Estimating Sheep Pain Level Using Facial Action Unit Detection

Yiting Lu, Marwa Mahmoud and Peter Robinson
Computer Laboratory, University of Cambridge, Cambridge, UK

***Abstract*— Assessing pain levels in animals is a crucial, but time-consuming process in maintaining their welfare. Facial expressions in sheep are an efficient and reliable indicator of pain levels. In this paper, we have extended techniques for recognising human facial expressions to encompass facial action units in sheep, which can then facilitate automatic estimation of pain levels. Our multi-level approach starts with detection of sheep faces, localisation of facial landmarks, normalisation and then extraction of facial features. These are described using Histogram of Oriented Gradients, and then classified using Support Vector Machines. Our experiments show an overall accuracy of 67% on sheep Action Units classification. We argue that with more data, our approach on automated pain level assessment can be generalised to other animals.**

I. INTRODUCTION

Pain level assessment is critical to the welfare of sheep. Severe pain in sheep often indicates diseases, such as footrot [16] and mastitis [17]. Recognising and quantifying pain are essential to the subsequent treatment and pain alleviation [18]. Moreover, efficient and reliable pain assessment tools would help with early diagnoses.

Facial expressions are often used as an indicator of pain level in animals [2], [15]. The Sheep Pain Facial Expression Scale (SPFES) [1] has recently been introduced. It is a standardised measure to assess pain level based on facial expressions of sheep, and has been shown to recognise pain in sheep faces with high accuracy. However, training of scorers and the scoring process can be time-consuming, and individual bias may lead to inconsistent scores [1].

In this paper, we have used computer vision techniques to automate the analysis of facial expressions in sheep. Our approach can improve efficiency and ensure consistency in estimation of pain. We have deployed techniques that are widely used in human emotion recognition to address the problem of automatically assessing pain in sheep.

The overall pipeline of our sheep pain level estimation system is shown in Fig. 1. The main contributions of this paper can be summarised as follows:

1) Introducing a preliminary taxonomy for sheep facial Action Unit (AUs) based on the SPFES.
2) Presenting an automatic multi-level approach for estimating pain level in sheep by extending computer vision techniques that have been widely used in human emotion recognition.
3) Demonstrating that our approach can successfully classify 9 facial action units of sheep and can automatically estimate pain levels. We also show that our approach is generalisable across different dataset of sheep faces. Finally, we argue that - with their pain scales calibrated - the proposed automatic pain level estimation approach can be generallised to other animals, such as mice [12] [5], rabbits [14] and horses [13].

978-1-5090-4023-0/17/$31.00 ©2017 IEEE

We start by reviewing the related work in Section 2. A description of our dataset is discussed in Section 3. Our methodology is described in section 4 followed by the experimental evaluation in Section 5. Finally, conclusions and future work are presented in Section 6.

II. RELATED WORK

Analysing facial expressions of animals was first introduced by Langford *et al.* [4] to facilitate detection of pain level in mice. This approach has been advanced and generalised to many other animals. Yet, manual scoring is the usual practice and automatic assessment of pain level is still an underdeveloped area.

Recently, a standardised sheep facial expression pain scale SPFES was developed by McLennan *et al.* [1]. They showed that their approach is able to recognise sheep pain face with high degree of accuracy. Since manual labelling was used, they found that for different scorers, the accuracy of the pain assessment ranged from 60% to 75%. Their work is the basis of our sheep AU taxonomy.

Sotocinal *et al.* [5] attempt to automate animal pain assessment. They introduced a partially automated approach for pain level assessment on rats. A Haar feature cascade classifier is used for real time eye and ear detection. The classifier served as a pre-screening tool so that only frames detected with the key features are kept as candidates for manual assessment. They found such partially automated pain recognition largely solves the labour-intensive problem of manual scoring.

