A Low-Cost Virtual Coach for Diagnosis and Guidance in Baseball/Softball Batting Training

Hou-Chin Liu
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
Email: johnnyliu505@gmail.com

Chung-Ta King
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
Email: king@cs.nthu.edu.tw

***Abstract*—Baseball and softball are popular sports worldwide. In baseball/softball, batting is a fundamental action, but it is also one of the most difficult skills to master. Batting requires constant practice and proper guidance. Experienced coaches can provide instant diagnoses and feedbacks tailored for individual players. In the absence of such coaches, tools such as motion tracking systems or sports bracelets may help to assist the players. Unfortunately, they are either too expensive and awkward to use, or too limited in providing useful diagnosis and guidance. In this paper, we introduce a low-cost diagnosis and guidance tool for batting training in baseball/softball. The tool requires only one wearable device and one camera, such as the one on the smartphone, to capture the player's motion. The collected data are summarized and analyzed to derive the distinct features of the player's actions in different swing stages. The tool then pinpoints the mistakes and discrepancies in player's batting actions by comparing with expert actions, which in turn guides the player to perfect the action. Evaluations on real users show the effectiveness of our tool in comparison with experienced coaches.**

Keywords–Motion evaluation; Motion segmentation; Sport training; Wearable device; Semantic guidance.

I. INTRODUCTION

Baseball and softball are popular sports in the world. People not only watch the games but also play the games, from professionals to amateurs. In baseball/softball, batting is commonly accepted as one of the most difficult skills to master. Batting is not only swinging the bat. It requires the coordination of the whole body, from wrists, arms, waist, knees, to ankles, to concentrate the force on the bat to hit at the ball. It also involves the use of the muscle strengths at the right time.

To learn the skill, players need to practice constantly. During practice, an experienced coach or expert by the side can evaluate the batting actions and provide instant diagnoses and feedback tailored to the individual players, thereby shortening the learning curve. Unfortunately, such guidance is not always available. Many coaching tools are thus developed to assist the players. At one end, there are sophisticated motion tracking systems that use multiple surrounding cameras to capture detailed motions of a player for experts to analyze [1]–[4]. However, these tools are very expensive, mostly used in indoor and specially controlled environments and requiring specialists to operate and analyze. At the other end, wearable devices or smartphones are used for tracking players' physical and physiological status [5]–[7]. However, they are mainly purposed for data collection and can only give very crude information about the player's postures, not to mention to provide useful guidance to improve the skill.

In this paper, we introduce a low-cost coaching tool for baseball/softball batting training that can be used by amateur and novice players in the field and provide useful and immediate suggestions on improving the batting action, down to each stage of the action. Our tool not only evaluates the posture of the player but also the strength exercised by the player. To do so, the tool requires only one wearable device on the wrist to measure the strength and one low-cost camera, such as the one on the smartphones, by the side to capture the player's motion. Note that the proposed tool serves different purposes from those more expensive systems, which aim for professional player training.

Our tool has to perform two main tasks: (1) segmenting the recorded data to correspond to different stages of a batting action, and (2) identifying discrepancies of the player's action in each stage and providing suggestions to improve. There are challenging issues in each task. First, novice players may not perform the batting action right, making it very difficult to partition their motions into stages. The problem is aggravated by the low-cost cameras and very short duration of the batting action. For example, a batting action takes only a second or less, and a camera recording at 30 FPS (Frames Per Second) will have only a handful of frames per batting stage, increasing the difficulty in segmenting the video. Worse yet, low-cost cameras often drop frames. For the second task, the challenge lies in the fact that action evaluation is very subjective and the professional judgments of coaches are very difficult to quantize. How to extract coaches' experiences and program the tool to make similar judgment remains an issue.

