:

$$\mathbf{S}_W = \sum_{i=1,2} \sum_{x \in C_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \tag{3}$$

Lastly, $\mathbf{w}$ is the optimal projection matrix. The criterion $J(\mathbf{w})$ can be maximized by differentiating with respect to $\mathbf{w}$ to find:

$$(\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w} = (\mathbf{w}^T \mathbf{S}_W \mathbf{w}) \mathbf{S}_B \mathbf{w} \tag{4}$$

From Equation 2 it can be seen that $\mathbf{S}_B\mathbf{w}$ will always be in the direction of $(\mathbf{m}_2 - \mathbf{m}_1)$. Furthermore, since only the direction of $\mathbf{w}$ is important, the scalars $(\mathbf{w}^T\mathbf{S}_B\mathbf{w})$ and $(\mathbf{w}^T\mathbf{S}_W\mathbf{w})$ can be dropped. By multiplying by $\mathbf{S}_W^{-1}$, Equation 4 reduces to the transformation vector:

$$\mathbf{w} \propto \mathbf{S}_W^{-1}(\mathbf{m}_2 - \mathbf{m}_1) \tag{5}$$

Fisher's linear discriminant tries to project D-dimensional data to a lower dimension in the direction which that maximises the projected difference in between-class means while minimising the within-class variance. After the data has been projected to a lower dimension, a linear (or quadratic) boundary can be selected to separate the classes and new data can then be classified.

2) The multi-class case: Fisher's linear discriminant can be generalised to multiple classes ($K > 2$). To do this, the between-class and within-class covariance matrices must be reformulated to incorporate all of the K classes. Equations 2 and 3 can be rewritten as [8]:

$$\mathbf{S}_B = \sum_{k=1}^{K} N_k(\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \tag{6}$$

and

$$\mathbf{S}_W = \sum_{k=1}^{K} \sum_{x \in C_k} (\mathbf{x} - \boldsymbol{\mu}_k)(\mathbf{x} - \boldsymbol{\mu}_k)^T \tag{7}$$

Where N_k is the number of samples in class k, $\boldsymbol{\mu}_k$ is the mean for each class and $\boldsymbol{\mu}$ is the total mean of all the classes. Again the criterion in Equation 1 must be maximized in order to find the decision boundary. This can be obtained by solving the generalised eigenvalue problem [9]:

$$\mathbf{S}_B\mathbf{W} = \lambda\mathbf{S}_W\mathbf{W} \tag{8}$$

Eigenvalue decomposition of $\mathbf{S}_W^{-1}\mathbf{S}_B$ can be used to find the projection matrix. $\mathbf{W}$ will contain the eigenvectors resulting from this decomposition.

E. Data preprocessing and feature extraction

The accelerometer data was preprocessed using Python's Scipy library (version 0.14.0) [10] as shown in Figure 2. Each segment file was loaded and examined to make sure the segment contains at least 512 samples in order to have a epoch length of 5.12 seconds. An epoch length of 5.12 seconds was selected to window the segments as it was found in [11] that a 3-5 second epoch was optimal. The three acceleration values were then low-pass filtered using an 8th-order Butterworth filter with a cut-off frequency of 10Hz before applying a Hamming window. Consecutive features were obtained by allowing a 50% overlap in this data window. The features extracted were based on those recommended in [4] and [12] and included the following 10 features that consisted of 31 variables:

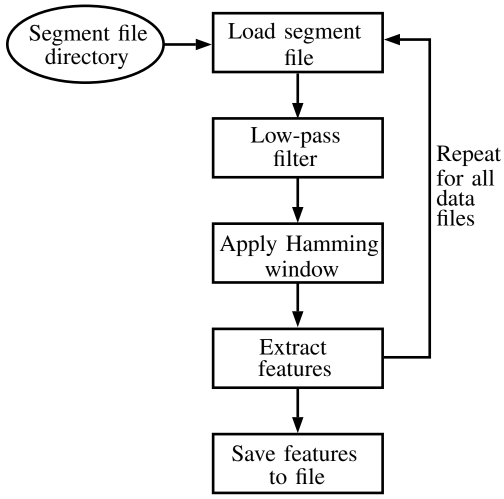


Fig. 2. Flow chart of the feature extraction process.