Yang *et al.* [6] analysed sheep faces and proposed a novel approach to localise sparsely distributed facial landmarks, which uses triplet-interpolated feature (TIF) extraction scheme under the cascade pose regression (CPR) framework [7]. They applied the TIF model on sheep, and reported good results regardless of sheep breed, head pose, partial occlusion, etc. Yet, their work assumed sheep face bounding boxes are known. In our work, we implement sheep face detection before applying the TIF model, then we use the localised sheep facial landmarks for later AU detection.

III. DATA

Unlike human AU analysis, facial expression recognition of sheep is still an underdeveloped area. Very few datasets are available on sheep and fewer include ground truth labels

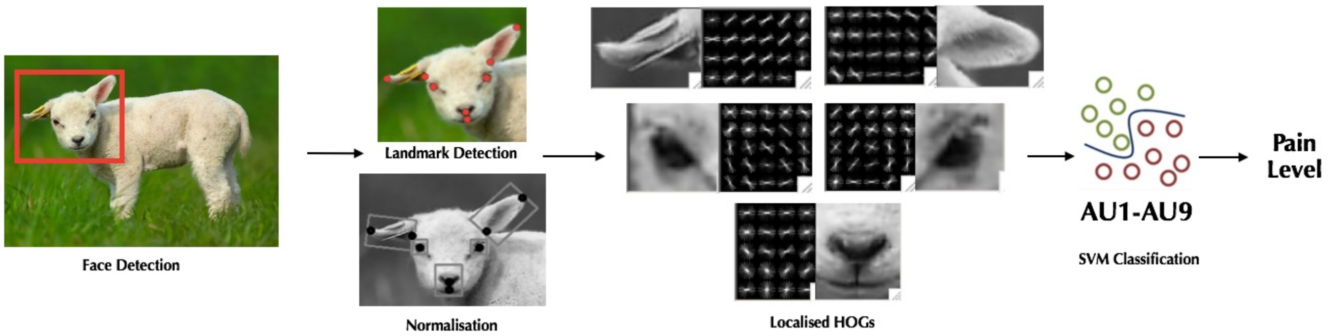

Fig. 1. The pipeline of our automatic approach to estimate pain level in sheep

of facial expressions. In this section, we describe our main dataset and discuss the sheep facial AU taxonomy that is used in our experimental evaluation.

Fig. 2. Sheep facial AU taxonomy with their description & sample. The taxonomy is based on the SPFES [1]

A. Dataset

We've used the same dataset which has been described by Yang *et al.*[6]. This dataset consists of a total of 480 images containing sheep faces. The face bounding boxes are given, yet there are no labels for sheep facial expressions. Therefore, we labelled the facial expressions. The labelling criteria is discussed in the next section.

For the sake of our work, we divide the dataset into two subsets:

1) The sheep from farm (SFF) dataset: this subset includes 380 photos taken from a farm. This set includes sheep of different breed, skin colour and head pose. The photos vary in lighting conditions with their background being either barn or fenced grassland. The image resolution is consistent throughout.
2) The sheep from the Internet (SFI) dataset: this subset contains 100 images collected from the Internet. This set is more diverse than SFF. Sheep of different breed, skin colour and head pose are included. The lighting condition, background and image resolution are all different from one another.

B. AU Taxonomy and Labelling

Facial Action Units (AUs) has been widely used in human facial expression analysis [19], [20]. Human AUs have been indexed in the Facial Action Coding System (FACS) [3], which forms the standard for automatic analysis of human facial expression and emotion recognition. In contrast, sheep facial expression is yet to be catagorised. We first discuss the sheep AU taxonomy, then present our labelling approach of SFF and SFI datasets accordingly.