To address the above challenges, we make several key observations. First, the impact point when the bat hits the ball has a very distinct feature that can easily be detected by a wearable device worn on the dominant wrist. The impact point can be leveraged to segment the batting action into stages. The second observation is that the batting motions of expert players are very similar and consistent. Therefore, we can develop a reference out of their batting actions. The reference is then used to assist segmenting the motions of novice players into stages and to evaluate the mistakes in the batting action of a player. The third observation is that sport coaches can often tell whether the player performs an action right or wrong, but

 ISBN: 978-1-61208-696-5

they only provide imprecise suggestions such as harder, wider, or higher. To translate these experiences and judgments into numbers that our tool can use to evaluate a batting action, we build a statistic model based on the evaluation results of real baseball coaches and set appropriate thresholds for common mistakes of players in batting.

The main contributions of this paper are as follows.

- We introduce a novel tool to diagnose bat swing motions and provide useful guidance for amateur baseball/softball players down to each stage of the action. The tool is low-cost and can be operated by ordinary users in the field for immediate feedback.
- The tool can properly segment the whole batting sequence into batting stages and extract corresponding motion features, even for novice players whose motions may be prone to errors and from low-quality videos taken by the low-cost camera.
- A statistic evaluation model is developed that reflects the judgment experiences of human coaches for evaluating the batting actions and providing improvement suggestions.
- Experiments by real players and coaches are conducted to show that the proposed tool can effectively detect motion mistakes and provide useful guidance to players comparable to the guidance provided by experienced coaches.

The remainder of the paper is organized as follows. In Section II, we discuss related works on motion assessment and bat swing motion analysis. Section III introduces the design and implementation of the system and the different phases in analyzing a batting action. Section IV presents the experimental setup and results. Section V concludes the paper and gives directions for future works.

II. Related Works

In motion assessment, people usually evaluate physiological and physical motion performances of the subjects. An intuitive approach is to analyzing their motion videos. Leightley et al. [8] propose a framework to automatically recognize and evaluate human motions using a depth camera. Patrona et al. [9] present a real-time framework for action detection, recognition and evaluation based on captured motion data. The outputs of the framework are semantic feedback generated by fuzzy logic. Parmar et al. [10] present multiple frameworks that use visual information for action quality assessment in evaluating and scoring Olympic sports. Qiao et al. [11] use the principle of gesture distance to develop a real-time 2D human gesture evaluation system.

The systems constructed by visual sensors are limited in sensing and evaluating detailed movements, which can be critical in practical motion training. Wearable sensors, on the other hand, are able to collect high-quality and fine physical and physiological data of specific parts of the body. In fact, for sports training, many motion assessment systems prefer wearable sensors. In [12], an ambulatory motion analysis framework is introduced that uses wearable inertial sensors to accurately assess an athlete's activities in an outdoor training environment. Sharma et al. [13] use a smart watch to capture and store inertial sensor data and develop a phase-based analytic system for tennis serving. The system can provide feedback for players to improve their serving performances. A wearable platform is presented in [14], which collects dominant body parts of the bat swing motion to provide baseball players with corrective feedback.

Hybrid approaches that combine wearable sensors and visual sensors to provide postural correction guidance and physical motion assessments have also been introduced. Kwon et al. [15] introduce a framework that combines wearable sensors and visual sensors for real-time motion training. They found out that the visual sensors are less effective for assessment feedback. Hirayama et al. [3] qualitatively compare batting motions using a motion capture system. Unfortunately, such motion capture systems are typically expensive and can only be used in the laboratory environments.

In the specific application domain of baseball batting, most motion assessment systems use wearable sensors to provide physical and physiological information [5]–[7]. In [14], baseball swing motions are evaluated using the motion transcripts to measure line segments and joints of the body. The swing motion quality is then assessed by comparing the intersegment coordination of a test swing to that of the template swing. In [16], Nakata et al. present detailed analyses of physiological status data for each phase in the swing motion between skilled and unskilled players. Other similar works, such as [17] and [18], propose different methods for segmenting bat swing motion and analyzing the stages to assess the motions.

Our work is different from previous studies. Our tool requires only one wearable device and one camera, such as the one on the smartphones, to provide not only assessments but also improvement suggestions for baseball/softball batting practice. The tool is low cost and can be operated by ordinary players in the field, unlike the expensive motion capture system.