- Mean
- Standard deviation (STD)
- Variance
- Skewness (Skew)
- Kurtosis
- Maximum and minimum value (MaxMin)
- Energy
- Frequency-domain entropy (FreqEntropy)
- Pairwise correlation between the axes (Corr)
- Average signal magnitude vector (ASMV)

The skewness and kurtosis was calculated for each axis. The skewness provides insight on the symmetry of a frame while the kurtosis serves as an indication of peakedness relative to the normal distribution. The aforementioned was accomplished by calculating the third and forth standardized moment of each axis per frame. The "MaxMin" feature contains the maximum and minimum value for each axis in a frame. This results in a vector with 6 values which represents the dynamic range. The energy was calculated as the sum of the squared FFT magnitudes with the DC component excluded. The sum was then normalised by the window length. The frequency-domain entropy was calculated by normalising the FFT and then treating the FFT magnitudes as discrete probabilities. The frequency entropy has been found useful to discriminate between behaviours with similar energy values [12]. The cross-correlation between each axis pairs were calculated as the ratio between the cross-covariance and the product of the standard deviations. The average signal magnitude vector (ASMV) for each frame was calculated by summing the vector length over the frame and normalising the result with the frame's length.

The distribution of each feature was visualised in a 1-dimensional histogram to visually inspect the underlying distribution. It was found that the "Skew" and "MaxMin" features resembled a Gaussian distribution, while the rest of the features had a closer resemblance to a Rayleigh distribution.

TABLE III. NUMBER OF USABLE FRAMES EXTRACTED FOR EACH SHEEP AND BEHAVIOUR.

Sheep	Lying	Standing	Walking	Running	Grazing	Total
1_tag1	0	59	145	65	229	498
1_tag2	0	46	141	72	330	589
1_tag3	0	27	198	88	359	672
1_tag4	0	73	123	82	157	435
1_tag5	0	58	121	84	125	388
2_tag1	7202	248	1003	94	1214	9761
2_tag2	8237	75	1291	101	1803	11507
2_tag3	1807	23	560	76	392	2858
2_tag4	1954	82	357	88	201	2682
2_tag5	494	11	84	74	53	716
3_tag1	697	64	447	69	260	1537
3_tag2	913	12	464	69	402	1860
3_tag3	707	130	610	71	508	2026
3_tag4	1281	46	608	63	390	2388
3_tag5	934	14	559	73	340	1920
Total	24226	968	6711	1169	6763	39837

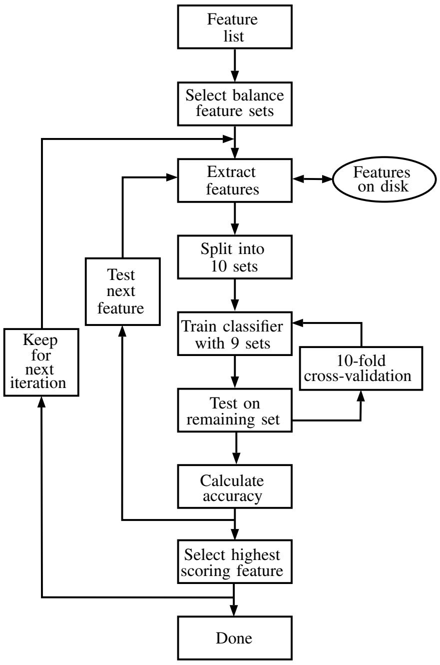


Fig. 3. The greedy feature selection procedure.

A breakdown of the number of usable frames extracted for each sheep and class is presented in Table III

F. Classification

The Python LDA en QDA implementations in Scikit Learn (version 0.14.0) were used to determine the best features to use for sheep behaviour classification.

A greedy feature selection procedure, as illustrated in Figure 3, was used to determine the best combination of features. For each iteration of the procedure, a balanced subset (an equal number of frames for each class) was selected for classification. The procedure tested each feature's classification accuracy and then retained the feature with the highest score. The accuracy was defined as the overall correctly classified feature vectors as a proportion of the total classified feature vectors [3]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Where TP, TN, FP and FN are the true positive, true negative, false positive and false negative counts respectively. The procedure iterated until all the features were used in classification. 90% of the balanced subset was used for training the classification algorithm and the remaining 10% was used as a testing set. 10-fold cross validation ensured best utilisation of the limited data. The total accuracy for each cross validation iteration was the average over the 10 cross-validation iterations. Both LDA and QDA classification were evaluated using this procedure.

TABLE IV. CONFUSION MATRIX RESULT WHEN USING THE QDA CLASSIFIER WITH ALL THE FEATURES USED FOR CLASSIFICATION.