The sheep facial AU taxonomy used in our work is based on the SPFES [1]. As a preliminary AU taxonomy, only frontal faces are considered. The key features that considered are ears, eyes and nose. Although cheek and lip profile are discussed in the SPFES, they are omitted in our work because those features can hardly be seen on a frontal face. The main differences between the SPFES and our AU taxonomy are illustrated as follows:

- **Ears**: In the SPFES, three pain levels are defined regarding the extent of the ear rotation with both profile and frontal faces taken into account. In our work, we map the three pain levels but only consider the frontal faces.
- **Nose**: In the SPFES, three pain levels are defined according to the nostril shape. In our work, we map the three pain levels as they are.
- **Eyes**: In the SPFES, three pain levels are defined in terms of the eye narrowing extent. In our work, we define only two pain levels, namely pain and no-pain, because the dataset is strongly biased towards the no pain case. We also define a separate class for non-classifiable pain level (AU8) when not enough infor-

mation can be drawn from the frontal face due to head pose deviation.

Fig. 2 shows the detailed description of our taxonomy. Based on these rules, we labelled SFF and SFI datasets with AU numbers. The mapping between AU numbers and feature-wise pain scores was developed based on the SPFES. Each frontal face is labelled with five features, namely left ear, right ear, left eye, right eye and nose. Although the SPFES scores for symmetric features are expected to be the same, our facial AU label might differ due to poor lighting, partial occlusion or head pose deviation. The overall pain rating was calculated from the feature-wise pain scores using the same rule-based approach adopted by experts.

IV. METHODOLOGY

In our work, we propose a full pipeline for automatic detection of pain level in sheep. We first present face detection and facial landmark localisation. We then extract appearance descriptors from the normalised facial features, followed by the AU classification. The overall pain level is estimated based on the classification results of facial features. This pain assessment approach is not specific to sheep and can be generalised to other animals if the proper taxonomies are developed.

A. Face Detection

We use the Viola-Jones object detection framework [8] to implement the frontal face detection. SSF dataset is used to provide the ground truth. Due to the small number of ground truth images, we adopt a boosting procedure to achieve larger number of training samples. Sheep faces are clipped from the ground truth images with ears excluded, then rotations and intensity deviation are applied to each sheep face. Finally, the processed sheep faces are put on top of some random background images. A fixed window size of 32×24 is used for positive samples. The final collection of positive images consists of 5000 image windows boosted from 250 ground truth images. The face detector gives an averaged accuracy of 71% using a 10-fold cross validation approach.

B. Facial Landmark Detection

The Cascade Pose Regression (CPR) [7] scheme is used for the facial landmark localisation. Due to the sparsely-distributed nature of sheep facial landmarks, the TIF [6] approach is adopted in our work. Compared with Robust Cascade Pose Regression (RCPR) [9], which accesses the features on the line segments between two landmarks by linear interpolation, the TIF model is able to draw feature from larger area. The shape indexed feature location is defined as:

$$\mathrm{p}(S,i,j,k,\alpha,\beta) = y_i + (\alpha \cdot \overrightarrow{v}_{i,j} + \beta \cdot \overrightarrow{v}_{i,k})$$

where S is the current shape, i, j, k are landmark indices and α, β are randomly generated constants. With $\overrightarrow{v}_{i,j}$ denoting the direction from landmark y_i to y_j ($\overrightarrow{v}_{i,k}$ from y_i to y_k), it can be shown that any feature is accessible on the area spanned by these two vectors. Such approach is robust against large head pose deviation and sparsely distributed facial landmarks, which suits well with sheep facial landmark localisation problem.

The localised sheep facial landmarks are: both ear tips (p1, p6), both ear roots (p2, p5), both eyes(p3, p4), the crossing of the nostrils (p7) and the mouth (p8). See Fig. 3 for an illustration.

