CV_Draft_4.docx

Golf Shot Swing Recognition Using Dense Optical Flow

Jyoti Madake, Shreyas Sirshikar, Swagat Kulkarni, Shripad Bhatlawande
Dept. of E&TC
VIT
Pune, India
{jyoti.madake, shreyas.sirshikar20, swagat.kulkarni20, shripad.bhatlawande}@vit.edu

***Abstract* - This paper introduces a computer vision-based methodology for identifying golf swings. The proposed approach utilizes videographic data of golf swings and recognizes swing type and postures such as Backswing, Downswing, and Follow Through. The system experiments with techniques such as Dense and Sparse Optical Flow to identify the Golf club's trajectory. The features of these trajectories are then extracted using Histogram of Oriented Gradients. To create a resilient model, the feature of golfer's posture during a specific swing is extracted using Scale-Invariant Feature Transform with morphological image frame. The extracted features are normalized, concatenated, and classified using various classification techniques. When evaluated on the standard GolfDB dataset, the system achieves an Accuracy of 84.43% for Random Forest classification with a Precision Score of 0.7949. This study proposes a promising approach to identify golf swings using computer vision techniques, which could potentially have implications in enhancing performance analysis in the field of sports.**

***Keywords* - Swing Recognition, Golf Shot, trajectory, Gunnar-Farneback, SIFT, HOG, Feature Extraction, Body Posture, Club Movement.**

I. INTRODUCTION

Golf Swing Shot recognition finds significance as an analytical tool which assists in identification of patterns or techniques that can help golfers improve their swing and overall performance on the golf course. Computer Vision facilitates the development of systems capable of recognizing the motion by marking trajectories and classifying the corresponding movement. A golf swing is a combination of different movements viz. Setup, Backswing, Transition, Downswing, Impact, and Follow Through. The recognition of the swing enables one to perform mechanical analysis as well as performance estimation of the golfer. The existing approaches of the models for the problem employ deep learning algorithms, however machine learning algorithms along with classifiers provide enhanced results [1]. The system finds application in training tools like performance enhancement models, automated scoring systems where scoring is to be done on the basis of movements of body parts or human posture, virtual reality-based gaming, development of smart golf clubs with self-analytical properties [2]. Most existing approaches tend to build the basis of the model based on the posture of players and employ assistance of sensors to support the generated outputs. A stand-alone system that contains abundant yet significantly relevant features would be desirably ideal. This paper proposes a novel system which functions on the basis of HOG and SIFT for feature extraction followed by Random Forest Classifier Based on its high accuracy among other notable classifiers. As a fully functional application, the system demands a video input from the user, it then divides the video into frames, matches the individual images with the feature vector and predicts its class. After subsequent iterations throughout the video, the final count of instances of each class is given based on which the type of swing is recognized.

II. RELATED WORK

This section describes all previously performed work by various authors to detect positions and types of swings occurring in an event of hitting a golf shot. Various concepts of Computer Vision, Embedded Systems and Machine Learning are reviewed in order to find the most optimal methodology for the swing detection in a golf shot. Below given is the summary of methodologies described by authors.
The study proposed by Z. Jiang, H. Ji, S. Menaker and J.-N. Hwan introduces a golf swing analysis system that uses a monocular camera and human pose estimation to track the player's body movement during the swing [3]. The method proposed in [4] for tracking a golf ball and club in a video using a combination of HOG and STV features whereas [5] presents a system for golf ball detection and tracking using a CNN model. The authors of [6] propose a system which implements sensors like Inertia Measurement units (accelerometer and gyroscope) to detect the relative changes in position of the golfer's body. This data from the sensor is then also validated using a stereo camera and thus a model is trained to obtain accurate results. From [7] a solution can be summarized that combines the use of CNN for the detection of objects and the Kalman filter for prediction. The research presented in [8] aims on monitoring the posture of the player from a video. The problem of overlapping movements arises when relying solely on the golfer's silhouette. To overcome this issue, this paper presents an optical flow-based approach to recognize the position of a golf club during any instant golf shot. This approach enables the tracking of the most concerned and moving part of the golf stick over time in different contours, eliminating the overlap problem.
In [9] the analysis is performed from a down the line perspective and calculations are carried out using a low-cost