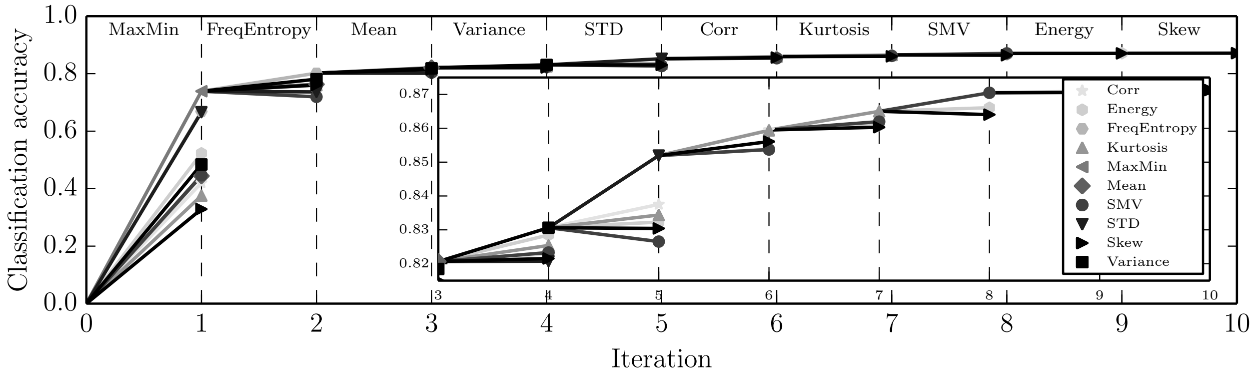
Observed behaviour	Predicted behaviour Lying	Standing	Walking	Running	Grazing	Total
Lying	**898**	26	0	0	36	960
Standing	10	**914**	34	0	2	960
Walking	0	18	**900**	12	30	960
Running	0	0	4	**946**	0	950
Grazing	270	7	45	0	**638**	960
Total	1178	965	983	958	706	4790

III. RESULTS AND DISCUSSION

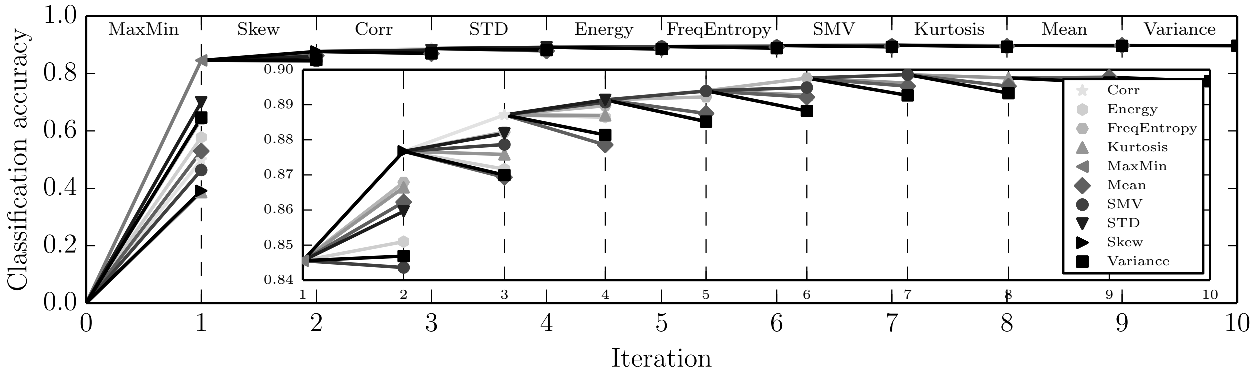
The LDA and QDA classifiers were both tested using the greedy feature selection procedure described in Section II-F. The results are presented in Figure 4.

Figure 4 shows the winning feature at each iteration of the greedy feature selection procedure. The accuracy of the LDA classifier was lower than the QDA classifier at each iteration of the procedure. From Figure 4 it can be concluded that the most important feature in classifying the sheep's behaviour was the "MaxMin" feature. By just using the first feature the LDA and QDA classifiers achieved a classification accuracy of 73.8% and 84.5%, respectively. On the other hand, the LDA classifier achieved a final accuracy of 87.1% when all of the features were used for classification, compared to an accuracy of 89.7% for the QDA classifier. Even though the LDA and QDA classifiers only had a 2.6% accuracy difference when all of the features were used, the QDA classifier showed larger improvements when using a smaller number of features.

