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 an animal's behaviour is usually a difficult and tedious job and can prove challenging when working in inaccessible environments. This problem can be solved by using animal attached accelerometers and automatic classifiers in order to determine an animal's behaviour. This study considers the feasibility of using specially designed hardware to capture three-dimensional accelerometer data from sheep and subsequently automatically classify their behaviour. Five common behaviours have been identified, namely: 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 algorithm 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. In the case of the QDA classifier, grazing was misclassified in 33.5% of cases 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 an animals behaviour provides us with a better understanding and insight to their health, which is a priority in conservation attempts. Monitoring animals in a supervised environment is relatively simple when compared with unsupervised animals in the wild. Unsupervised animals could be monitored by attaching sensors to the animal which will provide information on their behaviour and state. Depending on the type of the sensors, anything from the animals energy expenditure, location, speed, heart rate, temperature and behaviour could be monitored. This provides researchers with information on the 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 three axis provides a more complete picture of animal movement patterns and could reveal smaller movements in an axis that would have otherwise been missed.

A few studies have investigated the possibility of classifying animal's behaviour with the aid of accelerometers, using various machine learning techniques. Goat grazing behaviour has been automatically classified with high success rate using tri-axial accelerometers [2]. Moving averages was used in conjunction with selected threshold values to distinguish between resting, eating and walking. A recent study extended this word by using a basic machine learning algorithm, namely the k-nearest neighbour classifier [3]. It found that the classification accuracy was comparable to other complex techniques. The recognition of cow behaviour was also investigated with the use of a three-dimensional accelerometer and support vector machines [4]. This study attempted to distinguish between eight behaviours and achieved an overall precision of 78%. Another study used a combination of support vector machine and hidden-Markov models to classify the behaviour of cheetahs. The study was successful in classifying five minute segments between three behaviours (feeding, mobile and stationary) [5]. Finally, vultures' behaviours were classified comparing classification and regression trees, random forest methods, support vector machines and artificial neural networks with linear discriminant analysis (LDA) [6]. The LDA classifier was used as a baseline for comparisons and 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 sheep's behaviour using LDA, as well as quadratic discriminant analysis (QDA). The classifiers was used to distinguish between five common behaviours in the sheep, namely: lying, standing, walking, running and grazing.

The general idea behind LDA is to project high dimensional data down to a lower dimension in order to determine the best decision boundary that would separate the data [7]. This is accomplished by projecting the data to a lower dimension with a transformation that would maximise the between-class means and minimize the within-class variance.

II. Material and methods

A. Site and animals

The data collection was done 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 of testing. The sheep were held in a camp large enough to allow for natural movement throughout the day.



(a) (b)

Fig. 1. (a) A photograph showing how the tag was attached around the sheep's neck. (b) A sketch of the accelerometer axis 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 or fast pace.
Running	The sheep running at a fast pace.
Grazing	The sheep is grazing with its head down while standing still or walking slowly.

B. Hardware

The hardware (referred to as a tag) consisted of a 3-axis accelerometer, SD card, micro controller and battery. The tag was designed to allow the accelerometer to constantly sample at 100Hz and store it on the SD card. The tag was enclosed in a durable casing and fitted around the necks of the livestock using collars as seen in Figure 1a. The hardware was placed in approximately the same orientation on the livestock.

Figure 1b shows the orientation of the accelerometer's axis 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 separated days. Usually the days the sheep were collected in the mornings and the tags were fitted around their necks. The sheep were lead to the camp left undisturbed for the day. The hardware logged the acceleration and vibration data of the sheep at a constant sample frequency of $100Hz$ throughout the day. Five common behaviours (Tabel I) were chosen namely: lying, standing, walking, running and grazing. The sheep were monitored from a distance on the three testing days where 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 unprovoked.

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 row in a file is a data point and the following five columns were of interest: runID of the current logging session, timestamps in milliseconds and the three accelerometer axes.

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	0:54:54	6:08:15	1:11:30	6:27:17	1d 7:59:58

Each data file was manually segmented and labelled with the corresponding behaviour by reconciling the logged data with the manual recordings. A distribution of data collected of each behaviour and sheep is displayed in Table II with the corresponding totals. The sheep's id consists of the testing session when the data was logged and the tag's number. "Lying" behaviour was not observed for the sheep on the first testing session.

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. One assumption that should be noted is that all the classes shares the same covariance matrices when using LDA. LDA can be expanded into quadratic discriminant analysis (QDA) if a covariance matrix for each class is calculated. This will change the decision boundary between classes from linear to quadratic [8].