TPU for data processing and inference. The limitations and inaccuracies of the hardware and pose estimation are addressed by using a Savitzky-Golay filter. Authors of [10] employ Hough transform and Kalman Filter to enhance the accuracy of recognition. The effectiveness of the model is demonstrated through experiments on the NAO robot. Similar to earlier presented systems, authors of [11] employ the assistance of sensors and sports equipment systems for golf swing data classification. The authors investigate the use of CNN to process data from embedded multi-sensors in golf swings and classify the data into different categories based on the players and swings. Four different CNN-based classifiers are customized and tested on real-world swing data collected from multiple sensors, including strain gauge sensors, accelerometers, and gyroscopes. The model presented in [12] is a vision-based solution for detecting a golf ball in videos. The authors in [13] propose a system that incorporates a golfer identification mechanism that is automated and utilizes conventional machine learning algorithms like CNN to categorize swing events. Classification based on conventional algorithms like LinearSVM produces results with 88.3% recall, whereas the pruned VGGNet achieves 87.9% recall. Overall, the model outperforms the Bi-directional Long Short Term Memory approach, which has a recall score of 76.1%. Chen et al discuss the development of intelligent glasses for improving golf swing skills. The major focus is to detect posture faults and the system achieves an accuracy of 95.1%, 94.7%, and 86.1% in head, hip, and body lines respectively [14]. [15] presents a similar model based on 2D joint point information and analysis of swing shots based on human posture movement. The model presented in [16] finds significance in its approach of using an unsupervised learning model in which the system compares 2 motion sequences and identifies significant differences between the 2 frames. The system then analyses the human posture between 2 intermediate frames by synchronizing and detecting the discrepancies.

Similarly, multiple proposed systems are reviewed and the methodology of implementation is understood. Most of the methods involved the implementation of CNN interfaced with the Kalman filter, stereo camera, IMU sensors, and HOG feature descriptor. The integration of sensors on a sportsman's body can be invasive or cumbersome leading to discomfort for the professional while playing. Also, this type of feature extraction involves data loss and is prone to more inaccurate data acquisition which can be a potential limitation when implemented in real-time. The proposed system also aims to have low computation cost which can be a challenge when considered to implement using Neural Network based approach. Thus, this survey is concluded by selecting the HOG feature descriptor for golf club detection integrated with Scale-Invariant Feature Transform (SIFT) for body posture identification and classifying the features using multiple supervised learning models to design a robust and accurate system for identifying golf shot and swings.

III. METHODOLOGY

This paper presents an optical flow method implemented by Gunnar-Farneback technique to detect trajectory of the golf club along with body posture of a golfer while playing a golf shot and classify the detected trajectory and posture into backswing, downswing and follow through using multiple feature classification techniques. Fig. 1. depicts the presented model schematically.

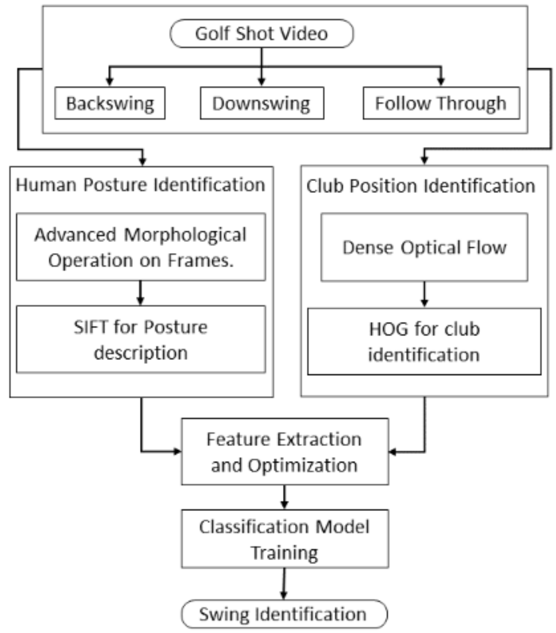

Fig. 1. Schematic Representation of the presented model.