The confusion matrix in Table IV was obtained using the QDA classifier and all the features. Most of the behaviours could be identified with high accuracy except grazing, which was misclassified as lying in 33.5% of cases. This outcome can be understood by viewing the raw accelerometer segments associated with lying and grazing, illustrated in Figure 5. The y-axis indicates the sheep's head position. A typical grazing segment consisted of the sheep moving his head down to the ground to graze then either standing still or walking. Finally, the sheep lifts his head up again. A lying segment usually corresponded to the sheep lying on its stomach, with its head

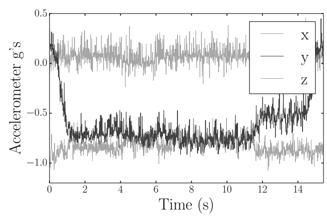


(a)

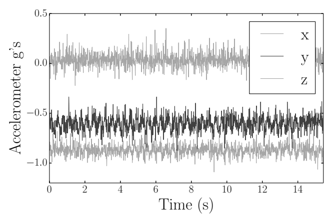


(b)

Fig. 4. The accuracy of the LDA (a) and QDA (b) classifiers at each iteration of the greedy feature selection procedure. The winning features at each iteration are indicated at the top of each plot. The inset shows a magnified view of the top segment of the plot.



(a)



(b)

Fig. 5. Raw accelerometer segments displaying the similarities in the y-axis (head position) between grazing and lying. (a) Grazing segment. (b) Lying segment.

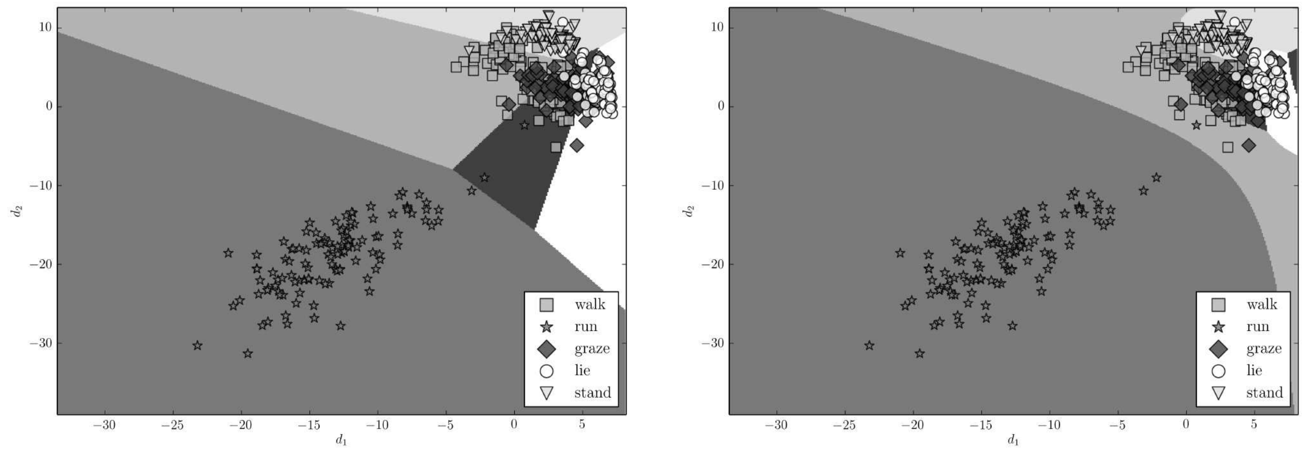
down on the ground. Since the accelerometer data is limited to the placement around the neck, it is difficult to distinguish between these two behaviours. Standing, walking and running were correctly classified in 95.2%, 93.7% and 99.5% of cases respectively. The high classification accuracy for "running" is important if our aim is to determine when animals are being pursued, as may be the case for stock theft.

For further analysis, LDA was used to reduce the feature space to the two most important dimensions. These are shown in Figure 6. The figure illustrates the difference between the LDA and QDA classifier's decision boundaries. The quadratic boundary is more flexible compared to a straight line leading to a higher classification accuracy. Clusters of data points corresponding to each behaviour can also be observed. Again, the ambiguity between lying and grazing is indicated by the overlapping clusters. Even though walking and standing are also closely grouped, it was still possible to separate them with a high degree of accuracy.