III. System Design and Implementation

In this section, we describe the design and implementation of the proposed system. The wearable device is worn on the player's dominant hand to record the motions of the bat and the wrists. Any wearable device that contains an accelerometer and a gyroscope and provides a proper API to retrieve the collected data can be used. The camera is positioned in front of the player to record the full swing motion of the whole body in video. The height of the camera and its distance to the player can be normalized by preprocessing steps and thus may be flexible. However, the angle between the camera and the player may affect the evaluation accuracy and hence should be placed more carefully.

The clocks on the wearable device and the camera should be synchronized so that their recorded data can be timestamped and aligned along time. There is a host to synchronize the clocks and collect the recorded data for further processing. The host can be the smartphone that installs the camera and connects to the wearable device through Bluetooth. After the player performs the batting actions and the recorded data are collected, our system starts to analyze the data and provide suggestions to improve the actions.

Our system consists of three phases. The first phase is preprocessing, which transforms and normalizes collected raw data to make them consistent across different plays. In the second phase, the collected video and sensor data are processed to segment the batting motion into five stages in order to provide a stage-based analysis. Finally, in the third phase, features from the segmented data are extracted and a statistic evaluation model is applied to find out common mistakes in

 ISBN: 978-1-61208-696-5

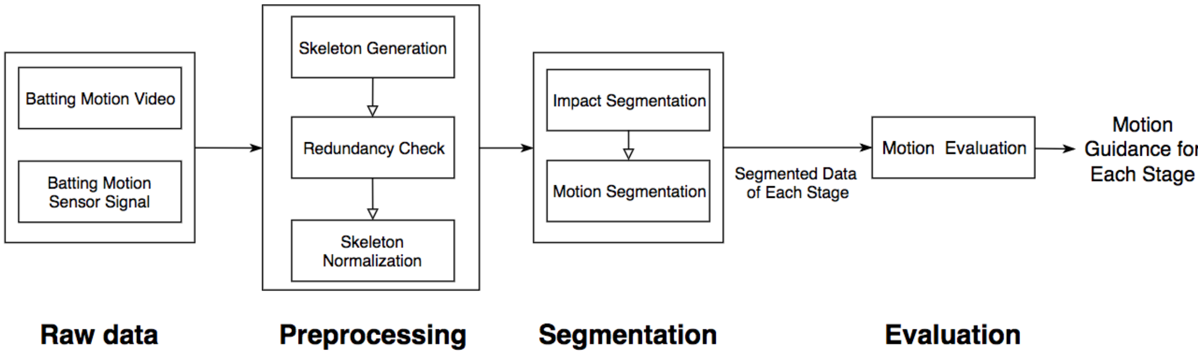


Figure 1. Overview of the proposed system.

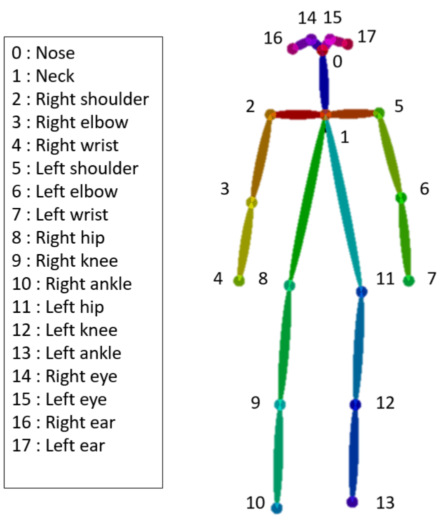


Figure 2. The 18 body joints generated from OpenPose.

the batting motion. Appropriate guidance and suggestions for improving the batting action in each stage are then provided according to the detected mistakes. The overview of the system is shown in Figure Figure 1 and the details of each phase are described next.

A. Preprocessing

The collected raw data need to be processed before they can be analyzed to segment the batting motion into stages. To extract postural information from the video, we first utilize the OpenPose library to generate body skeletons in each frame [19]–[21]. The library takes a color image as input and produces two-dimensional body, hand and facial keypoints for all the people in the image. From our single batting motion video, the OpenPose library can provide the positions of 18 body joints in each frame, including 2D skeleton joint coordinates and their corresponding confidence scores. Figure 2 shows the 18 body joints generated from OpenPose.