A. *Dataset Selection and Pre-Processing*

[17] is a good quality video dataset of golfers having a complete golf shot is used in the presented model. The dataset consists of 1400 unique golf shot videos recorded and resized to 160x160 pixels from two different angles namely down-the-line and face-on. These videos mainly depict a golf shot in 8 different posture sequences as depicted in Fig. 2. First image indicates the posture to address golf shot, second indicates posture for toe-up, third is the posture presenting a middle of backswing movement, fourth is the posture of having golf club at the topmost position in a swing, fifth posture is depicts having the golf stick at middle of a downswing, sixth posture indicates an impact of golf club and the golf ball, seventh indicates the posture of golf stick following through and the eighth posture indicating the finish of golf shot.

Considering these 8 postures the swing of a golfer during the entire event can be classified into 3 major classes as backswing, downswing and follow through as the occurrence of these postures is significant throughout the videos. This dataset is filtered in order to have all videos recorded in face-on angle and those videos recorded in down-the-line angles are removed. Further the new dataset is divided into three different subsets of video datasets on the basis of classes defined. As the model is to be trained on this dataset and it has to be tested over same, the dataset is divided into 4:1 ratio with 80% of data used for training and 20% data being used for testing.

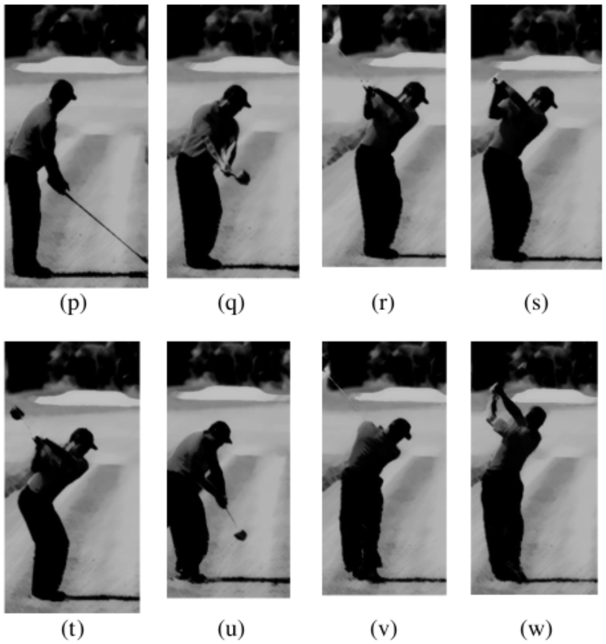

Fig. 2 Uniquely identified 8 golf swing postures: (p) addressing a golf shot, (q) having toe up for a shot, (r) middle of backswing, (s) golf club at topmost position, (t) middle of a downswing, (u) an impact of golf club and ball, (v) golf stick following through and (w) finishing a golf shot.

B. Feature Extraction and Optimization

Extraction of unique features is implemented using multiple feature identification and description methods. An optical flow approach is experimented to identify the swing shot by predicting the trajectory key points for the club part of a golf stick throughout the motion and thus evaluating HOG for individual frames to identify the position of golf club in this trajectory.

Optical flow analysis is a computer vision technique implemented to trace the sequence of movement for different objects in a video. It works by analyzing the apparent motion of pixels between consecutive frames of the video and computing the displacement vectors of each pixel. These displacement vectors represent the optical flow, which can be used to estimate the velocity and direction of object in motion. Mathematically, Optical Flow can be expressed as:

$$\frac{dI}{dx}p + \frac{dI}{dy}q + \frac{dI}{dt} = 0 \quad (1)$$

Here, I represent the pixel intensity, p is the speed of a pixel moving horizontally (x axis), q is the speed of a pixel moving vertically (y axis) and $\frac{dI}{dt}$ represents the change of intensity w.r.t time.

