IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Utilization of Support Vector Machine and Histogram of Oriented Gradients in Squat Posture Classification

Sakuntala Dilhani Wipulasundara
Department of Mechanical Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Abstract— **The squat exercise is widely used as a muscle strengthening exercise in regular fitness exercising in the gym and at home on a daily basis. If not performed accurately there can be repercussions of the squat, which can cause physical complications to subjects. This research looks at classifying the bodyweight squat into one correct and three incorrect posture classes. Videos of the four postures have been obtained and image frames have been extracted from them. The Histogram of Oriented Gradients (HOG) descriptor has been applied on each image to extract features. These features and the grey image have both been combined to make the feature set rich. Support Vector Machine (SVM) with the linear kernel, has been utilized for classification and the method has shown a good level of accuracy on our data set and also on the data set used by another research group.**

Keywords—Squat, Posture, HOG, SVM, Linear Kernel

1. INTRODUCTION

There is a variety of exercises carried out by individuals in gymnasiums, in parks and at home. Certain exercises improve the flexibility and coordination of the body in addition to increasing fitness. Therefore, they are also used as rehabilitation exercises for patients recovering from illnesses such as stroke and patients who are suffering from dementia and Alzheimer's disease. The squat exercise is one such example. One main reason why the squat exercise is popular is because it closely resembles activities of daily living (ADL). There are many variations of the squat exercise including the body weight squat, goblet squat, barbell squat, overhead squat and dumbbell squat [1], [2]. The scope of this research is limited to the body weight squat.

However, exercising needs to be carried out with care. If performed in an inaccurate posture on a daily basis, exercising may cause pains and distortions in muscle movement. This will retard the ability to get involved in routine exercise and also require medical attention. As a result, there are set guidelines for participants to follow, when indulging in exercise, in order to ensure the posture is accurate. In most gymnasiums, an instructor guides the participants during the exercise procedure. During exercise in a home setting, or in the absence of an instructor, there needs to be a mechanism to guide participants. Therefore, there have been previous research carried out to detect and correct the posture during exercise [3] . During the squat exercise the main focus is the posture of the hip, the knees and the feet. Previous research has clearly differentiated between inaccurate postures and accurate postures taking the position of the hip, knees and feet into consideration.

Posture related research has been carried out involving ADL, sports and fall detection. Various sensing and classification methods have been used in past research in order to differentiate between postures. Input has been obtained through sensor arrays, balance boards and cameras [4]-[5] . Classification has been performed by feeding data into algorithms such as, support vector machine (SVM), K nearest neighbors (KNN), artificial neural network (ANN) and convolutional neural networks (CNN) [6].

But considering the classification of squat postures there is a requirement for a method that has better accuracy and can be used in routine activity and not in a laboratory setting. The use of expensive equipment will prevent common usage. In order to cater to these requirements, a method to classify the body weight squat has been developed in this research. SVM has been used in this research as it had shown a high accuracy in classification when applied on postures. The input data set was constructed in such a way as to make the feature set better so that classification accuracy increases. The Histogram of Oriented Gradients (HOG) descriptor was utilized in this regard [7], [8].

The main contributions of this method are;

i) Utilization of a commonly available camera to obtain videos

ii) Application of HOG descriptor to extract features and it being used along with the grey image to make the feature set rich

iii) Demonstrating a classification accuracy of 95% on the data set of this research as well as the data set of another research group

This paper is organized as follows. Section 2 explains related work carried out in similar research areas. Section 3 explains the methodology followed by us. Section 4 demonstrates the results obtained in this research. Section 5 is the conclusion and explains future work in this area.

2. Related Work

A guidance system for squat posture has been developed in [4]. Here the Microsoft Kinect camera and the Wii balance board have both been utilized. The shoulder width, ankle width and knee angle were measured using joint coordinates obtained from the Kinect camera. The coordinated of the center of pressure was obtained from the Wii balance board. 11 subjects participated in the study. SVM and Naïve Bayes were the algorithms used. It was concluded that SVM was accurate for large amounts of data, but that Naïve Bayes was fast.

In [9] a two-stage fall recognition algorithm has been proposed based on human posture features. Here, SVM, KNN, decision tree (DT) and random forest (RF) have been utilized for classification. Experiment results have shown that SVM with linear kernel function can distinguish falling actions best. The results demonstrated that their algorithm not only was able to distinguish falls from stable ADL actions but also could separate fall actions and confusing movements like lying and bending over. Compared with other state-of-art fall detection methods, the proposed method had advantage of accuracy.

A system for posture recognition of lying-down human bodies using a low-resolution pressure sensor array was proposed in [10]. A support vector-machine was used to perform the classification of pressure maps. Three databases were constructed in order to represent the pressure maps: pressure raw-data, HOG and SIFT (Scale Invariant Feature Transform) image descriptor vectors. It was found that the classification of human postures has a high accuracy using an SVM classifier with a linear kernel.

In [11] a novel and non-invasive method to automatically detect and classify awkward working postures based on foot plantar pressure distribution data measured by a wearable insole pressure system was proposed. Ten participants performed five different types of awkward working postures (i.e., overhead working, squatting, stooping, semi-squatting, and one-legged kneeling) in a laboratory setting. Four supervised machine learning classifiers (i.e., ANN, decision tree (DT), KNN, and SVM) were used for classification performance using a 0.32s window size. SVM classifier obtained the best results.

A preliminary study of posture recognition system has been developed in [5], to rectify the user’s sitting posture by alerting him/her. In this research, Kinect sensor has been used for the sitting posture detection, and then the postures data has been fed in to the posture recognition models such as SVM and ANN, for training purpose. Out of these two models the SVM with linear kernel had the highest accuracy.

In [6] temporal distance matrices from video data obtained from a single camera are applied for detection of squat posture. The 3D pose data, are utilized based on, is extracted from each frame, taken from monocular camera image. This gives a representation independent from pixel information. The distances between 19 joints are measured and normalized. CNN with 1D convolutions were used for this. Six mistakes in squats have been detected here. Some posture differences such as warped and round back were not correctly differentiated as there was no key point in the middle of the back. Inward bending of knees was sometimes missed as it needed instantaneous detection.

Since SVM has shown better accuracy in posture related research involving classification problems, it has been selected as the classification method for this research. The data set creation method of this research has been modified in order to obtain the best possible accuracy while maintaining simplicity.

3. METHODOLOGY

The videos of the subjects performing the bodyweight squat were used as input as shown in Figure 3.1. The video frames were extracted and converted into the grey images. The HOG descriptor was applied to the frames. This was done in order to make the feature set rich. Both image data were stored, followed by a window size in the form of an array of pixels. The array was reshaped and fed in to the SVM and classified into the four postures of body weight squat identified in this research.

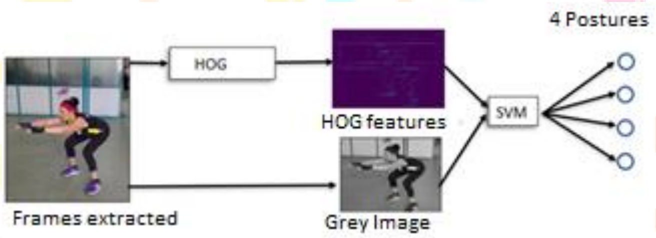

Figure 3.1: Methodology followed in summary

.

3.1 Squat postures and subjects

Squat posture classifications used in this research are explained in Figure 3.2. These have been decided based on previous research. The hip should not be unnecessarily stressed during lowering or rising [12]. Especially during rising, there is a tendency to exert excess force on the hip by over-flexing it. This unnecessary load on the hip, when exerted continuously may result in pain in the hip and spine. The participant will get exhausted and terminate exercising early.

The best posture is to bend the knees so that the thighs are parallel to the ground [13] . The knees should not come beyond the toes when lowering in the body weight squat [12]. The knees should remain at or behind the furthermost point of the toes. The heals should not lift off the ground at any instance of the squat. There is a tendency to raise the heals off the ground when lowering and this is considered a mistake. In such situations the weight acts on the front part of the bent foot and the body may also become imbalanced. This should be avoided.

Sometimes, there is a tendency to squat insufficiently [12]. Insufficient lowering happens here. In such situations, the leg muscles are not sufficiently exercised. This would lessen the effectiveness of the body weight squat. The parallel squat is recommended for athletes over the half squat. Correct and incorrect squat postures have also been identified in the research [6] Four postures were identified in our research. One was the correct posture for body weight squat. The second was flexing the hip, the third lifting the heals and the last was insufficient squat.

The postures explained in Figure 3.2 are namely, the accurate squat posture “CORR”, over flexing the hip, ”HFLEX”, knees coming beyond toes and lifting the heal “HLIFT” and finally, insufficient squat ”ISQU”. Out of these four postures the first was considered to be the accurate posture to engage in the body weight squat, whereas, the latter three postures were considered erroneous.