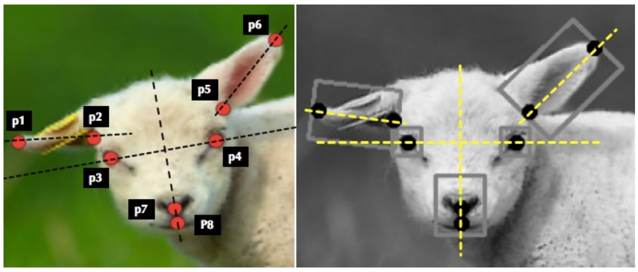

Fig. 3. Left: Localised facial landmarks (Note: the eight facial landmarks are labelled from p1 to p8) Right: Normalised sheep face marked with feature bounding boxes

C. Feature-wise Normalisation

Normalisation is commonly used in human face recognition [22] to ensure faces taken from various view points are registered [21] and comparable. In our work, feature-wise normalisation is applied on sheep face. Ears, eyes and nose are extracted and normalised separately.

Eye normalisation is achieved by rotating the image to keep the two eyes (p3, p4) aligned horizontally. The nostril crossing and the mouth (p7, p8) are then automatically aligned vertically since they are inherently in right angle to the eye alignment regardless of the head pose. The scaling factor for both eyes and nose is defined as the interpupilary distance. The feature bounding boxes (see Fig. 3) can then be drawn according to their dominant directions. The optimal box size is determined by optimising the AU classification accuracy. The bounding box sizes we have used are listed as follows: Eyes - 50×50 pixels; Nose - 100×80 pixels (all are $rows \times cols$) with 172 pixel interpupilary distance.

Unlike human ears, sheep ears vary greatly in size depending on their breed, and are able to show large rotations regardless of the head pose. The dominant direction of each ear is defined as the alignment of the ear tip and the ear root (p1 with p2 and p5 with p6). The scaling factor for each ear is the distance between the paired-up tip and root. The normalised bounding box size for Ears is 56×80 pixels.

D. Feature Descriptor

Histogram of Oriented Gradients (HOG) [23] has been widely used as an appearance feature descriptor for human facial expressions. We make use of HOG to describe sheep facial features. We used Dlib [11] implementation of HOGs. As proposed by Felzenswalb et al.[10], each block of HOG stands for a 31 dimensional vector: 4 normalisation masks are applied on top of the 9-orientational histogram, followed by PCA dimensional reduction [10]. Each HOG descriptor spans $(total\ number\ of\ blocks) \times 31$ dimensions. In Fig. 1, HOG

TABLE I

CLASSIFICATION ACCURACY OF OUR 3-CLASS AU CLASSIFIERS COMPARED TO MAJORITY VOTE CLASSIFIER. WE COMPARE SVM LINEAR, REF KERNEL AND SIGMOID FUNCTION. AS SHOWS, LNR OUTPERFORMS RBF AND SIG FOR MOST AU'S. LNR ALSO HAS THE HIGHEST OVERALL DETECTION RATE. [TRAINED ON SFF, TESTED ON SFF]

Feature	Ear (Left)			Ear (Right)			Nose			Eye (Left)			Eye (Right)			
AU Number	AU1	AU2	AU3	AU1	AU2	AU3	AU4	AU5	AU6	AU7	AU8	AU9	AU7	AU8	AU9	Mean
Sample size	210	**80**	**40**	200	**80**	**50**	100	160	**70**	230	90	**10**	220	100	**10**	-
Majority Vote	0.64	0.24	0.12	0.61	0.24	0.15	0.30	0.48	0.21	0.70	0.27	0.03	0.67	0.30	0.03	0.33
LNR SVM	0.80	**0.61**	0.83	0.85	**0.65**	0.72	**0.64**	0.49	**0.63**	0.72	**0.82**	**0.50**	0.77	**0.88**	**0.20**	**0.67**
RBF SVM	**0.96**	0.60	0.80	0.94	0.58	0.76	0.58	0.71	0.59	**0.91**	0.68	0.10	**0.93**	0.85	0.00	0.66
SIG SVM	**0.96**	0.55	**0.88**	**0.97**	0.35	**0.82**	0.47	0.64	0.36	0.85	0.60	0.30	0.82	0.60	0.10	0.62

descriptors are visualised, showing the block dimensions for ear, eye and nose. It can be seen that HOGs are able to depict the shape and texture of each feature.