This trajectory estimation technique can be implemented in two ways: firstly, sparse optical flow as implemented by Lucas-Kanade algorithm which focuses mainly on the corners of the object and predict the trajectory of these corners and second technique being dense optical flow as implemented by Gunnar-Farneback algorithm. It is a differential optical flow method that uses a pyramid approach to improve the accuracy of the flow estimation while maintaining efficiency. The flow is computed by solving the system of linear equations derived from the polynomial expansion using a least-squares method. Figure 3 presents the frames detecting golf club trajectory when experimented using Lucas-Kanade algorithm and Gunnar-Farneback algorithm.

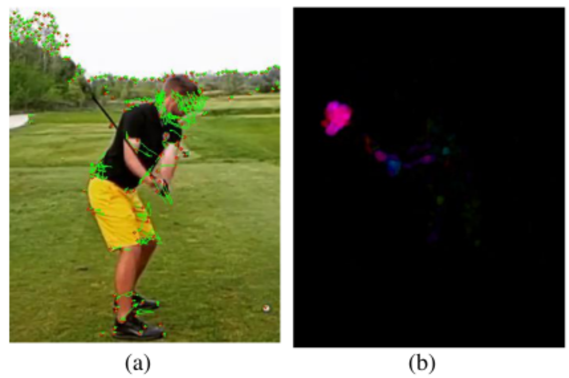

Fig. 3. Frame of a video implemented by (a) Optical Flow using Lucas-Kanade and (b) Optical Flow using Gunnar-Farneback Algorithm.

From Fig. 3, it can be concluded that Gunnar-Farneback effectively identifies the trajectory key points of the golf club whereas Lucas-Kanade algorithm mainly focuses on corners of the content present in individual frames. This result of optical flow visualization is stored in the form of HSV image with the saturation being maximum. The magnitude and angle of the optical flow vector is calculated and is normalized to fit into 0-π radians for angle whereas magnitude is normalized to fit into 0-255 values. For every image thus formed, histogram of intensity is plotted and divided into 10 bins. The set of features extracted using optical flow are namely mean intensity, standard deviation, variance and histogram bin values of intensity for each frame from the input video. Further to identify the Golf Club based on the intensity values of pixels, Histogram of Oriented Gradient is applied on frames of video. Images from every class consist of golf club moving in a specific direction, thus the count of gradient orientations will also be significantly different for every class. This set of HOG features are thus extracted having variable length depending upon type of images and class of the swing. The value of these features mainly ranges from 0 to 1. Therefore, these values are classified into 10 different sets of height 0.1 and the count of pixels having value in a given range is stored as a feature of HOG.

In order to design a robust model, the posture of the golfer at a given instant along with the swing class is taken into consideration. Scale-Invariant Feature Transform (SIFT) is applied on a frame after advanced morphological operations performed, to detect the posture of the golfer. The key points and feature descriptors are obtained for the arm, elbow joint, knee joint and leg. These obtained 128 features of SIFT for an image frame are then concatenated into existing feature vector. Algorithm 1 summarizes the steps involved in feature extraction and concatenation.

The concatenation is done for features extracted from optical flow trajectory, HOG and SIFT. Initially, 8 features of trajectory are extracted with 14 features of HOG using steps mentioned in Algorithm 1. Further including 128 features of SIFT the total of 150 features for every image frame is taken into consideration. These features are further normalized and stored as a 1-dimensional array in a CSV file with a row indicating an image and the columns storing distinct feature values.

Algorithm 1: Extracting SIFT and HOG features of trajectory

Input: Labelled hsv and morphological images of a swing class.

Output: Feature vector of image frame identified for a trajectory and golfer's body posture.

