Poster Abstract: Mobile Golf Swing Tracking Using Deep Learning with Data Fusion

Hong Jia
UNSW & Data61-CSIRO
Sydney, Australia
h.jia@unsw.edu.au

Yuezhong Wu
UNSW & Data61-CSIRO
Sydney, Australia
yuezhong.wu@student.unsw.edu.au

Jun Liu
UNSW
Sydney, Australia
jun.liu@student.unsw.edu.au

Lina Yao
UNSW
Sydney, Australia
lina.yao@unsw.edu.au

Wen Hu
UNSW & Data61-CSIRO
Sydney, Australia
wen.hu@unsw.edu.au

ABSTRACT

Swing tracking is one of the key information for many sports such as golf. One approach to track swing is to use IMU to measure linear acceleration then get position by two-time integration. However, the complex noise model of the IMU limit the accuracy of the tracking. Another approach is to use depth sensor to measure 3D location of a point of interest directly. Unfortunately, the depth sensor-based approach cannot accurately measure the trajectory of a swing when the sensor is occluded, which happens regularly. To overcome these limitations, we develop a novel solution to make use of these two sensor modalities (i.e., IMU and depth sensor) by a novel deep neural network to produce high precision swing trajectory tracking. The learned network automatically makes use of the IMU when the depth sensor is occluded, and relies on depth sensor when IMU signal is noisy. Our experiment shows that the proposed method outperforms state-of-the-art swing tracking method by 62% of error reduction.

CCS CONCEPTS

• **Computer systems organization → Sensor networks**.

KEYWORDS

mobile computing, swing tracking, neural networks, sports analytics

ACM Reference Format:
Hong Jia, Yuezhong Wu, Jun Liu, Lina Yao, and Wen Hu. 2019. Poster Abstract: Mobile Golf Swing Tracking Using Deep Learning with Data Fusion. In *SenSys '19: Conference on Embedded Networked Sensor Systems, November 10–13, 2019, New York, NY, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3356250.3361968

1 INTRODUCTION

Sports analytics is a thriving industry whose market size is estimated to reach five billion dollars within the next five years. Swing tracking is a key information within many sports such as golf, tennis and baseball.

Current solutions for swing tracking are broadly categorized to vision-based and IMU-based solutions. One example concerns benchmark OpenPose [1], which is a real-time library to extract multi-person skeleton points from videos. However, OpenPose can only reach 6 to 12 frames per second (fps). This is too slow for many swing sports such as golf, which professional golfers can reach speed up to 67 m/s. The resolution for OpenPose here can only reach $67 \div 12 = 5.58$ m. Which is challenging to achieve accurate golf swing tracking. Another problem for vision-based methods is occlusion. During the swing, the depth-sensor's view is occluded in many swing stages; as a result, the reconstructed 3D positions are usually inaccurate and ambiguous.

Another solution is to use one or multiple IMU for swing tracking. However, the complex noise model within the IMU limits its accuracy for swing tracking. Shen et. al. propose a method for IMU tracking, which eliminates the accumulated error for motion tracking[5]. However, this work still suffers from the complex noise model within the IMU and its accuracy is close to that of IONet, which is based on deep learning model [2].

To address the above limitations, we propose a novel solution to fuses IMU and depth sensor measurements to produce high accurate swing tracking. The proposed method relies on depth sensor when the IMU signal is noisy and relies on IMU when the depth sensor is occluded or misses key tracking information due to low sampling rates. By tailoring a deep neural network, the proposed method can calibrate different sensor modalities automatically at different golf swing stages to reach high accuracy swing tracking, and produce highly accurate tracking results.

2 PROPOSED FRAMEWORK

The proposed system takes the IMU and depth sensor's measurements as the input. The IMU ($\mathbf{I} \in \mathbb{R}^{6\times T}$) includes 3 axis of raw acceleration signals and 3 axis of raw gyroscope signals, where $\mathbf{T}$ is the length of time stamps. Not that the proposed method does not take compass measurements as input as it lowers down the sampling rate of the IMU. The depth-sensor $\mathbf{K} \in \mathbb{R}^{3\times T}$ is a 3 axis of raw location of a point of interest (being tracked) signal.

The proposed deep neural network features an encoder-decoder framework. We denote concatenation as a square bracket to group

SenSys '19, November 10–13, 2019, New York, NY, USA

ACM ISBN 978-1-4503-6950-3/19/11…$15.00
https://doi.org/10.1145/3356250.3361968

the signals from two sensor modalities as the input. Thus, the input can be written as:

$$\mathbf{X} = [\mathbf{K}, \mathbf{I}]. \quad (1)$$

Encoder: The encoder first fuses the inputs signals of IMU and depth-sensor through one Convolutional Neural Network (CNN) layer with 3×3 kernel size. This is followed by batch normalization [3], a Relu activation function and an average pooling layer.