However, because the camera used to record the bating motion is low-cost and readily available on devices such as smartphones, some frames in the video may be dropped and adjacent frames are repeated to fill in the gap. To detect redundant frames, we compute the Mean Square Error (MSE) value of each frame and its previous one. If the MSE is lower than a threshold, we mark the frames as repetition and replace the redundant frame by the one that is interpolated based on the 18 body joints from the adjacent frames. The redundancy threshold is empirically set. A video containing four consecutive repeating frames is considered broken and will not be processed further.

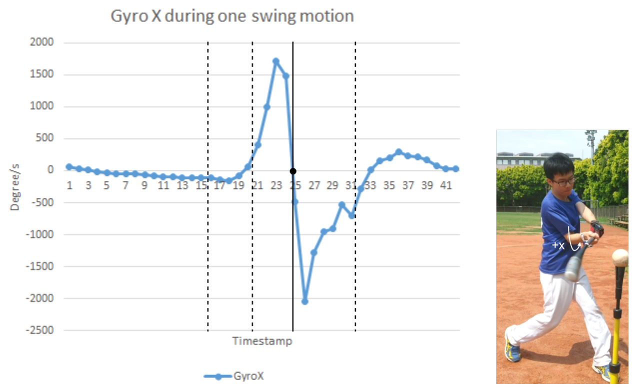
Finally, we need to handle the case in which the players have different heights and widths, and the camera may be positioned inconsistently across plays, e.g., the distance to the player. Our solution is to normalize the raw data. The original 2D positions of the 18 body joints are provided in a camera coordinate system. We first transform the camera coordinate system to body coordinate system with the origin at the neck joint. Next, we measure the distance from the neck joint to the line across the right ankle joint and left ankle joint. That distance is then used to scale all the body joints to obtain normalized body joint positions.

B. Segmentation

A baseball/softball batting action can generally be divided into several stages [16]–[18]: *waiting*, *shifting body weight*, *stepping*, *landing*, *swing*, *impact* and *follow through*. Since the waiting stage has very little effect on the batting action and different players have very different poses at the waiting stage, we thus do not consider this stage in our system.

As mentioned earlier, there are two challenges to meet in segmenting a batting action. First, the batting action lasts for a second or less, but we need to segment the action into six stages. On average, a stage contains only a few frames using a low-cost camera recording at 30 FPS. If we segment the video sequentially starting from the shifting-body-weight stage, errors can easily accumulate towards the last stage. Second, novice players may not perform the batting action right. Thus, their actions may lack of distinguishable stages and lead to incorrect segmentation and wrong analyses.

To address the first challenge, we leverage the obvious characteristic of the impact of the bat on the ball [22] to identify the impact stage. From there, we can divide the segmentation problem of the batting motion into segmentations before and after the impact stage. This dramatically improves the accuracy of motion segmentation. The second challenge is addressed by using the batting actions of expert players as reference. For expert players, their batting motions are very similar and consistent, which can be segmented by applying Hidden Markov Models (HMMs) on sensor readings [22]–[25]. The result then serves as a reference, by which the motions of



(a) (b)

Figure 3. (a) Variations of the GyroX value, (b) wrist rotation before the impact stage.

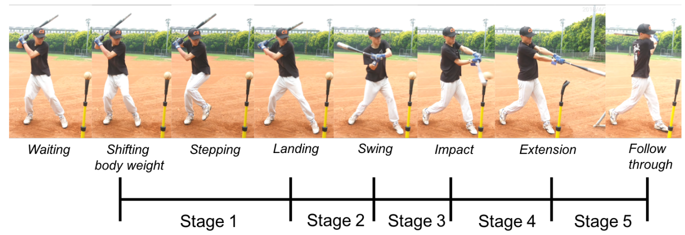


Figure 4. The five-stage batting motion based on HMM.