Initialization

```
no_bin ← 10
for i in image_folder1:
  img1 ← read optical flow image in grayscale
  Compute mean, std. deviation, variance of img1
  hist, bins ← Compute histogram(img1) for no_bin
  fv ⟵ mean, std. deviation, variance, hist
  orient ← 9, ppc ← (16,16), cpb ← (2,2)
  h ← hog(img1, orient, ppc, cpb)
  Compute mean_h, std. dev_h, var._h, sum_h of h
fv ⟵ fv + mean_h + std. dev_h + var._h + sum_h
for i in range(0,1,0.1):
  for j in h:
    if j ≤ i then:
      hoghist[i * 10] ⟵ hoghist[i * 10] + 1
    end if
  end for
 end for
 fv ⟵ fv + hoghist
end for
for i in image_folder2:
  img2 ← read morphological image in grayscale
  sift ← Create SIFT object
  key, descriptors ← Compute sift(img2)
end for
fv ⟵ fv + descriptors
return fv
```

end

C. *Feature Classification and Model Training*

This section presents the feature classification technique, which is done through experimenting different supervised learning models.

To test and train the feature vectors a new column is introduced in the data frame which stores the class of image in a binary form as 1 for the swing class under consideration and 0 for other swing types. All the feature vectors along with its class are then divided in 4:1 ratio, having 80% of training data. Based on this data 4 models are trained while performing hyperparameter tunning of these models. Steps involved in classification process is depicted in Fig. 4.

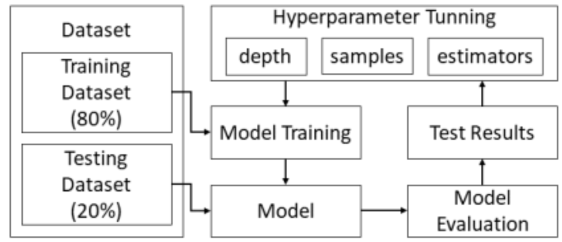

Fig. 4. Feature classification procedure implemented in model training.

For Decision Tree, parameters grid is defined that consists list of values that the model considers while performing classifications. The max depth parameter of this grid, which has potential values of 2, 4, and 6, determines the highest possible depth of the tree. Second, the minimum sample split parameter, which can have a value of 2, 5, or 10, indicates the least count of samples desired to divide at an internal node. Third, the minimum leaf samples parameter, which shows the least count of samples expected to be at a leaf node, can have a value of 1, 2, or 4. Fourth, the criterion parameter is the metric that is used to gauge the split's quality. It can be of two types, Gini impurity or entropy, and it determines the split's quality using the data gathered. Similar to this, the parameter tuning of Random Forest also depends on parameters like max depth, minimum sample splitting, minimum sample for a leaf, and criterion, along with a few new parameters like the quantity of estimators, and the maximum quantity of features, which can have values of "sqrt" indicating square root, and "log2" indicating logarithmic value of total number of features in order to perform data splitting. The parameter grid of K Nearest Neighbour have parameters like number of neighbors which can have value like 5, 7, 13, 17 or 23. Weight parameter that can be assign either uniform weights to all neighbors or based on their distance and metric parameter that indicates the metric used for distance calculation and can have value either Euclidean or Manhattan. The Support Vector Machine is also trained for different kernels like, linear kernel, 2-degree polynomial kernel, 3-degree polynomial kernel, rbf kernel and sigmoid kernel.

IV. Results and Discussion

The models experimented are tested and evaluated using various performance parameters like, train accuracy, test accuracy, precision score, F1 score and confusion matrix.

After tunning the hyperparameters for each experimented model the best hyperparameters found for Decision tree are, entropy criterion, and best splitter method with 6 being the maximum possible depth tree should have, minimum of 2 samples required to split as a leaf node and 10 samples required for splitting data into separate node. For Random Forest hyperparameters with best accuracy are, criterion being entropy, maximum features having value 'sqrt', minimum samples for leaf being 1 and minimum samples for split being 5 and number of estimators being 200. Similarly, the hyperparameters with best results obtained for KNN is Manhattan distance measurement metric, 5 nearest neighbors with distanced based weights of neighbors.

