Swing and Head Synchronisation in Golf using Video Analysis

Marc Marais
Supervisor: Dr. Dane Brown
Computer Science Department, Rhodes University

April 23, 2021

Abstract

Computer vision in golf is commonly found using ball-tracking technology, but what about extending this to player swing and head posture sequencing? Thus, enabling the amateur golfer to pick up when their swing and head position are out of sync and improve this aspect of their game. Using an existing dataset GolfDB and a combination of image processing and machine learning techniques, an attempt will be made to achieve this. Action recognition is the core aspect of making this possible. Therefore, recognising the golf club will be essential. Various image processing techniques may be needed to extract the golf club from the video frames. The machine learning model will need to train the dataset on the correct and incorrect head position during the swing sequence from the preprocessed data. The machine learning model will need to use neural network models that first need to be determined. From this, the best model will be trained and tested to improve prediction accuracy when checking for synchronisation between head position and swing sequence.

1 Introduction

There is no doubt that golf is one of the more complex and challenging sports around today. The fine-tuning and complexity behind the golf swing itself comes under much scrutiny. With little research or studies on the effect of different swings, much of the debate regarding the correct body posture and swing technique stems from professional golfers and coaches personal opinions (Smith et al., 2012). Correct head and golf ball alignment is one of the key elements to a good swing (Glazier and Lamp, 2013).

The golf swing action can be recognised using machine learning, deep learning, and neural network algorithms. This study will attempt to pick up correct and incorrect head alignment during a golf swing. Golf swing sequencing has been achieved using a lightweight deep neural network (McNally et al., 2019). Using this study, which uses a labelled video database to sequence golf swing events, head alignment will be checked for synchronisation with the swing sequence events.

According to Glazier and Lamp (2013) head movement during the golf swing is linked to body movement. It is virtually impossible to keep the head completely still during the golf swing but avoiding excessive head movement is necessary.

Many studies around golf swing sequencing use sensors and optical cameras to detect the swing pattern and events (Ko and Pan, 2020; McNally et al., 2019). Sensors and optical cameras are costly implementations to study swing events and sequencing. Thus, the use of computer vision and deep learning to analyse golf swing events is more practical and accessible to amateur golfers.

Therefore, there is a need for a cost-effective and easily accessible solution in the field of computer vision to aid golfers in the analysis of their swing and head movement.

2 Related Studies

The use of computer vision to aid sports is growing, used by both broadcast viewers and in training and coaching. Tracking players and ball tracking forms a large part of computer vision (Thomas et al., 2017). In the golfing world, the use of ball-tracking to track a shot's trajectory has come in handy in assisting viewers in following the shot trajectory (Thomas et al., 2017).

Gehrig et al. (2003) presented an image processing technique to extract a golf club from a video frame, by first looking for motion detection and then applying a binary mask over the motion detection (Gehrig et al., 2003). The next step was to grayscale the frame. Using the Canny method of edge detection, extraction of the straight aspects of the image such as the golf club shaft and arms can take place (Canny, 1986; Gehrig et al., 2003). From the extracted edges, close parallel segments are merged into one component.

The success rate of this study was promising, although complications did occur. As the golf swing is a fast action itself and a club shaft is thin, a high-velocity swing may blur the club shaft (Gehrig et al., 2003).

Research conducted, in the field of golf swing analysis, by McNally et al. (2019) aimed to create a sequence of golf swing events through localising each event to a single frame. Events were made up of the address, toe-up, mid-backswing, top, mid-downswing, impact, mid-follow through and finish—these events were used to analyse and enable consistent evaluation of the golf swing. The image processing methods to extract these events frame-by-frame were implemented using the ImageNet database McNally et al. (2019).

Using a benchmark database GolfDB, which consists of 1400 labelled videos and a combination of SwingNet, McNally et al. (2019) implemented a lightweight deep learning neural network to analyse the performance of golf swings. SwingNet compromises a network architecture design incorporating lightweight convolutional neural networks (CNNs) to enable effective mobile deployment (McNally et al., 2019).

Convolutional neural networks form part of the artificial neural network

(ANN) architecture. CNNs focus area is to assist with solving complex image pattern recognition tasks (O'Shea and Nash, 2015). CNNs are very effective when implemented with labelled datasets and combined with feature extraction systems such as SIFT (scale-invariant feature transform) or HoG (Histogram of Oriented Gradients (LeCun et al., 2010).

O'Shea and Nash (2015) differentiates ANNs from CNNs in that CNNs use neurons that self-optimise through learning. The only drawback with CNNs is the large datasets required to train the model, which opens the door to overfitting. Overfitting reduces the model's ability to generalise and predict effectively in training, test, and prediction data sets (O'Shea and Nash, 2015).

A comparison was drawn regarding the accuracy of real-time vs slow-motion videos, and it was found the detection of the golf swing events was more accurate using real-time videos (McNally et al., 2019). Overall the study found that the CNN SwingNet average a rate of 76.1% at detecting all eight events in the golf swing and 91.8% at picking up six of the eight events. This provides a basis to build on a database that works towards predicting human posture and synchronising head movement with golf swing events.

Another study conducted by Ko and Pan (2020) also looked into swing sequencing but included body sway analysis of a swing. Ko and Pan (2020) made use of a single frontal facing camera as well as a motion capture to perform 3D analysis of the swing and body motion.

Classification of the swing events and quantitative information on twisting angles of the head, upper body, pelvis and shoulder are done through a deep learning regression model (Ko and Pan, 2020). The regression model is based on a bi-directional long short term memory (Bi-LSTM) neural network (Ko and Pan, 2020). In the forward direction, sequential temporal features are implemented (Ko and Pan, 2020). In the backward direction, temporal features are incorporated. Both forward and backward features improve classification performance (Ko and Pan, 2020).

Ko and Pan (2020) used a concept of golf known as head-up analysis, the changing of the central axis height and raising the head before impact with the golf ball. This common mistake made by amateur golfers causes the golfer's swing to become unbalanced, leading to a poor shot in either the form of an offset trajectory or a missed shot (Ko and Pan, 2020).

Swing events are captured and extracted as a sequence of images using the sequence feature extraction and classification network (SFEC-net) (Ko and Pan, 2020). SFEC-net is made up of three convolution layers, three pooling layers, and two fully connected layers. After extracting the swing event images, head-up analysis is conducted, creating three-dimensional and three-axis rotation angles of head movement (Ko and Pan, 2020).

Table 1: Results for golf swing event classification using SFEC-net (Ko and Pan, 2020)

Section	Accuracy	Recall	Precision	F1-score
1	93.9192	0.9392	0.9782	0.9583
2	96.9170	0.9692	0.9303	0.9493
3	92.1648	0.9216	0.8939	0.9076
4	89.2368	0.8924	0.9451	0.9180
5	84.8765	0.8488	0.7618	0.8029
6	97.0079	0.9701	0.9509	0.9604
7	99.1911	0.9919	0.9949	0.9934
Total	95.4403	0.9333	0.9221	0.9277

The results in Table 1 show that the SFEC-net accurately predicts the swing events with the network achieving an average accuracy of 95.44% and precision of 92.21%. Although under conditions where there is a large number of hidden units, there is an increased possibility of overfitting and therefore the need to adapt parameters (Ko and Pan, 2020). This study performed well, but a way to remove the need for the motion capture suit needs to be found to make the technology readily available to all golfers.

Both studies use different techniques to achieve swing sequencing. The advantage of McNally et al. (2019) is the access to the large dataset to use when training and testing the model.

As the golfer is a crucial aspect to extract and track from the video frames, Shinde et al.'s proposal to use YOLO (You Only Look Once) combined with Liris

Human Activities dataset for fast human action recognition and localisation is an alternative to general CNNs (Shinde et al., 2018).

Object detection in YOLO uses a single CNN instead of the general two-stream CNN using fewer frames to achieve action recognition and localisation(Shinde et al., 2018). YOLO's architecture comprises 24 convolutional layers and two fully connected layers (Shinde et al., 2018).

The LIRIS dataset (Wolf et al., 2014) contains action categories relating to Human-Human interactions, human-object interactions and human-human-object interactions (Shinde et al., 2018).

The results obtained from Shinde et al.'s research indicate all three indicator values, precision, recall and f-score, to achieve a good classification model are above 80%. When contrasted to other state-of-art methods such as Mukherjee et al. (2015)'s hybrid motion optical flow combined with a binary support vector machine for classification, which used the same LIRIS dataset, YOLO's f-score of 88.358% outperformed Mukherjee et al.'s f-score of 81.27%. Therefore YOLO achieved better performance while using fewer frames to do so.

Looking further into state-of-the-art models such as YOLO and Region-based Convolutional Neural Networks (R-CNN's) (Girshick, 2015) is a worthwhile endeavour.

3 Research Statement

A labelled dataset with a supervised machine learning model can be used to analyse swing and head synchronisation in golf in videos. A system, based on this model, can be constructed to assist both amateur and professional golfers in preventing incorrect head position during their swing by detecting these anomalies. High accuracy in anomaly detection can be compared to related system to measure the systems effectiveness.

4 Research Objectives

- Analyse the existing dataset GolfDB and apply various image processing techniques to extract the required data need for the machine learning algorithm.
- Extract the golf club object from video frames.
- Develop a deep learning algorithm, using existing studies neural networks (CNNs, Bi-LSTM or YOLO) as a building platform.
- Test and evaluate the model using the dataset to ensure better prediction models. This involves dividing the dataset into training and testing data.
- Predict the moment when the head position and swing sequence are out of sync.

5 Limitations

What is too much head movement? As most golf swing posture analysis is based on coaches and professionals opinions and no actual studies have been done, this adds an element of uncertainty. To overcome this, only significant changes in head movement will be observed. An example of this would be the complete lifting of the head and taking eyes of the golf ball during the swing sequence, with particular focus on the golf club and ball impact event.

6 Plan of Action

The initial phases of this project will be time spent reviewing literature related to machine learning in analysing golf swing sequencing, with a particular focus on the types of neural networks implemented to achieve the most accurate sequencing events.

Another focus area will be experimenting with image processing techniques to extract the requiring information to sequence the swing events. Assessment of possible techniques by Gehrig et al. and action recognition through convolutional neural networks such as SwingNet and YOLO will be vital in extracting and analysing swing events and head position.

Once the desired technique and model has been accessed and selected, machine learning model design will begin. Training and testing the model will be done using the existing database compiled by McNally et al. (2019).

Critique of the model's performance will be based on precision, recall and f-score ratings. These ratings will then compare to the results achieved in McNally et al. and Ko and Pan.

7 Timeline

Deadline	To Do
2 April 2021	Draft proposal.
6-20 April 2021	Seminar Series 1 presenting draft proposals.
8 April 2021	Completion of image processing task set to gain practical competency.
7 April 2021	Review feedback from seminar series 1 and apply relevant changes.
23 April 2021	Final proposal.
31 April 2021	Be able to extract the required information from video frames or images. i.e. golf club, arm and head.
1 May 2021	Start applying machine learning knowledge gained to working towards developing a machine learning model as machine learning course will be nearing completion.
28 May 2021	Literature review submission.
31 May 2021	Suitable machine learning model selected by this stage.
15 June 2021	Final chance to change model.
31 July 2021	Working machine learning model, may still need some adapting at this stage. Testing will be done at this stage as well.
9-11 August 2021	Seminar Series 2: Presenting Project Progress.
30 August 2021	Completion of testing phase. Finalisation of technical project work.
20 September 2021	Progress report for supervisor.
4 October 2021	First full draft thesis.
11 October 2021	Submit short ACM-style paper.
18-20 October 2021	Seminar Series 3: Final Oral Presentations.
29 October 2021	Final Project Submission.
17-22 November 2021	Oral Examination.
3 December 2021	Corrected short ACM-style paper Corrected project

References

John Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. ISSN 01628828. doi: 10.1109/TPAMI.1986.4767851.

Nicolas Gehrig, Vincent Lepetit, and Pascal Fua. Visual golf club tracking for enhanced swing analysis. In *BMVC*, pages 1–10, 2003.

Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

Paul Glazier and Peter Lamp. Golf Science: Optimum Performance From Tee to Green. In *Golf Science: Optimum Performance From Tee to Green*, chapter The swing. University of Chicago Press, 2013. ISBN 9780226001272.

Kyeong Ri Ko and Sung Bum Pan. CNN and bi-LSTM based 3D golf swing analysis by frontal swing sequence images. *Multimedia Tools and Applications*, pages 8957–8972, 2020. ISSN 15737721. doi: 10.1007/s11042-020-10096-0.

Yann LeCun, Koray Kavukcuoglu, and Clément Farabet. Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE international symposium on circuits and systems*, pages 253–256. IEEE, 2010. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5537907`.

William McNally, Kanav Vats, Tyler Pinto, Chris Dulhanty, John McPhee, and Alexander Wong. GolfDB: A video database for golf swing sequencing. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2019-June:2553–2562, 2019. ISSN 21607516. doi: 10.1109/CVPRW.2019.00311.

Snehasis Mukherjee, Apurbaa Mallik, and Dipti Prasad Mukherjee. Computer Vision Systems. In Lazaros Nalpantidis, Volker Krüger, Jan-Olof Eklundh, and Antonios Gasteratos, editors, *International Conference on*

Computer Vision Systems, volume 9163 of *Lecture Notes in Computer Science*, pages 477–487, Cham, 2015. Springer, Springer International Publishing. ISBN 978-3-319-20903-6. doi: 10.1007/978-3-319-20904-3. URL `http://link.springer.com/10.1007/978-3-319-20904-3`.

Keiron O'Shea and Ryan Nash. An Introduction to Convolutional Neural Networks. *arXiv preprint arXiv:1511.08458*, pages 1–11, 2015. URL `http://arxiv.org/abs/1511.08458`.

Shubham Shinde, Ashwin Kothari, and Vikram Gupta. Yolo based human action recognition and localization. *Procedia Computer Science*, 133:831–838, 2018. ISSN 18770509. doi: 10.1016/j.procs.2018.07.112.

Aimée Smith, Jonathan Roberts, Eric Wallace, and Stephanie Forrester. Professional golf coaches' perceptions of the key technical parameters in the golf swing. In *Procedia Engineering*, volume 34, pages 224–229. Elsevier Ltd, jan 2012. doi: 10.1016/j.proeng.2012.04.039.

Graham Thomas, Rikke Gade, Thomas B. Moeslund, Peter Carr, and Adrian Hilton. Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, 159:3–18, jun 2017. ISSN 1090235X. doi: 10.1016/j.cviu.2017.04.011.

Christian Wolf, Eric Lombardi, Julien Mille, Oya Celiktutan, Mingyuan Jiu, Emre Dogan, Gonen Eren, Moez Baccouche, Emmanuel Dellandréa, Charles Edmond Bichot, Christophe Garcia, and Bülent Sankur. Evaluation of video activity localizations integrating quality and quantity measurements. *Computer Vision and Image Understanding*, 127:14–30, 10 2014. ISSN 1090235X. doi: 10.1016/j.cviu.2014.06.014.