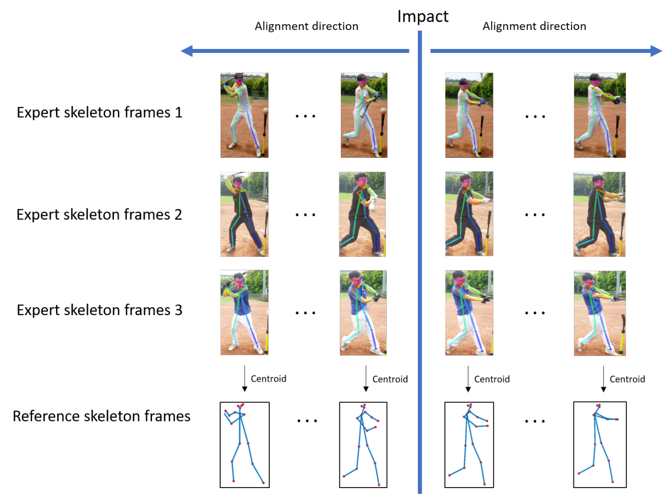


Figure 5. Reference skeleton frames generated from the videos of all expert players.

the novice players can be segmented through matching their characteristic features with those in the reference. Details are given in the following subsections.

1) Impact Stage Detection: The impact stage detection is based on the observation that when the bat hits the ball, there is a very fast twist of the wrist, which causes a drop in the angular acceleration of the X-axis (GyroX) towards zero on the wearable sensor. Figure 3(a) shows the variations of the X-axis values of the gyroscope around the impact point. The dot indicates the impact point. Figure 3(b) shows the rotational twist of the wrist just before the impact stage. With such a distinct feature, the sensor data and the corresponding skeletons of a batting action can thus be segmented by the impact stage.

2) Reference Segmentation from Expert Players: The batting motions of expert players are very similar and can serve to build a reference for segmenting the batting actions. Based on [22] for tennis serving, we first derive HMMs for the batting action from the wearable sensor readings. In fact, two HMMs are obtained, one before and another after the impact point. For the example gyroscope readings in Figure 3(a), the five stages identified by the HMMs are marked by vertical lines. The solid line indicates the impact point.

From the HMMs and timestamps of the wearable sensor data, we next partition the skeleton frames from the video also into five stages. Since the bat swing speeds of different expert players may be different, causing a small variation in the bat swing postures, the skeleton frames of expert motions need to be temporally aligned. To do it, we observe that the skeleton frames of expert players just before the impact point have almost the same postures. Thus, we align the skeleton frames before the impact point in a reverse order, starting with the last skeleton frame just before the impact point and working towards the beginning of the video.

The resulting segmentation of the video identified by the HMMs of the wearable sensor is shown in Figure 4. To compare with the five stages marked by a human, which are shown in the top half of the figure, we can see that Stage 1 in the first HMM covers the shifting-body-weight, stepping, and landing stages. Stage 2 overlaps with landing and swing, while Stage 3 covers swing and impact. The second HMM that uses sensor data after the impact point segments the video frames into two stages. Stage 4 extends from impact to extension stages and Stage 5 covers extension and follow through. Note that even though the batting stages identified by our tool are different from those identified by a human, what matters is a consistent reference segmentation that can be used to partition the motions of novice players and to detect their mistakes, which is discussed next.

To build reference segmentation using data from expert players, we first extract two features from the sensor readings of expert players: (1) the average time spent in each stage, and (2) the maximum strength value before and after the impact stage. It is found that these two features have very distinct values for expert and novice players. The time spent on each stage indicates the bat swing speed, and the maximum strength value indicates the effectiveness in concentrating the muscle forces. The strength value is determined by the X-axis acceleration, which is dominated by the bat swing strength.

Next, we build the *reference skeleton frames* of the five stages. The skeleton frames of all expert players with the same aligned timestamp are examined to calculate, for each skeleton joint, the centroid point of that joint in all those skeleton frames. The central points of all the 18 skeleton joints then produce a reference skeleton frame corresponding to that timestamp. Figure 5 shows the generated reference skeleton frames.

3) Novice Player Motion Segmentation: Novice players may not perform the batting action right. It is thus difficult to partition their batting motions using HMM due to the irregular sensor readings. Our solution is to find the most similar posture to match the expert reference skeleton frames.

Then, the corresponding frame is classified into the batting stage indicated by the reference skeleton frame. To match the skeleton postures, we reference the work in [26], which proposes trainable skeleton pose detectors to automatically learn a representation of skeleton poses by modeling the spatial arrangement of skeleton joints with respect to a reference point. The pose detectors are trained to learn actions formed by skeleton frames and are able to classify an action.

Given a prototype skeleton frame from our reference skeleton frames, the pose detector learns a model S to determine the position (x_i, y_i) of its skeleton joints, $j_i, i = 0...17$. Note that OpenPose outputs 18 body joints. The skeleton joints in the prototype frame can be described by (j_i, w_i), where w_i is the weight of that joint. Generally, the more a joint position varies, the more important it is. Therefore, the weight of a skeleton joint is determined by the variations of the joint position in the prototype skeletons.

Given a skeleton frame of a novice player, we want to find out the most similar frame from our reference skeleton frames. The measurement of skeleton similarity is a summary of the *similarity scores* of all the joints, which are calculated from the distance between the joint in the reference frame and that in the novice skeleton. The distance is weighted with a Gaussian function, which allows for spatial deformations. The score $r(t_i, j_i)$ is computed as:

$$r(t_i, j_i) = e^{-\frac{D(t_i, j_i)}{2\sigma_i^2}}, \tag{1}$$

where t_i is a joint in the novice skeleton, j_i is its corresponding joint in the reference skeleton, and $D(t_i, j_i)$ is computed by the Euclidean distance between positions of t_i and j_i. The σ_i is the standard deviation of the Gaussian weighting function for i-th joint. This value regulates the tolerance to the position of the i-th joint with respect to the position of its homologous. It is determined by the skeletal distance $\hat{d}(j_i, j_b)$ between the position of the reference point j_b and that of the reference skeleton joint j_i, where

$$\sigma_i = \sigma_0 + \alpha \cdot \hat{d}(j_i, j_b). \tag{2}$$

The value of σ_i increases with the skeletal distance of the i-th point from the reference point. The reference point j_b is denoted by the barycenter, which is determined by three joints, including neck, right hip and left hip. The principle of σ_i is that terminal joints have more mobility than those joints close to the body, and the value σ_i shows various tolerances in those joint positions. The distance $\hat{d}$ is the sum of line segments that connect the joints j_i and j_b. Notice that σ_0 and α regulate the tolerance values for deformation and are both tunable in the application phase. The weighting function shows the tolerance range in the position of skeleton joints and contributes robust deformations of the prototype skeleton.

After the similarity score of each joint is calculated, they can then be combined to determine the total skeleton similarity R as:

$$R(T, S) = \left(\prod_{i=1}^{|S|} r(t_i, j_i)^{w_i}\right)^{1/\sum_{i=1}^{|S|} w_i}, \tag{3}$$

where T denotes the novice skeleton and S denotes the reference skeleton.

Once the pose detectors are developed, they can then be applied to match similar frames in each stage. The pose detectors are computed with the reverse direction before the impact stage and the in-order direction after the impact stage. We have set the similarity threshold of skeleton detection to ensure the detection precision. The starting and ending frames are determined by the previous stage. After we identify proper skeleton frames for each stage for novice players, the sensor data for each stage can then be segmented by the identified frames. Figure 6 shows an example of frame matching on stage 3. The upper row shows the reference skeleton frames, and the lower row shows the skeleton frames of a novice player. The frames with the dotted line are the frames that are the most similar to the reference ones.

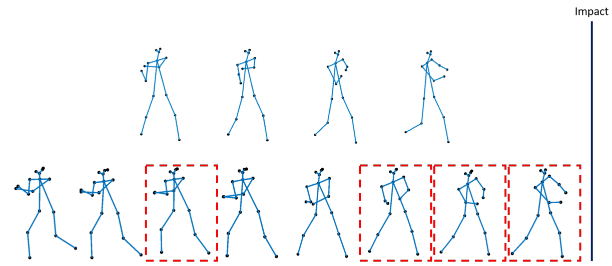


Figure 6. Frame matching result of stage 3.