As mentioned earlier the dataset is split for testing of the model as well as training of the model in a ratio of 1:4, which means 20% from entire data is used in test the presented system. While testing the computation time of all the model is also accounted. Model framed using KNN is found to be least time-consuming during testing.
Table 1 indicates, the test accuracy of multiple classifiers where random forest has an accuracy of 84.43% while Decision Tree having test accuracy of 80.66% and KNN having accuracy 75.94%. Various performance parameters involved in evaluation are presented through the graph shown in Fig. 5.

TABLE I. COMPARATIVE PRESENTATION OF CLASSIFIERS.

Sr. No.	Classifier	Accuracy
1)	Support Vector Machine	75%
2)	K Nearest Neighbour	75.94%
3)	Decision Tree	80.66%
4)	Random Forest	84.43%

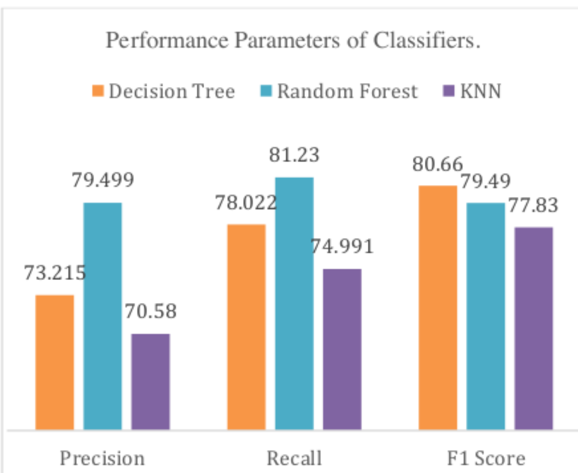

Fig. 5. Graphical representation of performance parameters evaluated for three different models.

Considering other performance parameters like precision and F1 Score along with accuracy it is found that Random Forest is the best performing model in terms of classification. Further the entire system is also tested and evaluated by manually providing an input of video for a specific class apart from the ones provided in either testing or training datasets, where the model has estimated the swing correctly.

V. CONCLUSION

Computer Vision enables the development of the model for golf swing shot recognition. Based on the experimentation and results obtained, Gunnar-Farneback Optical flow model provides better results in terms of the significant features for trajectory estimation based on the motion of the golf club against the conventional deep learning or sensor-based approach. The features extracted from the HOG are supported by SIFT which provides better uniqueness for training the model. The Random Forest Classifier provides the best testing accuracy of 84.43% among other notable classifiers. The system is insubstantial in extracting unique features for certain instances where the downswing follows the same trajectory as backswing, and instances where the club comes at same position in follow through where it was in backswing. The scope of further development in the system includes calculation of mechanical metrics viz speed of the club, angle of trajectory, distance of the ball based on the velocity at the impact stage.