Then, a multiple domain block encodes the filtered signal before concatenating into hidden features. The Encoding block is shown in Figure 1. Here, Conv 1×1 and 3×3 denote 2D Convolutional layers with 1×1 and 3×3 Kernel sizes, respectively. Each convolutional layer is followed by batch normalization and a Relu activation function.

After that, a shuffle block shuffles the learned features to make the system robust. Finally, the encoder uses a Conv 3×3 to convert the hidden features for the input format of the decoder.

For convolutional layer l, filter i, and denote j as the index of domain shown in Figure 1, the encoder process can be viewed as:

$$P_l^{ij} = ReLu(BN(\sum \omega_l^{ij} P_{l-1}^{ij} + b_l^{ij})). \quad (2)$$

where BN is the batch normalization [3], and ω_l^{ij} and b_l^{ij} represents the weight and bias parameters, respectively, which will be learned during the training process. The input is X for the first convolutional layer (i.e., $P_1 = X$) and the output (feature map) of previous layers for the other layers.

Decoder: The decoder learns the spatial relationship of the learned features and predicts the swing trajectories. The decoder is a typical one layer Long Short-Term Memory (LSTM) neural network with 128 hidden units. The output is the position of point of interest (being tracked) in 3D space (**Pos** $\in \mathbb{R}^{3 \times T}$). Recall that $\mathbb{T}$ is the length of time stamps in a swing segment.

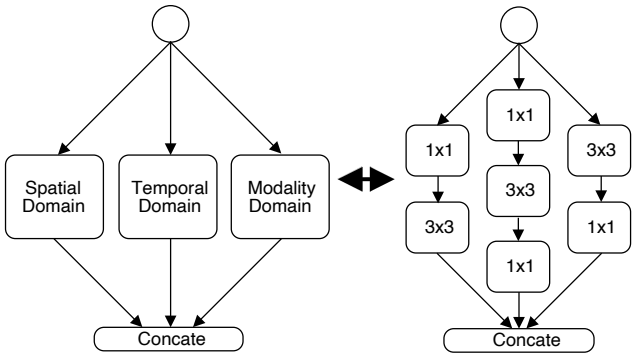

Figure 1: Encoding Block

3 EVALUATION

We collect a dataset of golf swings with a motion-capture system comprising 25 high-fidelity cameras called OptiTrack [4] as the ground truth. We select Sparkfun M0 as the IMU with a sampling rate of 240 Hz, and the Xbox One V2 as the depth-sensor with a sampling rate of 30 Hz. We have collected 1,000 swing segments from four subjects, and each segment has $T = 152$ samples.

The overall performance of different tracking approaches are shown in Table 1. The table shows how the proposed method outperforms the state-of-the-art approach IONet that relies on the IMU measurements only by reducing tracking error by approximately 60%. Figure 2 is randomly selected example to compare predictions of different models. We can see how the proposed method outperforms other models.

Table 1: Overall 3D position tracking comparison.

Model	MAE (cm)	RMSE (cm)
IONet	11.02 ± 0.91	17.00 ± 1.59
Proposed method	**4.67 ± 0.30**	**6.51 ± 0.30**

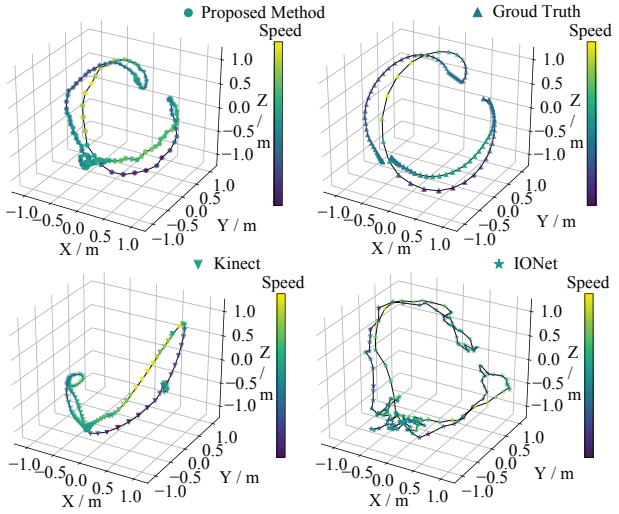

Figure 2: 3D swing trajectory comparison

4 CONCLUSION

We propose a swing-tracking CNN-LSTM model that fuses measurements from different sensor modalities and further addresses the challenges in single-sensor modalities, including occlusion, low sampling rates, and complex sensor noise models. The proposed method showing 62% of improvements in swing tracking compared to the state-of-the-art approach (IONet).

REFERENCES

[1] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv preprint arXiv:1812.08008* (2018).
[2] Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. 2018. IONet: Learning to cure the curse of drift in inertial odometry. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
[3] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).
[4] Natural Point. 2009. Inc.: Optitrack-optical motion tracking solutions.
[5] Sheng Shen, Mahanth Gowda, and Romit Roy Choudhury. 2018. Closing the gaps in inertial motion tracking. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 429–444.