TABLE I. COMMON MISTAKES OF BATTING ACTIONS AND CORRESPONDING FEATURES

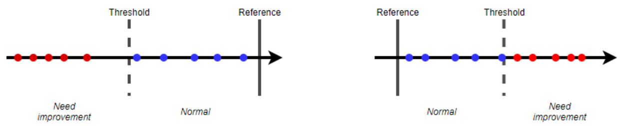
Common mistakes	Features
Center of body weight moved	Center of body weight
Right arm turned straight too early	Right arm angle
Left arm turned straight too early	Left arm angle
Right leg not bent	Right leg angle
Left leg not straight	Left leg angle
Swing too slow	Time spent on stages 2 and 3
Not enough swing strength	AccX of stage 3
Not enough extension strength	AccX of stage 4

C. Batting Action Evaluation

After the batting motion of the player is segmented into stages, we need to evaluate the actions in each stage for possible mistakes and for providing suggestions for improving the actions. As mentioned earlier, the problem is challenging because the judgments of human coaches are very subjective and hard to quantized. Our solution is to collect statistics of the judgments of human coaches to establish thresholds for tolerating the deviations of the player's actions from the reference skeletons. Since stage 1 has more personalized features and stage 5 is the finishing motion after the extension, both are less helpful when assessing a batting motion. Therefore, we focus on the remaining three stages.

By consulting experienced baseball/softball coaches, we first establish a list of common mistakes in batting action, as shown in Table I. The table also shows the features that can be used for detecting the mistakes. The mistakes are stated from the perspective of a coach and are thus descriptive and lacking precise quantitative definitions.

To derive meaningful measures for detecting mistakes in batting actions, we propose to use assistance from human coaches and the reference segmentation introduced above. The idea is to ask coaches to view the videos of batting actions and label the parts that the players perform incorrectly according to Table I. The skeleton features of the labeled frames are then extracted to be compared with those from the reference



Threshold type	Threshold value < Reference value	Threshold value > Reference value
Feature	Left leg angle, Center of body weight, AccX of Stage 3, AccX of Stage 4	Right/Left arm angle, Right leg angle, Center of body weight, Time spent on stage 2 and stage 3

Figure 7. Types of threshold values illustrated.

Figure 8. Evaluation results of a batting action.

skeleton frames. Distribution of the differences for each feature is then examined to determine a threshold value for tolerance. The resultant threshold values for all the features in Table I constitute a *statistic evaluation model* in our tool.

The threshold values can be classified into two types: one defines the upper bounds and the other defines the lower bounds, as shown in Figure 7. For example, according to Table I, the left leg should be as straight as possible. Hence, its value should not exceed that of the reference feature, and the threshold value should bound that feature value from above.

The features extracted from the player's motion data are compared with the reference values based on the thresholds to give a correct/incorrect mark. The sensor data of the player are also compared with the reference sensor data in terms of two features: time spent on each stage and the maximum strength value before and after the impact stage. Next, frame-level marks of the same stage are combined by majority voting to form mistake vectors at the stage level. These vectors are then used to determine the motion correction guidance to provide at each stage.

Correction guidance is actually the opposite side of motion mistakes. Once a motion mistake is identified, the corresponding correction guidance can be fed back immediately. Figure 8 shows an example of a batting action and the correction guidance provided by our system.

IV. EXPERIMENT

A. Experimental Setup

One wrist-worn device and one camera on a mobile phone (HTC A9) are used to record a player's batting action. The wearable device, NuMaker TRIO, is a low-cost wireless device that consists of a 32-bit low-power microcontroller, a 6-axis Microelectromechanical sensor, a wifi and a Bluetooth module (BT3.0 and BLE). To measure the physical status of a batting motion, the 6-axis Microelectromechanical sensor is used to collect sensor readings of 3-axis acceleration and angular velocity. The data are then sent back to a computer through Bluetooth. The sampling rate of the wearable device is 30Hz, and its measuring ranges are ± 16g for the accelerometer and ± 2500 degree/s for the gyroscope. The camera on the HTC A9 is used to record the whole batting action in video with a frame rate of 30 FPS, which is the same as the sampling rate of the wearable device. We also put timestamps in the video for time synchronization with the wearable device later.