REFERENCES

[1] C Sungkuk., K Donghoon., C Hyeong-Rae. et al. "A sensor-aided self-coaching model for uncocking improvement in golf swing." Multimedia tools and applications, 72, pp.253-279.
[2] G Nicolas, V Lepetit, and P Fua. "Golf club visual tracking for enhanced swing analysis tools." In British Machine Vision Conference.
[3] J Zhongyu, J Haorui, M. Samuel and H. Jenq-Neng, "GolfPose: Golf Swing Analyses with a Monocular Camera Based Human Pose Estimation," 2022 IEEE ICMEW, pp. 1-6.
[4] W. Li, X. Lou, M. Dong, and L. Zhu. "Golf video tracking based on recognition with HOG and spatial–temporal vector." International Journal of Advanced Robotic Systems 14, no. 3.
[5] X. Zhang, Z. Tianxiao, Y. Yiju, W. Zongbo, and W. Guanghui. "Real-time golf ball detection and tracking based on convolutional neural networks." In 2020 IEEE ICSMC, pp. 2808-2813.
[6] N. K. Nam Cao, K. Hee Jun, and Young S. S. "Golf swing motion tracking using inertial sensors and a stereo camera." IEEE Transactions on Instrumentation and measurement 63, no. 4, pp. 943-952.
[7] Z. Tianxiao, X. Zhang, Y. Yiju, W. Zongbo, W. Guanghui. "Efficient golf ball detection and tracking based on convolutional neural networks and kalman filter."
[8] K. F. Sim, K. Sundaraj. "Human motion tracking on broadcast golf swing video using optical flow and template matching." In 2010 IEEE ICCAIE, pp. 169-173.
[9] Theodore T. K., Z. A. Mohamed, and B. P. Michael "Applying pose estimation to predict amateur golf swing performance using edge processing." IEEE Access 8: 143769-143776.
[10] S. Erwei, C. Long, X. Yingli, H. Bin. "Target Detection in NAO Robot Golfing." In Journal of Physics: Conference Series, vol. 1828, no. 1, p. 012171. IOP Publishing.
[11] J. Libin, B. Rongfang, W. Hao, et al. "Golf swing classification with multiple deep convolutional neural networks". IJDSN. 2018, 14, no. 10.
[12] K. Akshay, S. Patrick, N. Mircea, and H. C. Steven. "Vision-based Assessment of Instructional Content on Golf Performance." In Proceedings of the 4th International Conference on Image Processing and Machine Vision, pp. 78-83.
[13] Marais Marc. . Golf Swing Sequencing using Computer Vision.
[14] Chen, Hua-Tsung, Tzu-Wei Huang, Chien-Li Chou, Hou-Chun Tsai, and Shih-Yin Lee. "Improving golf swing skills using intelligent glasses." In 2015 Visual Communications and Image Processing (VCIP), pp. 1-4. IEEE, 2015.
[15] Wang, Meng, and Tong Zang. "Research on Intelligent Sports Training System for Golf Based on Body Sense Recognition." Computational Intelligence and Neuroscience 2022 (2022).
[16] Liao, Chen-Chieh, Dong-Hyun Hwang, and Hideki Koike. "AI Golf: Golf Swing Analysis Tool for Self-Training." IEEE Access 10 (2022): 106286-106295.
[17] McNally, William, Kanav Vats, Tyler Pinto, Chris Dulhanty, John McPhee, and Alexander Wong. "Golfdb: A video database for golf swing sequencing." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0-0.

CV_Draft_4.docx

ORIGINALITY REPORT

8%
SIMILARITY INDEX

PRIMARY SOURCES

1	Yanting Zhang, Qing'Ao Wang, Fuyu Tu, Zijian Wang. "Automatic Moving Pose Grading for Golf Swing in Sports", 2022 IEEE International Conference on Image Processing (ICIP), 2022 Crossref	34 words — 1%
2	www.researchgate.net Internet	27 words — 1%
3	arxiv.org Internet	26 words — 1%
4	aircconline.com Internet	24 words — 1%
5	people.cs.nctu.edu.tw Internet	23 words — 1%
6	vuir.vu.edu.au Internet	22 words — 1%
7	Young Jin Kim, Kyun Dae Kim, Sung Hoon Kim, SeungGwan Lee, Hyon Soo Lee. "Golf swing analysis system with a dual band and motion analysis algorithm", IEEE Transactions on Consumer Electronics, 2017 Crossref	20 words — < 1%
8	elcvia.cvc.uab.cat Internet	

19 words — < 1%

9 link.springer.com
Internet
19 words — < 1%

10 www.hindawi.com
Internet
19 words — < 1%

11 www.atlantis-press.com
Internet
17 words — < 1%

12 export.arxiv.org
Internet
15 words — < 1%

13 ijsret.com
Internet
15 words — < 1%

14 Xiaoxiao Zhang, Jie Zhong. "Thermal Environment Analysis and Thermal Comfort Evaluation of Huizhou Folk Dwelling Houses Guided by Auxiliary Interior Space Layout", International Journal of Heat and Technology, 2022
Crossref
14 words — < 1%

15 dblp.dagstuhl.de
Internet
14 words — < 1%

16 doaj.org
Internet
13 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF