POSTPRINT VERSION. The final version is published here :
Islam, M. R., Massicotte, D., Nougarou, F. et Zhu, W. P. (2018, 24-27 June 2018). HOG and Pairwise SVMs for Neuromuscular Activity Recognition Using Instantaneous HD-sEMG Images. Communication présentée 2018 16th IEEE International New Circuits and Systems Conference (NEWCAS), Montreal, Canada. doi: 10.1109/NEWCAS.2018.8585731

HOG and Pairwise SVMs for Neuromuscular Activity Recognition Using Instantaneous HD-sEMG Images

Md. Rabiul Islam[1], Daniel Massicotte[1], Francois Nougarou[1] and Wei-Ping Zhu[2]

[1] Dept. of Electrical and Computer Engineering, Université du Québec à Trois-Rivières, QC, Canada
[2] Dept. of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada

Abstract- **The concept of neuromuscular activity recognition using instantaneous high-density surface electromyography (HD-sEMG) image opens up new avenues for the development of more fluid and natural muscle-computer interfaces. The state-of-the-art methods for instantaneous HD-sEMG image recognition achieve prominent performance using a computationally intensive deep convolutional networks (ConvNet) classifier, while very low performance is reported using the conventional classifiers. However, the conventional classifiers such as Support Vector Machines (SVM) can surpass ConvNet at producing optimal classification if well-behaved feature vectors are provided. This paper studies the question of extracting distinctive feature sets, thus propose to use Histograms of Oriented Gradient (HOG) as unique features for robust neuromuscular activity recognition, adopting pairwise SVMs as the classification scheme. The experimental results proved that the HOG represents unique features inside the instantaneous HD-sEMG image and fine-tuning the hyper-parameter of the pairwise SVMs, the recognition accuracy comparable to the more complex state of the art methods can be achieved.**

Index Terms— Neuromuscular activity recognition, HOG, HD-sEMG, Gesture recognition, SVM, Muscle-computer interface

1. INTRODUCTION

The precise characterization and recognition of neuromus cular activities present a great challenge [1] The high-density sEMG (HD-sEMG) based methods have been proposed in the recent years [2][3]. The HD-sEMG records myoelectric signals using two-dimensional (2D) electrode arrays that characterize the spatial distribution of myoelectric activity over the muscles that reside within the electrode pick-up area [4]. The collected HD-sEMG data are spatially correlated which enabled both temporal and spatial changes and robust against malfunction of the channels with respect to the previous counterparts [3]. However, the existing HD-sEMG based neuromuscular activity recognition methods are still depending on the windowed sEMG which demands to find an optimal window length otherwise influence the classification accuracy. To overcome this problem and developa more fluid and natural muscle-computer interface, more recently, W. Geng et al. [4], explored the patterns inside the instantaneous sEMG images spatially composed from HD-sEMG enables neuromuscular based gesture recognition solely with the sEMG signals recorded at a specific instant. In their approach, the instantaneous values of HD-sEMG signals at each sampling instant were arranged in a 2D grid in accordance with the electrode positioning. Afterwards, this 2D grid was converted to a grayscale sEMG image. A computational model based on deep convolutional neural networks (ConvNet) [5] has been employed for sEMG image classification. However, the potential drawback is the classification method based on ConvNet, is computationally very expensive to be practical for real-world applications for neuromuscular activity recognition. Moreover, the studies conducted in [4] reported of attaining recognition rate as low as 20% using the conventional classifiers such as support vector machines (SVM). However, the conventional classifiers such as SVM can surpass ConvNet at producing optimal classification if well-behaved feature vectors are provided [6]. However, this aspect is totally overlooked in [4]. Therefore, developing computationally efficient distinctive feature extraction and classification algorithms for instantaneous sEMG image based neuromuscular activity recognition is highly demanded.

For instantaneous sEMG image based neuromuscular activity recognition, the challenge remains open because very limited research has been done on it. This paper studies the histogram of oriented gradients (HOG) for the improved characterization of the instantaneous sEMG image. HOG is one of the state-of-the-art methods for object recognition [7]-[9]. However, this important characterization method is ignored for sEMG signal classification. This paper proposed to use a HOG based feature extraction method for instantaneous sEMG image classification. According to our best knowledge, no one performed similar studies before for sEMG signal classification.

The rest of the paper is organized as follows. Section 2 provides the computational details of the proposed feature extraction method. Section 3 describes the testing database and the experimental validation. Section 4 offers some conclusive remarks.

2. THE PROPOSED NEUROMUSCULAR FEATURE EXTRACTION AND CLASSIFICATION ALGORITHM

The proposed neuromuscular feature extraction and classification algorithm has three computational components: (i) preprocessing and sEMG image generation, (ii) feature extraction, and (iii) classification. A schematic diagram of the proposed muscular activity recognition method by instantaneous sEMG images are shown in Fig. 1. The sketches of hand and gestures in Fig.1 are adapted from [4].

First, the acquired HD-sEMG signals at each sampling instant were arranged in a 2-D grid according to their electrode positioning. This grid was further transformed into an instantaneous sEMG image by linearly transforming the values of sEMG signals from mV to color intensity as $[-2.5mV,\ 2.5mV]$ to $[0\ 255]$. Thus, an instantaneous grayscale sEMG image was formed with the size of 16×8. The gradient image $\nabla f(x, y)$ is obtained by convolving an estimation filters over x and y axis of the instantaneous sEMG image $f(x, y)$. The magnitude $|\nabla f(x, y)|$ and orientation $\theta(x, y)$ for each pixel of the sEMG image are computed from the gradient image $\nabla f(x, y)$. The sEMG image is divided into a dense grid with a spatial $\eta \times \eta$ pixels cells. For each cell, a local 1-D histogram of gradient over all pixels in the cell are computed as features. This aggregated cell-level 1-D histogram builds the HOG feature vector for the unique representation of the instantaneous sEMG image. Finally, these HOG feature vectors are fed to a computationally effective learned pairwise SVM classifier for instantaneous gesture recognition.

Section 2.1 presents the HOG feature extraction technique for sEMG image representation and Section 2.2 presents the classification schemes respectively.

2.1. Histogram of Oriented Gradients (HOG) Feature Extraction

After generating the instantaneous sEMG image by linearly transforming the values of sEMG signals to color intensity as mentioned above, the crucial task is to extract distinctive features to represent the instantaneous sEMG image for robust classification of the performed hand gesture. However, the main research question is what makes the different gestures distinctive performed by the same or different subjects? For example, the hand gestures explained in Section 3 and shown in Table I can be differentiated by their shape and orientation features. The color might not be a reliable feature because the portrayed hand gestures have the same color. Therefore, any method that can precisely describe the shape and orientation information will solve the problem. Nevertheless, the problem in our hand is even more challenging because the instantaneous sEMG image is formed by linearly transforming the values of sEMG signals from mV to color intensity which reflects the intensity distributions of the performed hand gestures. The different hand gestures produce different spatial intensity distributions, thus also make the structure of the instantaneous sEMG image different. These discriminative attributes have been capitalized and used as features in this work.

Both intuitive observation and preliminary experimental results indicate that the gradient of the intensity distributions or edge directions provides the discriminative features for instantaneous sEMG image classification. HOG precisely captures this notion. Therefore, we propose to use HOG as features for instantaneous sEMG image classification. HOG features are calculated by taking orientation histograms of intensity distributions from all locations of a dense grid on a sEMG image region and combined features are used for classification. HOG features are assumed to be designed for

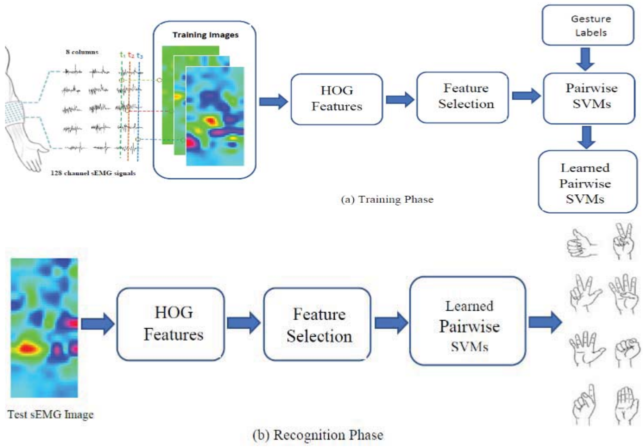

Fig. 1. Schematic illustration of the proposed muscular activity recognition by instantaneous sEMG images.

imitating the visual information processing of the brain and have robustness against local changes of position. This important property of HOG can be exploited to cope with the electrode shifting problem encountered between two different HD-sEMG recording sessions. HOG is like scale-invariant feature transform [11] in the sense that a local region is described by deriving gradient orientations from the orientation histogram.

Consider the gradient estimation filters $h_x = [-1, 0, 1]$, and $h_y = [-1, 0, 1]^T$. The gradient information of an instantaneous sEMG image can be obtained by

$$\nabla f(x, y) = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}\right]^T = \begin{bmatrix} f(x, y) * h_x \\ f(x, y) * h_y \end{bmatrix} \qquad (1)$$

where, $*$ denotes an operation of a 1-dimensional (1-D) convolution. The x and y stand for height and width of the instantaneous sEMG image. The magnitude of a pixel is calculated by

$$|\nabla f(x, y)| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \qquad (2)$$

and the orientation of a pixel is calculated by

$$\theta(x, y) = \tan^{-1}\left(\frac{\partial f}{\partial x} \Big/ \frac{\partial f}{\partial y}\right) \qquad (3)$$

These magnitude $|\nabla f(x, y)|$ and orientation $\theta(x, y)$ at each pixel are then used for calculating HOG.

The main intuition behind HOG feature extraction is that, while individual $|\nabla f(x, y)|$ and $\theta(x, y)$ are highly variable and subject to significant variations across nearby (x, y) locations, even for the sEMG images generated by the same hand gesture, the cumulative statistics of the spatial distribution of the gradient orientation and magnitudes over small region of the sEMG images derived from the same gesture provide quite robust descriptors of the instantaneous sEMG image.

To compute orientation histograms, the obtained instantaneous sEMG image gradient is divided into $8 \times 4 = 32$ non-overlapping rectangular cells, and each cell is of size $\eta \times \eta$ pixels ($\eta = 2$). Four $\eta \times \eta$ neighboring cells form a block of size $\varsigma \times \varsigma$ ($\varsigma = 2$). A schematic diagram of HOG extraction process is illustrated in Fig. 2. There are total $v\varsigma \times$

$h\varsigma = 21$, overlapping blocks are formed over an instantaneous sEMG image (where $v\varsigma = 7$ and $h\varsigma = 3$, denotes the number of vertical and horizontal block respectively). In each $\eta \times \eta$ cell, the orientation histogram has β bins ($\beta = 7$), which correspond to orientations $i \times \pi/\beta$, where $i = 0,1,\ldots,\beta$. Thus, each of the block contains $\varsigma \times \varsigma \times \beta = 28$ dimensional HOG feature vectors and each instantaneous sEMG image contains $v\varsigma \times h\varsigma \times (\varsigma \times \varsigma \times \beta) = 588$ dimensional HOG feature vectors.

This 588-dimensional HOG feature vector is used to represent the instantaneous sEMG image. It is noteworthy that η, ς and β are parameters and selecting values of these parameter tradeoff with the overall instantaneous sEMG image classification performance. Therefore, it is significant to select the optimum values of these parameters for extracting most discriminant HOG features.

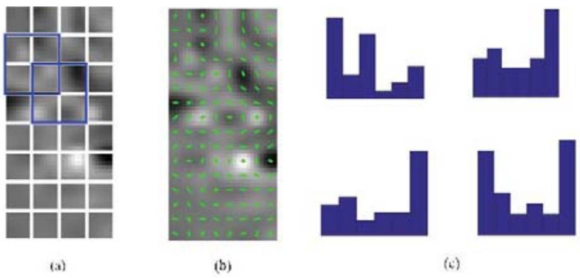

Fig. 2. HOG extraction process (a) An instantaneous sEMG image is partitioned by non-overlapping cells and overlapping blocks (each block has (2×2) four cells). (b) Gradients information are overlaid over an instantaneous sEMG image (c) HOG in each block. The horizontal axis represents angle information and the vertical axis bears weighted histogram.

Now, we calculate the 28-D HOG feature vector from a block of $\varsigma \times \varsigma$ cells. Consider $|\nabla f(x,y)|$ and $\theta(x,y)$ in one block as shown in Fig. 2(a) and 2(b). In Fig. 2(b), the orientation of the arrow represents $\theta(x,y)$ and the length of the arrow stands for $|\nabla f(x,y)|$. In the experiments, the gradient orientation is transformed from $-\pi \leq \theta \leq \pi$ to $0 \leq \theta \leq \pi$ and then evenly quantized into β bins. The HOG feature vector $h_1 \in \mathbb{R}^{\beta}$ of the first cell (top left in Fig. 2(a)) can be calculated by voting

$$h_1(i) \leftarrow h_1(i) + |\nabla f\theta_i(x,y)|, \ \ i = 1,\ldots,\beta \tag{4}$$

where $|\nabla f\theta_i(x,y)|$ indicates the magnitude from the gradient and θ_i is the quantized orientation. In the same way as h_1, the three-feature vectors (h_2, h_3 and h_4) can be generated from three other cells of a same block. By combining these feature vectors, the HOG feature vectors of a block turn into $h = [h_1^T, h_2^T, h_3^T, h_4^T]^T \in \mathbb{R}^{\beta \times 4}$.

It is to be noted that the equation (4) is a simplified form. However, in our implementation, the trilinear interpolation is used to calculate the HOG features [12]. The trilinear interpolation smoothly distributes the gradient to $\varsigma \times \varsigma$ cells of a block to reduce the aliasing effect caused by the pixels near to the cell boundaries. This technique can also be robust against small distortions between sEMG images derived from the same gesture.

Moreover, the gradient strengths vary over an instantaneous sEMG image owing to local variations. Therefore, the overlapped blocks on sEMG image are normalized individually so that each scaler cell-response contributes several components to final HOG feature vector. The normalization is performed by

$$h = h/\sqrt{||h||_2^2 + \epsilon^2} \tag{5}$$

where, ϵ is a small normalization constant used to avoid divided by zero [12]. This normalized HOG representation is used for instantaneous sEMG image classification.

2.2. Pairwise SVM Classifier

After the HOG feature extraction for representing an instantaneous sEMG image, the most important task is to employ a computationally effective classifier which has the high generalization ability for solving a multi-class classification problem. SVM [13][15] is essentially a binary classifier, however, multi-class classification problem is solved by training several binary SVM classifiers and an optimal global decision function is obtained by fusing the outputs of each of these binary classifiers. In addition, the decision function of SVM's is fully determined by the number of support vectors (SVs) which is substantially lower than the actual number of samples used in training, makes SVM computationally very efficient. Moreover, SVM trained on HOG features has become a popular method for across many visual perception tasks due to the performance and robust theory [14]. Why do SVM's trained on HOG features perform so well is still an open research issue in the literature. However, it is pointed out in [14] that preserving second-order statistics and locality of interactions are fundamental to achieve good performance. All these motivated us to use and train pairwise SVM's classifiers on HOG features extracted from the instantaneous sEMG image.

3. EXPERIMENTS

We tested our feature characterization method on CapgMyo data sets [10] (this database is made available from following website http://zju-capg.org/myo/data/index.html). This dataset was developed for providing a standard benchmark database (DB) to explore new possibilities for studying next-generation muscle-computer interfaces (MCIs). Table I illustrates gesture in DB-a and DB-b. The CapgMyo database comprises 3 sub-databases (referred as DB-a, DB-b and DB-c). However, as followed by the [4], DB-a has been used in our preliminary experiments to evaluate the performance of our proposed methods. In DB-a, 8 isotonic and isometric hand gestures were obtained from 18 of the 23 subjects and each gesture was also recorded for 10 times. For each subject, the recorded HD-sEMG data is filtered, sampled and the instantaneous sEMG image is generated using the method mentioned in Section 3. More explicitly, 8 different hand gestures are performed by every subject and each hand gestures are recorded for 10 times with a 1000 Hz sampling rate, which in total generates $(8 \times 10 \times 1000 = 80000)$ instantaneous sEMG images. Then, our HOG-based proposed feature extraction technique elaborated in Section 2.1 is applied to each of the instantaneous sEMG images. Thus, an

Table I. Gestures in DB-a and DB-b (8 isotonic and isometric hand configurations) [10]

Label	Description	Gesture	Label	Description	Gesture
1	Thumb up		5	Abduction of all fingers	
2	Extension of index and middle flexion of others		6	Finger flexed together in fist	
3	Flexion of ring and little finger, extension of the others		7	Pointing index	
4	Thumb opposing base of little finger		8	Adduction of extended fingers	

$80000 \times M$ dimension HOG feature vectors are obtained. The each of the HOG feature vectors dimension M depend on the different HOG parameters such as η, ς and β. However, considering the low resolution instantaneous sEMG image and based on our preliminary experiments, we select $\eta = 2$, $\varsigma = 2$ and $\beta = 7$ respectively. Hence, we obtained $v\varsigma \times h\varsigma \times (\varsigma \times \varsigma \times \beta) = 588$ dimension HOG feature vectors of an instantaneous sEMG image.

Now, for every subject in DB-a, a pairwise SVMs classifier is trained to predict the desired hand gestures for each incoming sEMG images. The pairwise SVMs framework is based on LIBSVM, a library for support vector machines [16]. To conduct the above-mentioned gesture classification task, the obtained $80000 \times M$ dimension HOG feature vectors are randomly divided into three subsets such as training, validation and testing set. In this preliminary investigation, 50% of the HOG feature vectors from the entire feature set are randomly selected and used as a training set. In the same way, the remaining 50% of the HOG feature vectors are divided into validation and testing set. The validation set is used for model/kernel and parameter selection for pairwise SVMs. Due to computationally effective and reducing searching space for parameter selection, the RBF kernel $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \gamma > 0$ is used to train the obtained HOG feature set. There are two parameters for an RBF kernel which is a cost parameter (C) and kernel parameter γ. It is not known in advance which C and γ are the best for a given problem. Therefore, the parameter selection is performed. We used a grid search along with this υ-fold ($\upsilon = 3$) cross-validation scheme to find the optimum (C, γ) on the validation set. It is recommended in [17] to use the exponentially growing sequences of C and γ to identify the good parameters. Hence, we use $C = [2^5, 2^4, 2^3, \ldots, 2^{-1}]$ and $\gamma = [2^{-6}, 2^{-5}, 2^{-4}, \ldots, 2^2]$. Therefore, we examined with $7 \times 9 = 63$ combinations of (C, γ) pairs. Then, the whole training feature set is trained using the pair of (C, γ) that achieves the best cross-validation accuracy. Finally, this trained classifier is used to predict the test feature set.

Confusion matrix generated from the predicted classification results were used as a performance indicator. The correctly classified (%) gesture classes are listed along the diagonal line of the Confusion matrix as presented in Table II. The average classification accuracy of the proposed methods is 86.63% which is comparable to the state of the art methods. Using instantaneous values of HD-sEMG and SVM classifier, the average classification accuracy as low as 20% was reported in [4]. However, the average classification accuracy increased to 86.63% using proposed HOG and optimized parameter of pairwise SVMs. In addition, the recall or true positive rate (TPR) and the precision or the positive predictive value (PPV) [18] of each gesture classes are also computed and mentioned in Table III. The 86.62% average precision and recall of each class also indicate the potentiality of the HOG and pairwise SVMs for neuromuscular activity recognition. Finally, the experimental results demonstrate that: (i) HOG are effective features for unique representations of instantaneous HD-sEMG images (ii) Provided discriminant features and fine-tuning the hyper-parameter of the conventional classifiers such as pairwise SVMs, the state of the art recognition rate can be achieved for muscular activity recognition based on instantaneous HD-sEMG images.

Table II. Confusion Matrix of the Proposed Neuromuscular Activity Recognition Method.

Ture Class \ Predicted Class	CL01	CL02	CL03	CL04	CL05	CL06	CL07	CL08
CL01	87.35	0.39	1.05	0.04	0.08	7.85	3.24	0.00
CL02	0.40	86.03	4.72	4.44	2.06	0.61	0.12	1.61
CL03	1.38	3.88	89.24	1.74	1.29	0.44	0.89	1.13
CL04	0.04	3.51	1.24	82.84	4.51	0.60	0.40	6.86
CL05	0.04	1.76	1.12	4.31	89.99	0.28	0.08	2.43
CL06	8.77	0.57	0.89	0.53	0.81	83.76	4.14	0.53
CL07	2.27	0.20	1.34	0.20	0.24	4.05	89.02	2.67
CL08	0.04	1.93	1.26	6.52	2.63	0.51	2.32	84.79

Table III. Precision and Recall of every gesture classes.

Class	CL01	CL02	CL03	CL04	CL05	CL06	CL07	CL08
Precision	87.52	87.44	88.38	82.32	88.57	85.07	88.66	85.03
Recall	87.35	86.03	89.24	82.84	89.99	83.76	89.02	84.79

4. CONCLUSIONS

In this paper, we propose to use Histogram of Oriented Gradients (HOGs) as distinctive features and pairwise SVMs for robust neuromuscular activity recognition using instantaneous HD-sEMG images. 80000 instantaneous HD-sEMG image frames for 8 different gesture of each subject from CapgMyo database were examined. The experimental results demonstrate that HOG are effective features for unique representations of instantaneous HD-sEMG images. Also, provided discriminant features and fine-tuning the hyper-parameter of the conventional classifiers such as pairwise SVMs, the state of the art recognition rate can be achieved for neuromuscular activity recognition based on instantaneous HD-sEMG images.

ACKNOWLEDGMENT – This work was supported in part by the regroupement stratégique en microsystèmes du Québec (ReSMiQ) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES

[1] Farina, D. *et al., "*The extraction of neural information from the surface EMG for the control of upper-limb prostheses: merging avenues and challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22**,** 797–809, 2014.

[2] Amma, C., Krings, T., Boer, J. & Schultz, T., "Advancing muscle-computer interfaces with high-density electromyography," *Conference on Human Factors in Computing Systems,* 929–938, ACM, 2015.

[3] Stango, A., Negro, F. & Farina, D., "Spatial correlation of high density EMG signals provides features robust to electrode number and shift in pattern recognition for myocontrol," *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 23, 189–198, 2015.

[4] Geng, W., Du, Y., Jin, W., Wei, W., Hu, Y., Li, J., "Gesture recognition by instantaneous surface EMG images," *Scientific Report.*2016, *6*, 36571.

[5] Krizhevsky, A., Sutskever, I. & Hinton, G. E.,"Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems,* 1097–1105, 2012.

[6] F.J. Huang and Y. LeCun, "Large-scale Learning with SVM and Convolutional Nets for Generic Object Categorization," 2006 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pp. 284-291, 2006.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proc. IEEE Conf. Comput. Vis. Pattern Recog., vol. 2, pp. 886–893, 2005.

[8] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627-1645, Sept. 2010.

[9] Y. Tian, L. Cao, Z. Liu, and Z. Zhang, "Hierarchical filtered motion for action recognition in crowded videos," *IEEE Trans. Syst., Man, Cybern. C*, Appl. Rev., vol. 42, no. 3, pp. 313–323, May 2012.

[10] Du. Y., Jin. W., Wei. W., Hu. Y., Geng. W., "Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation," *Sensors*, 17, 458, 2017.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[12] N. Dalal,"Finding people in images and videos," Ph.D. thesis, INRIA Rhone-Alpes, 2006.

[13] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.

[14] Bristow, Hilton and Lucey, Simon, "Why do linear SVMs trained on HOG features perform so well?" in arXiv preprint arXiv:1406.2419, 2014.

[15] C.W. Hsu and C.J. Lin, "A Comparison of Methods for Multiclass Support Vector Machines," *IEEE Trans. Neural Networks*, vol. 13, pp. 415-425, 2002.

[16] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology, 2:27:1—27:27, 2011.

[17] C. Hsu, C. C. Chang, and C. J. Lin. A practical guide to support vector classification. Technical report, 2005.

[18] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The binormal assumption on precision-recall curves," in Proc. 20th ICPR, Aug. 2010, pp. 4263–4266.