Figure 7 shows the normalized confusion matrices for the QDA classification. The first matrix indicates only the first three features selected by the greedy feature selection procedure. The second matrix shows all ten features for classification. A classification accuracy of 88.7% was achieved by using only the first three features, while a 89.7% accuracy was achieved when all the features were used. Hence, by retaining only the first three features, less than 1% in classification accuracy is sacrificed.

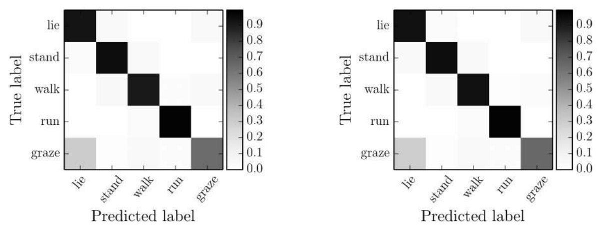
A. *Future work*

The hardware will be further developed in order to improve practicality of the tag. This includes using radio communication in order to retrieve the stored data on the tag without disturbing the animal. Future work is needed in order to increase the overall classification accuracy. The hardware could be improved in order to increase the quality of the accelerometer data. Smarter positioning of the accelerometer should also be attempted to remove or reduce ambiguity between behaviours such as grazing and lying. Additionally, different



(a) (b)

Fig. 6. (a) Linear decision boundaries. (b) Quadratic decision boundaries.



(a) (b)

Fig. 7. Confusion matrices of the QDA classifier (a) trained with only the first three features selected by the greedy feature test and (b) trained with using all 10 features.

and additional types of sensors could be considered to further remove ambiguity.

Classification success can also be influenced by different signal processing parameters, for example the window length as shown in [11]. Other machine learning classifiers should be investigated. Success has for example been demonstrated when using support vector machines [4] and k-nearest neighbour classifiers [3].

IV. Conclusion

The accelerometer was successfully used to classify between five behaviours in sheep. The quadratic discriminant analysis classifier proved to be superior compared to the linear discriminant analysis classifier. The quadratic discriminant analysis classifier also proved highly successful even when only a small number of features was used for classification. The feasibility study showed positive results when using specially-built hardware and a basic classification technique.

References

[1] E. L. Shepard, R. P. Wilson, F. Quintana, A. G. Laich, N. Liebsch, D. A. Albareda, L. G. Halsey, A. Gleiss, D. T. Morgan, A. E. Myers *et al.*, "Identification of animal movement patterns using tri-axial accelerometry," *Endangered Species Research*, vol. 10, no. 2.1, 2008.

[2] M. Moreau, S. Siebert, A. Buerkert, and E. Schlecht, "Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour," *Applied animal behaviour science*, vol. 119, no. 3, pp. 158–170, 2009.

[3] O. R. Bidder, H. A. Campbell, A. Gómez-Laich, P. Urgé, J. Walker, Y. Cai, L. Gao, F. Quintana, and R. P. Wilson, "Love thy neighbour: Automatic animal behavioural classification of acceleration data using the k-nearest neighbour algorithm," *PloS one*, vol. 9, no. 2, p. e88609, 2014.

[4] P. Martiskainen, M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, "Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines," *Applied Animal Behaviour Science*, vol. 119, no. 1, pp. 32–38, 2009.

[5] S. Grünewälder, F. Broekhuis, D. W. Macdonald, A. M. Wilson, J. W. McNutt, J. Shawe-Taylor, and S. Hailes, "Movement activity based classification of animal behaviour with an application to data from cheetah (acinonyx jubatus)," *PloS one*, vol. 7, no. 11, p. e49120, 2012.

[6] R. Nathan, O. Spiegel, S. Fortmann-Roe, R. Harel, M. Wikelski, and W. M. Getz, "Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures," *The Journal of experimental biology*, vol. 215, no. 6, pp. 986–996, 2012.

[7] B. Scholkopft and K.-R. Mullert, "Fisher discriminant analysis with kernels," *Neural networks for signal processing IX*, 1999.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer, 2006.

[9] T. Li, S. Zhu, and M. Ogihara, "Using discriminant analysis for multi-class classification: an experimental investigation," *Knowledge and information systems*, vol. 10, no. 4, pp. 453–472, 2006.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] B. Robert, B. White, D. Renter, and R. Larson, "Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle," *Computers and Electronics in Agriculture*, vol. 67, no. 1, pp. 80–84, 2009.

[12] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive computing*. Springer, 2004, pp. 1–17.