E. Pain level estimation

With HOGs extracted and AUs labelled, we use Support Vector Machine (SVM) [24] to train separate classifiers for each facial feature. The overall pain level estimation approach can be described as follows: we first map the predicted AUs to feature-wise pain scores. Then we average the scores for symmetric features (i.e.:eyes, ears) and average all three feature-wise scores (ear, eye, nose) to get the overall pain score. Finally, we define two thresholds (0.4, 0.8) to generate the overall pain score.

TABLE II

COMPARISON BETWEEN THE CLASSIFICATION ACCURACIES OF OUR 2-CLASS & 3-CLASS CLASSIFIERS FOR ACTION UNITS AU1 & AU3 [TRAINED ON SFF, TESTED ON SFF]

AU Number	AU1(L)	AU1(R)	AU3(L)	AU3(R)
3-class	0.80	0.85	0.83	0.72
2-class(relabelling)	0.83	0.83	0.87	0.84
2-class(exclusion)	0.84	0.86	0.98	0.98

V. EXPERIMENTAL EVALUATION

In this section, we evaluate our approach presented in the previous section. We compare 3-class and 2-class AU classification approaches. We also discuss the effect of data rebalancing as well as the generalisability of our AU classifiers.

A. AU classification results

We first evaluated our AU detection approach using a 3-class classifiers for each feature. The SFF dataset was used for both training and testing. Each face is given five labels (left ear, right ear, left eye, right eye and nose), and each label is associated with 3 AUs (Ears-AU1,2,3; Nose-AU4,5,6; Eyes-AU7,8,9). Altogether 15 SVM classifiers were trained for all five features using linear kernel (LNR), radio basis function (RBF) and sigmoid function (SIG). A 10-fold cross validation approach was used in all of our experiments.

Table. I shows the evaluation results, with the distribution of the ground truth and the corresponding majority vote classifier accuracies. The accuracy is defined as true positives divided by the total number of samples. With most AUs achieving more than 60% detection rate, our experimental evaluation confirms that the presented AU taxonomy is reasonable and that our proposed AU detection approach are able to classify AUs of sheep.

It can be seen that SVM with LNR outperforms RBF and SIG for most AUs (as highlighted in Table. I). Moreover, the overall accuracy of the LNR model is the highest among the three - achieving a 67% detection rate in average. We therefore use LNR SVM model for the rest of our experiments due to its good performance as well as high computation speed.

Among all three features, ear appears to be the strongest pain level indicator. Our approach achieved high accuracy on ear action units: AU1 (SPFES: no pain) and AU3 (SPFES: great pain). This is expected as AU1&AU3 classes can be unambiguously differentiated.

B. Confusion reduction

As seen in our classification results, our challenge is to map evolutionary features into a fixed number of AU classes.

Sheep facial expression changes gradually as their pain deteriorate. In manual scoring, the decision boundaries are inherently soft due to human nature and can easily be re-calibrated to fit in more pain levels (tighter decision boundary) or less pain levels (wider decision boundary) assuming the human scorer knows the trend of the evolution. However, in computer vision, such sense of trend is missing when those evolutionary features are simply split into different classes and used in a one-vs-all training approach. In this case, the number of classes, the taxonomy and the labelling of the training samples become crucial.

In this section, we attempt to reduce the confusion by reducing the number of AUs. As a sample feature, we focus on ear-related AUs because the intermediate state (AU2: 61%) shows obvious confusion compared with the AU1(82%) and AU3 (78%).

1) AU reduction by relabelling: Training samples labelled as AU2 are relabelled and split into AU1 and AU3. The rationale here is that: the facial symptom indicating pain would progressively become more obvious as the pain level deteriorates. Since there is no solid boundary between two consecutive pain levels, by splitting up AU2 (SPFES: slight pain) into AU1 (SPFES: no pain) and AU3 (SPFES: great pain), we are simply recalibrating the pain scale by making

TABLE III

COMPARISON BETWEEN THE CLASSIFICATION ACCURACIES OF OUR 3-CLASS LINEAR SVM CLASSIFIERS BEFORE AND AFTER DATA REBALANCING [TRAINED ON SFF, TESTED ON SFF]

Feature	Ear (Left)			Ear (Right)			Nose		
AU Number	AU1	AU2	AU3	AU1	AU2	AU3	AU4	AU5	AU6
Sample size before rebalancing	210	80	**40**	200	80	**50**	100	160	**70**
Majority Vote Accuracy	0.64	0.24	0.12	0.61	0.24	0.15	0.30	0.48	0.21
Accuracy	0.80	0.61	0.83	0.85	0.65	0.72	0.64	0.49	0.63
Sample size after rebalancing	40	40	40	50	50	50	70	70	70
Accuracy	0.85	0.53	0.73	0.84	0.60	0.76	0.66	0.51	**0.74**

TABLE IV

CROSS-DATASET TESTING, SHOWING THE CLASSIFICATION ACCURACY OF OUR 3-CLASS AU CLASSIFIERS. WE CAN SEE THAT OUR APPROACH IS GENERALISABLE ACROSS DIFFERENT DATASETS [TRAINED ON SFF, TESTED ON SFI]

Feature	Ear (Left)			Ear (Right)			Nose			Eye (Left)			Eye (Right)			
AU Number	AU1	AU2	AU3	AU1	AU2	AU3	AU4	AU5	AU6	AU7	AU8	AU9	AU7	AU8	AU9	Mean
Sample size	96	8	13	102	7	8	24	77	16	80	33	4	91	20	6	-
SVM LNR	0.65	0.63	0.62	0.77	0.43	0.63	0.54	0.65	0.31	0.60	0.39	0.00	0.37	0.10	0.67	0.49

each pain level cover a wider range of facial expressions. We then train a binary classifier for each ear. The resulting accuracy (see Table.II) exceeds our 3-class approach accuracy by 6% in average.

2) AU reduction by exclusion: In this section, we change the way of AU reduction: we exclude the confusing AU and the associated samples from the classification stage. AU2 training samples are excluded. Using this approach, we managed to get a 15% increase in detection rate (see Table.II). Such increase indicates that by excluding the confusing intermediate class, a more reliable classifier can be trained.

The results are reasonable since we are mapping between a continuous scale of feature changes to a set of discrete AU's.

C. Training sample rebalancing

Some AUs perform worse than the others, such as AU2,5&9. We have further explored our data to check if the exceptionally low accuracies are resulted from the imbalance in training samples. In this experiment, we enforce training sample rebalancing and investigate its effect on accuracy. Eyes are not examined because there are only 10 samples labelled as AU9 in SFF dataset. Three 3-class linear kernel SVM classifiers are trained.

By reducing the samples for AU4&5, the detection rate of AU6 improves by 17% (see Table.IV) and the detection rates of AU4&5 increase by about 4%. Note that the accuracy of AU5 is the lowest among AU4,5,6 despise having large number of samples, while among AU1,2,3, AU3 has the highest accuracy even with the smallest number of samples. The accuracy of AU2 is about 30% lower than AU1&3 in both the imbalanced and balanced cases. These results suggest that data rebalancing would, to some extent, improve the accuracy of the AU with the lowest majority vote accuracy. However, the nature of the AU definition still takes the leading role in affecting its detection rate.

D. Generalisation

Here we aimed to evaluate the generalisability of our approach. In this experiment, we test if a classifier trained on a specific dataset is generalisable to another dataset. Five 3-class classifiers are trained using the SFF dataset, and then tested on the SFI dataset. The SFI dataset varies a lot in resolution and is strongly unbalanced and biased towards AU1,5&7. This makes the cross-dataset testing a challenging task.

The test results are shown in Table. IV. The detection rates of AU1,2&3 show strong correlation with their test results on the SFF set, while the accuracies of AU4,5,6,7,8,9 is affected by the data distribution. Such result suggests that ear is a strong pain indicator and its classifier generalises well, whereas for noses and eyes, more data is needed in order to achieve better classification results.

The overall pain level estimation is also tested on the SFI and the confusion matrix is shown in Table. V. We adopt the same rule-based method as used in manual scoring: we use five classifiers to predict five feature-wise pain scores, then those scores are averaged and compared with two thresholds separating the three pain levels. No extra error is introduced during the estimation stage, therefore the overall pain level is expected to be a fair measure of our overall performance. An obvious trend favouring higher pain levels can be seen from the confusion matrix. It suggests that our automated pain level estimation approach is able to detect, though exaggerate, the pain level based on the five features. The inherited softness in human decision boundary is expected to be achieved by adjusting the two thresholds of the three pain levels.

E. Concatenated feature

In this experiment, we attempt to test an alternative approach to detect pain level in sheep. Instead of training

TABLE V

CROSS-DATASET TESTING, SHOWING THE CONFUSION MATRIX OF THE ESTIMATED OVERALL PAIN LEVEL [TRAINED ON SFF, TESTED ON SFI]

Truth / Label	P=0	P=1	P=2	Sample size	Accuracy
P=0	35	27	18	80	0.44
P=1	3	12	11	26	0.46
P=2	1	1	9	11	0.82

TABLE VI

CONFUSION MATRICES OF THE PAIN LEVEL ESTIMATION (M.V.: MAJORITY VOTE ACCURACY) [TRAINED ON SFF, TESTED ON SFF & SFI]

Truth / Label	P=0	P=1	P=2	Total	M.V.	Accuracy
Concatenated feature - Trained on SFF, Tested on SFF						
P=0	94	19	27	140	0.41	0.67
P=1	30	67	53	150	0.44	0.45
P=2	3	7	40	50	0.15	0.80
Concatenated feature - Trained on SFF, Tested on SFI						
P=0	43	43	5	80	-	0.54
P=1	17	9	0	26	-	0.35
P=2	5	4	2	11	-	0.18

five separate AU classifiers, we train a single pain level classifier. We concatenate all five feature descriptors to be a 3568 dimensional whole face descriptor and label the training samples with the overall pain levels. A 3-class pain level classifier is trained with linear SVM model on the SFF dataset. The classifier is tested on both SFF and SFI dataset. The confusion matrices are presented in Table. VI. In the generic dataset (SFF) test, The pain level classifier shows high accuracy on $'Pain = 0'$ & $'Pain = 2'$ classes regardless of the small number of samples on $'Pain = 2'$ class. Yet, in the cross-dataset (SFI) test, the detection rate leans to the majority vote accuracy. Larger amount of balanced dataset is required for further exploration of this method.

VI. CONCLUSIONS

In this paper, we present a multi-level approach to automatically estimate pain levels in sheep. We propose a preliminary sheep facial AU taxonomy based on the SPFES. We automate the assessment of facial expressions in sheep by adapting the techniques for human emotion recognition. We demonstrate that our approach can successfully detect facial AUs and assess pain levels of sheep. Our experiments also show that our AU classifiers are generalisable across different datasets.

For future work, we would like to further explore classifier training with the concatenated feature descriptor to map facial feature directly to pain levels. We would also like to add geometry features as well as appearance features. This will help our AU classifier to be more robust to head pose deviation as well as breed variation. Larger number of labelled data is needed to further investigate data balancing and generalisation. Ultimately, we would like to test our automatic pain assessment approach on different animals. However, this will again require more efforts in data collection and labelling.

VII. ACKNOWLEDGMENTS

Dr. Krista McLennan (University of Chester) facilitated the data collection and advised the SPFES. The authors would also like to thank the help of the Department of Veterinary medicine at University of Cambridge.

REFERENCES

[1] McLennan, K.M., et al., Development of a facial expression scale using footrot and mastitis as models of pain in sheep. Appl. Anim. Behav. Sci, 2016.

[2] Leach, M.C., Klaus, K., Miller, A.L., Scotto di Perrotolo, M., Sotocinal, S.G., Flecknell, P.A., The assessment of post-vasectomy pain in mice using behaviour and the mouse grimace scale. PLoS One 7, 2012.

[3] P. Ekman and W. V. Friesen. Manual for the Facial Action Coding System. Palo Alto: Consulting Psychologists Press, 1977.

[4] Langford, D.J., Bailey, A.L., Chanda, M.L., Clarke, S.E., Drummond, T.E., Echols, S., Glick, S., Ingrao, J., Klassen-Ross, T., Lacroix-Fralish, M.L., Matsumiya, L., Sorge, R.E., Sotocinal, S.G., Tabaka, J.M., Wong, D., van den Maagdenberg, A.M., Ferrari, M.D., Craig, K.D. and Mogil, J.S. Coding of facial expressions of pain in the laboratory mouse. Nature Methods, 2010.

[5] Sotocinal et al.: The Rat Grimace Scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions. Molecular Pain, 2011.

[6] H. Yang, R. Zhang and P. Robinson. Human and Sheep Facial Landmarks Localisation by Triplet Interpolated Features. 2015.

[7] P. Dollar, P. Welinder, and P. Perona. Cascaded pose regression. In CVPR, 2010.

[8] P.A. Viola, M.J. Jones. Rapid object detection using a boosted cascade of simple features. In CVPR, issue 1, 2001.

[9] X.P. Burgos-Artizzu, P. Perona and P. Dollar. Robust face landmark estimation under occlusion. In ICCV, 2013.

[10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object Detection with Discriminative Trained Part Based Models. IEEE TPAMI, 2010.

[11] D. E. King. Dlib-ml: A machine learning toolkit. JMLR, 2009.

[12] Matsumiya L. C. et al. Using the Mouse Grimace Scale to reevaluate the efficacy of postoperative analgesics in laboratory mice. J Am Assoc Lab Anim Sci 51, 2012.

[13] Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E, et al. Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration. PLoS ONE, 2014.

[14] Keating SCJ, Thomas AA, Flecknell PA, Leach MC. Evaluation of EMLA Cream for Preventing Pain during Tattooing of Rabbits: Changes in Physiological, Behavioural and Facial Expression Responses. PLoS ONE 7, 2012.

[15] Van Rysewyk, S. "Nonverbal indicators of pain". Animal Sentience: An Interdisciplinary Journal on Animal Feeling. 2016.

[16] Dolan S, Kelly JG, Monteiro AM, Nolan AM. Up-regulation of metabotropic glutamate receptor subtypes3 and 5 in spinal cord in a clinical model of persistent inflammation and hyperalgesia. Pain, 2003.

[17] Dolan, S., Field, L. C., and Nolan, A. M. The role of nitric oxideand prostaglandin signalling pathway is spinal nociceptive pro-cessing in chronic inflammation. Pain, 2000.

[18] Flecknell P. Analgesia from a veterinary perspective. Br J Anaesth, 2008.

[19] Freitas-Magalhes, A. Microexpression and macroexpression. In V. S. Ramachandran (Ed.), Encyclopedia of Human Behavior, 2012.

[20] Reed, L. I.; Sayette, M. A.; Cohn, J. F. "Impact of depression on response to comedy: A dynamic facial coding analysis". Journal of Abnormal Psychology, 2007.

[21] Lisa Gottesfeld Brown, A survey of image registration techniques (abstract), ACM Computing Surveys archive, 1992.

[22] R. Brunelli, Template Matching Techniques in Computer Vision: Theory and Practice, 2009.

[23] Dalal, N., Triggs, B., & Schmid, C. Human detection using oriented histograms of flow and appearance. In European conference on computer vision, 2006.

[24] Cortes, C.; Vapnik, V. "Support-vector networks". Machine Learning, 1995.