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 [7]:

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

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:

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

The criterion $J(\mathbf{w})$ can be maximized by differentiating with respect to $\mathbf{w}$ to find:

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

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, only the direction of $\mathbf{w}$ is important so $(\mathbf{w}^T\mathbf{S}_B\mathbf{w})$ and $(\mathbf{w}^T\mathbf{S}_W\mathbf{w})$ scalars can be dropped. By multiplying with $\mathbf{S}_W^{-1}$, Equation 4 reduces to the transformation vector:

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

Fisher's linear discriminant tries to project D-dimensional data to a lower dimension in the direction which will maximise the projected 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 for the multi-class $(K > 2)$ problem. 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 [7]:

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

and

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

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 has to be maximized in order to find the decision boundary. This can be obtained by solving the generalised eigenvalue problem [9]:

$$\mathbf{S}_B \mathbf{W} = \lambda \mathbf{S}_W \mathbf{W} \quad (8)$$

E. Data preprocessing and feature extraction

The accelerometer data was preprocessed with python's Scipy library (version $0.14.0$) 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 [10] that a 3-5 second epoch would be optimal. The three acceleration axis's columns were then low-pass filtered at $10Hz$ with an 8th-order Butterworth filter before windowing (Hamming window with a 50% overlap). The features extracted were based on those recommend in [4] and [11] and included the following 10 features that consisted of 31 variables:

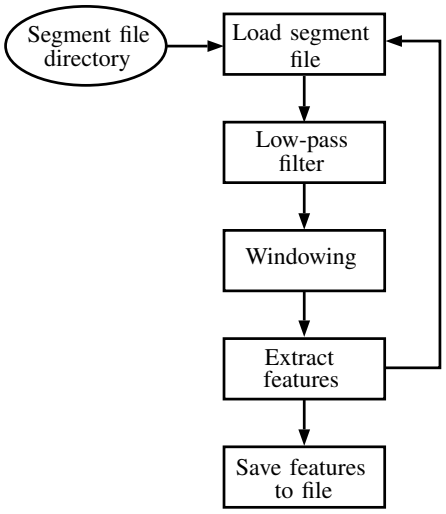


Fig. 2. Flow chart of the feature extraction process.

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

The energy was calculated as the sum of the squared discrete FFT of the window with the DC component excluded. The sum was then normalised by the window length. The frequency-domain entropy was calculated by normalising the discrete FFT and then using those coefficients to determine the frequency-domain entropy of the signal. The frequency entropy has been found useful to discriminate between behaviours with similar energy values [11].

A breakdown of the usable frames after preprocessing for each sheep and class can be viewed in Table III

F. Classification

The python LDA en QDA implementation in Scikit Learn's library (version 0.14.1) were used to determine the best features to use for sheep behaviour classification.

A greedy feature selection algorithm, as illustrated in Figure 3 was used to determine the most importation combination of features. For each iteration of the algorithm, a balanced subset (an equal number of frames for each class) of the usable frames was selected for classification. The algorithm

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

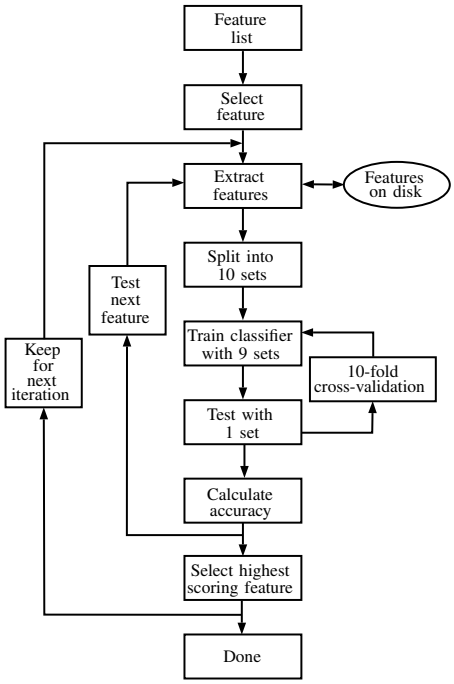


Fig. 3. Flow chart of the greedy feature selection.

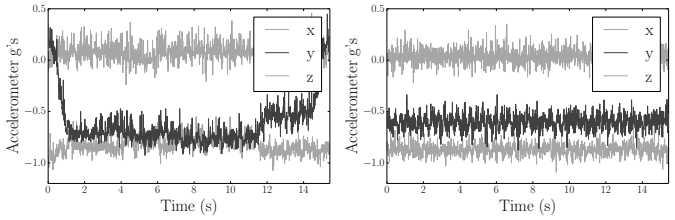
tested each feature's classification accuracy and then kept the feature which scored the highest. The accuracy was defined as the overall correctly classified points (true positive and true negative) proportioned to the total points [3]:

$$ACC = (TP + TN)/(P + N) \tag{9}$$

The algorithm iterated until all the features were used in classification. 80% of the balanced subset was used for training the classification algorithm and the remaining 20% was used as a testing set. Cross validation was achieved by using the 10-fold cross validation technique to make sure all the data has been included in a training or test set. The total accuracy for each cross validation iteration was the averaged and the feature with the highest accuracy was kept for the next iteration. Both LDA and QDA classification were used for this algorithm.

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

Observed	Predicted behaviour					Total
behaviour	Lying	Standing	Walking	Running	Grazing	
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



(a) (b)

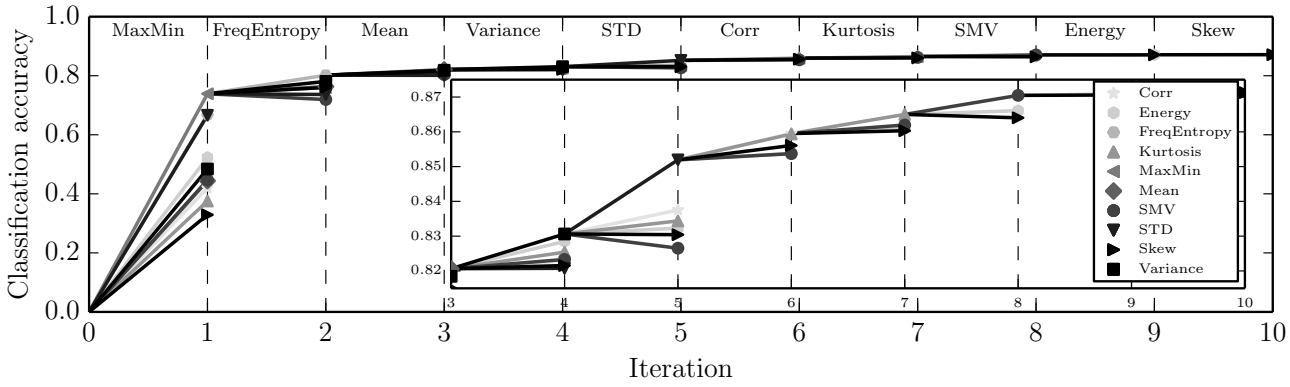
Fig. 5. Raw accelerometer segments displaying the similarities in the y-axis (head position). (a) Grazing segment. (b) Lying segment.

III. RESULTS AND DISCUSSION

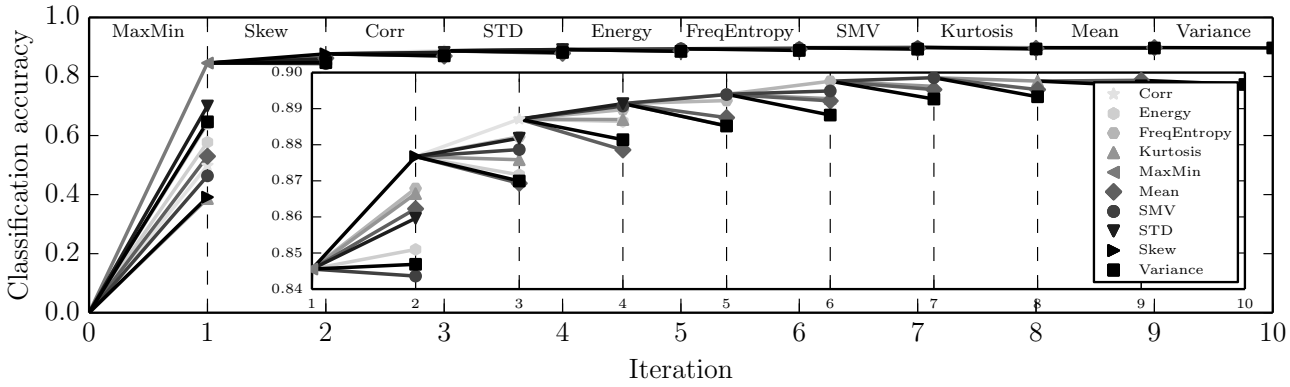
The LDA and QDA classifiers were both tested using the greedy feature selection algorithm described in Section II-F. The results are presented in Figure 4.