Figure 3.2. Squat postures explained

12 subjects aged between 22 years and 45 years with bodyweight varying between 62 kg and 90 kg participated in this research. Recording of videos was carried out in two gymnasiums, under the supervision of an instructor each. Permission for the experiments were obtained in writing from the ethics committee of Sri Lanka Institute of Information Technology. Written consent was obtained from each subject prior to the experiment. Since this is a classification problem, and there had to be samples to train the algorithm to classify, equal samples of the identified postures were obtained from each subject. Each sample consisted of a duration of 30 seconds where they performed the exercise and afterwards had a break of 1 to 2 minutes. Similarly, the four samples were recorded from the same angle and distance. A camera with a resolution of 1920x1080 pixels was used at a frame rate of 30 frames per second to record the videos. Samples were recorded under various lighting conditions that prevailed in the gymnasium at the time of the experiment. The participants wore clothing of different colours. Still images extracted from some video samples that were used for the experiment are shown in Figure 3.3.

Figure 3.3: Samples extracted from videos for the four postures

3.2 Feature extraction

The characterization of local object appearance and shape can be done quite well by the distribution of local intensity gradients or edge directions, even without exact knowledge of the relevant gradient or edge positions. HOG feature descriptors used to compute local intensity gradients or edge directions are similar to SIFT features. [7]

In order to extract HOG features the HOG algorithm was applied by following three steps according to [8] Initially, the Sobel filter which computes the gradient magnitude and orientation, was applied, to detect edges. Then the magnitude and orientation of gradients were calculated using equations (1) and (2)

$$m(x,y) = \sqrt{d_x(x,y)^2 + d_y(x,y)^2} \qquad (1)$$

$$\Theta = \arctan\left[\frac{d_y(x,y)}{d_x(x,y)}\right] \qquad (2)$$

Secondly, orientation binning was performed. The image was divided into a number of cells, and for each cell a histogram of gradient directions for the pixels within the cell was compiled. Each pixel casted a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation.
Thirdly, block normalization was carried out, where, the cells were grouped and the local histograms were contrast normalized according to equation (3), where, v is the un-normalized descriptor vector and ϵ is a small normalization constant to avoid division by zero.

$$v \rightarrow \sqrt{\frac{v}{(\|v\|_1 + \epsilon)}} \qquad (3)$$

Using this method, the HOG features were extracted using a python code. In addition to the HOG features the grey image was obtained by converting the colour images in to grey images.

3.3 Data set creation

The data set was created by using the HOG features as well as the grey image. The summary was shown in Figure 3.1. The HOG feature matrix was resized to a 50 x 50 matrix. The plane image was converted from colour to grey scale and resized to 50 x 50. These two planes were maintained as two separate layers and the entire data set consisted of the video being broken into window sizes of 20. When the window size was larger the time taken increased and when the window size was smaller the information was insufficient, the best efficiency and amount of information was obtained with 20. One unit of the data set was created as 50 x 50 x 2 x 20. The data set was created by applying a python code on the pre-recorded set of videos. The data was stored in the form of an array. Each data unit mentioned above had a target, which was the numeric value of its classification category. The data array was reshaped and fed into the SVM along with the target array. The algorithm was trained using 80% of the data and validated using 20% of the data.

3.4 Classification

The SVM algorithm was used for this multiclass classification problem related to posture. Here, the linear kernel was utilized. The linear kernel has been successfully utilized in [9] [10] and [14] The Google Colab environment was used for training and the coding was done using Python. SVM computes the hyper plane by solving the optimization equation (4) according to [10].

$$\min_{w,b,\xi} \quad (1/2)w^T w = C \sum_{i=1}^{l} \xi_i \qquad (4)$$

subject to $y_i(w^T \varphi(xi) + b) \geq 1 - \xi_i$, $\xi_i \geq 0$.

Where, $x_i \in R$ n is an instance of data set (x_i , y_i) with i = 1, . . . , l; yi ∈ {1, −1} is the label of instance x_i , and C > 0 is a penalty parameter to allow some misclassification as extracted from [10].
The Linear Kernel function φ is shown in equation (5)

$$K(x_i , x_j) = x_i^T x_j \qquad (5)$$

Initially, the SVM was trained and then tested on the test sample. The target was predicted and compared with the actual target to obtain the results tables and the confusion matrices. The exercise was performed on videos recorded by us. Then the videos recorded by another group which carried out similar research, was used to further validate our method. These are discussed under results.

4 Results

4.1 Results obtained on our data

The results obtained on our data set is shown in Table 4.1. The confusion matrix is shown in Figure 4.1. The results show that the proposed method has been successful on our data set showing an overall accuracy of 95%. The results in Table 4.1 also demonstrate that there has been no bias towards any particular posture and that all postures have been identified equally well.

Table 4.1 : Results obtained from our data set

	precision	recall	F1-score	support
HLIFT	0.96	0.91	0.94	57
CORR	0.98	0.97	0.98	62
ISQU	0.86	0.94	0.90	34
HFLEX	0.97	0.98	0.98	65
Accuracy			0.95	218
Macro avg	0.95	0.95	0.95	218
Weighted avg	0.96	0.95	0.95	218

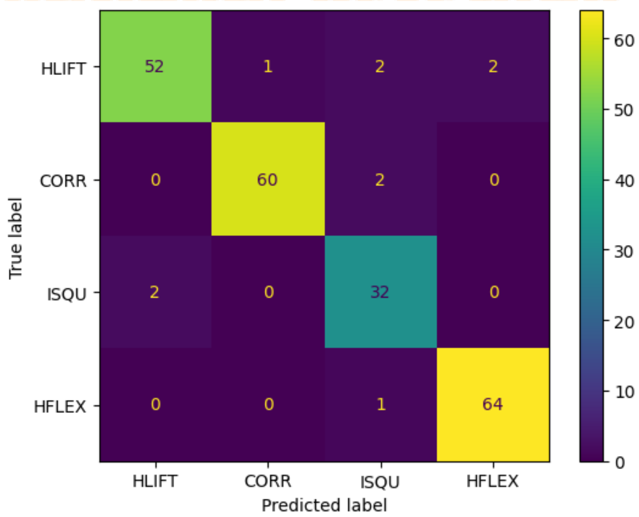

Figure. 4.1. The Confusion matrix of our data set using HOG image and grey image

The data set that was used in the research in [6] was also tried out in our research. There were 76 video samples for each posture category. The four classifications that were applied there were, namely, correct posture “CORR”, over flexing the hip ”HFLEX”, lifting the heal “HLIFT” and insufficient squat ”ISQU”. Although in our original research all participants were videoed from the same angle and same distance, in the research by [6] it was not the case. Participants were videoed from various directions and distances. Therefore, testing our algorithm on their data was an additional opportunity to investigate the suitability of our algorithm for distance and direction variation.

4.2 **Results obtained on the data set of** [6]

The results obtained for their data set by applying the method, are shown in Table 4.2. The confusion matrix is shown in Figure 4.2.

Table 4.2 : Results obtained from the data set of [6]

	precision	recall	F1-score	support
HLIFT	0.92	0.91	0.91	233
CORR	0.96	0.95	0.96	232
HFLEX	0.92	0.96	0.94	204
ISQU	0.98	0.98	0.98	228
Accuracy			0.95	897
Macro avg	0.95	0.95	0.93	897
Weighted avg	0.95	0.95	0.95	897

The results and confusion matrix show that the postures have been classified without any bias towards a particular posture.

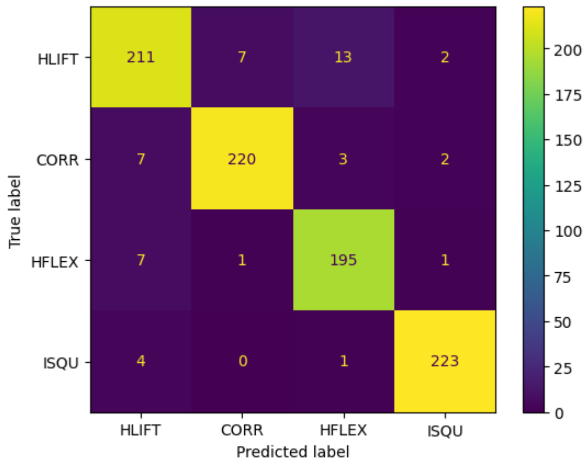

Figure 4.2. Confusion matrix of the data set of [6] using HOG image and grey image

A comparison has been made with the results of other research carried out to classify squat posture.

Table 4.3 shows the results of such research.

Table 4.3: Results of other similar research

Description	Accuracy
Squat posture classification [12] in using deep learning CNN-LSTM method	91.7%
Squat posture classification in [4] using SVM	95.61%
Squat posture classification in [15] using CNN	93.33%
Squat posture classification in [6] using CNN	81.05%

ACKNOWLEDGMENT

The author would like to acknowledge the authors of [6] for sharing their data set when requested. The author would also like to acknowledge the contribution by the staff of Pran gym Udahamulla and Booker's gym Colombo 10. Also, the participation of all subjects in the videos is acknowledged.
.

5 CONCLUSION AND FUTURE WORK

It can be concluded that using the HOG features and grey image in order to make the feature set rich, and training an SVM with a linear Kernel, has been successful in classifying different postures encountered in the body weight squat. Compared to other methods used in past research for squat posture classification, our method has the advantages of maintaining accuracy while utilizing a simple and cost-effective method that can be utilized in routine exercising.

This method has shown positive results when the subjects were photographed from the same direction and also when the subjects were photographed from various directions. The results have shown unbiased classification of all postures with nearly equal accuracy. Therefore, can be utilized in the gym or in a home setting for subjects who indulge in the body weight squat.

Further research needs to be carried out to modify the program in order to develop the interface, so that the subject is given real-time feedback during the exercise in order to rectify mistakes in posture. This will enable the subjects to identify and correct their mistakes during the exercise. This can be used in the absence of an instructor.

REFERENCES

[1] A. C. Fry, J. C. Smith, and B. K. Schilling, "Effect of Knee Position on Hip and Knee Torques During the Barbell Squat," 2003.

[2] M. A. O'Reilly, D. F. Whelan, T. E. Ward, E. Delahunt, and B. M. Caulfield, "Technology in Strength and Conditioning: Assessing Bodyweight Squat Technique with Wearable Sensors," *J Strength Cond Res*, vol. 31, no. 8, pp. 2303–2312, Aug. 2017, doi: 10.1519/JSC.0000000000001957.

[3] T. Kernozek, M. Schiller, D. Rutherford, A. Smith, C. Durall, and T. G. Almonroeder, "Real-time visual feedback reduces patellofemoral joint forces during squatting in individuals with patellofemoral pain," *Clinical Biomechanics*, vol. 77, Jul. 2020, doi: 10.1016/j.clinbiomech.2020.105050.

[4] K. Kim, S. Oh, and D. K. Kim, "Development of Squat Posture Guidance System Using Kinect and Wii Balance Board," *J. lnf. Commun. Converg. Eng*, vol. 17, no. 1, pp. 74–83, 2019, doi: 10.6109/jicce.2019.17.1.74.

[5] Sunway University College, IEEE Instrumentation and Measurement Society, and Institute of Electrical and Electronics Engineers, *HAVE 2019 : IEEE International Symposium on Haptic Audio-Visual Environments & Games : 2019 symposium proceedings : Kuala Lumpur, Malaysia, October 3-4, 2019*.

[6] R. Ogata, E. Simo-Serra, S. Iizuka, and H. Ishikawa, "Temporal distance matrices for squat classification," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE Computer Society, Jun. 2019, pp. 2533–2542. doi: 10.1109/CVPRW.2019.00309.

[7] IEEE Computational Intelligence Society and Institute of Electrical and Electronics Engineers, *Proceedings of the 2014 International Joint Conference on Neural Networks : July 6-11, 2014, Beijing, China*.

[8] M. M. F. Donia, A. A. A. Youssif, and A. Hashad, "Spontaneous Facial Expression Recognition Based on Histogram of Oriented Gradients Descriptor," *Computer and Information Science*, vol. 7, no. 3, Jul. 2014, doi: 10.5539/cis.v7n3p31.

[9] K. Han, Q. Yang, and Z. Huang, "A two-stage fall recognition algorithm based on human posture features," *Sensors (Switzerland)*, vol. 20, no. 23, pp. 1–21, Dec. 2020, doi: 10.3390/s20236966.

[10] IEEE Computational Intelligence Society and Institute of Electrical and Electronics Engineers, *Proceedings of the 2014 International Joint Conference on Neural Networks : July 6-11, 2014, Beijing, China*.

[11] M. F. Antwi-Afari, H. Li, Y. Yu, and L. Kong, “Wearable insole pressure system for automated detection and classification of awkward working postures in construction workers,” *Autom Constr*, vol. 96, pp. 433–441, Dec. 2018, doi: 10.1016/j.autcon.2018.10.004.

[12] J. Lee, H. Joo, J. Lee, and Y. Chee, “Automatic classification of squat posture using inertial sensors: Deep learning approach,” *Sensors (Switzerland)*, vol. 20, no. 2, Jan. 2020, doi: 10.3390/s20020361.

[13] R. F. Escamilla and M. W. Krzyzewski, “Knee biomechanics of the dynamic squat exercise,” 2001. [Online]. Available: http://www.acsm-msse.org

[14] I. Cohen and H. Li, “Inference of Human Postures by Classification of 3D Human Body Shape,” 2003.

[15] N. Rungsawasdisap, X. Lu, Z. Zhang, A. Yimit, M. Mikami, and Y. Hagihara, *Squat Movement Recognition Using Convolutional Neural Network The Perception Neuron Based Tele-Rehabilitation System View project Squat Movement Recognition Using Convolutional Neural Network*. 2019. [Online]. Available: https://www.researchgate.net/publication/336775404