The graph in Figure 4 shows the winning feature at each iteration of the greedy feature selection algorithm. The accuracy of the LDA classifier was lower than the QDA classifier at each iteration of the algorithm. From Figure 4 it can be concluded that the most important feature in classifying the sheep's behaviour was the maximum and minimum 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 shows the results of the QDA classifier when using all the features. Most of the behaviours could be identified with high accuracy except grazing, which was misclassified in 33.5% of cases because it was confused with lying. This outcome can be explained by viewing the raw accelerometer segments associated with lying and grazing, illustrated in Figure 5. The y-axis (highlighted in Figure 5) indicates the sheep's head position (up or down). A typical grazing segment consisted of the sheep moving his head down the the ground to graze then either standing still or walking. Finally, the sheep lift his head up again. A lying



(a)



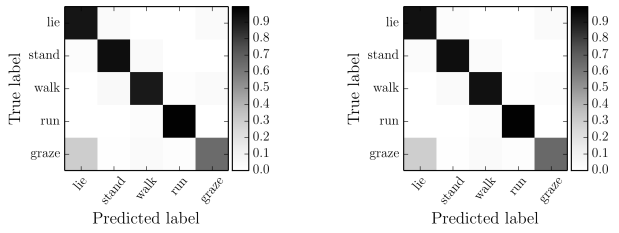
(b)

Fig. 4. The accuracy of the LDA (a, top) and QDA (b, bottom) classifiers at each iteration of the greedy feature selection algorithm. 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.

segment usually occurs when the sheep is lying on its stomach, with its head 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 we want to know when animals are being perused, 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 showing in Figure 6. The figure illustrates the difference between the LDA and QDA classifier's decision boundaries. The quadratic boundary between the behaviours allows for a more flexible decision boundary 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 algorithm. The second matrix shows all ten features for classification. A classification accuracy of 88.7% was achieved by only using the first three features, while a 89.7% accuracy when all the features were used. Hence, by only using the first three features, less than 1% in accuracy is sacrificed.



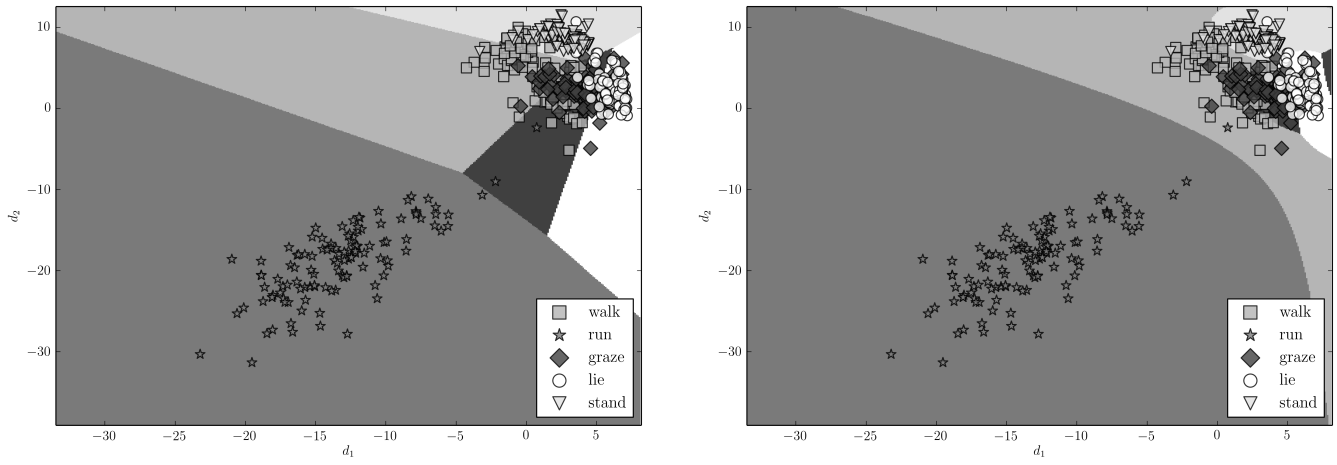
(a) (b)

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

A. Future work

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 taken in consideration to remove or reduce ambiguity between behaviours the system is trying to classify. Additionally, different type of sensors could be investigated to further remove ambiguity.

Classification success can also be influenced by different signal processing parameters, for example the window length



(a) (b)

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

as shown in [10]. 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 few features was used for classification. The feasibility study showed positive results by using the custom hardware and a basic classification technique to classify sheep's behaviour. However, there is definitely room for improvement by refining the hardware and investigating alternative classification techniques.

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

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

[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] 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.

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