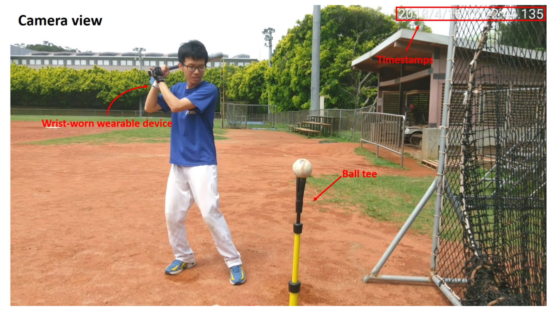


Figure 9. Experimental setup for motion data collection.

We have collected a set of batting motion data for training and evaluating the proposed system (see Figure 9). Twelve players were invited in the data collection. They were asked to wear the wearable device and perform batting motion in front of the camera. Four of them were experienced players, who were on the softball team of our department. They had played softball for about 4 to 6 years and were all sluggers on the team. The other three of the twelve players had baseball or softball experiences, but they seldom played the sport. The remaining five players had no prior experience. We had collected 1139 batting actions, including 673 from experienced players and 466 from the other players. After removing broken files, the remaining data count was about 950.

B. Experimental Results

The four experienced players are considered expert, and their batting motions are used to generate the reference motion and build the statistic evaluation model. The hitting coach, who served on the softball team, was asked to label the videos of the other eight players for possible mistakes according to Table I. The labeling results are denoted as the ground truth, which can be compared against the results from our tool. The *recall*, *precision*, and *F-measure* (F1 score) metrics are used for comparison, where:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (4)$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (5)$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

The recall metric calculates the proportion of actual positives that are identified correctly, and the precision metric calculates the proportion of positive identifications that are actually correct. To combine these two metrics, we apply the F1 score to compute the harmonic average of the recall and precision. These metrics are very useful in verifying the ability of model classification.

 ISBN: 978-1-61208-696-5

TABLE II. COMMON MISTAKES OF BATTING ACTIONS AND CORRESPONDING INDICES

Index	Common Mistakes
m1	Left leg not straight
m2	Right leg not bended
m3	Center of body weight moved backward
m4	Center of body weight too high
m5	Right arm turned straight too early
m6	Left arm turned straight too early
m7	Swing speed too slow
m8	Not enough swing strength
m9	Not enough extension strength

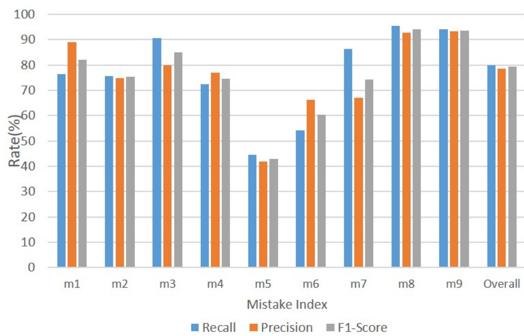


Figure 10. Detection rate for each common motion mistake.

The collected batting actions are separated into two parts, 80% for training and 20% for testing. To train the proposed system, we use 5-fold cross-validation to have optimal effects. The system results are used to tune the threshold parameters of the statistic evaluation model. We use majority voting for combining stage-level judgments from our system into a final mark for the whole batting action in order to compare with the ground truth for error detection accuracy.

For evaluate purposes, we consider the common mistakes listed in Table II and assign an index to each. We removed two mistakes in Table I due to insufficient samples (less than 10).

The detection rate for each common motion mistake is shown in Figure 10. This experiment shows the detection abilities of our system for all common mistakes. We can see that most of the detection rates are higher than 70%. However, for m5 and m6, the detection accuracy is lower than the others. The F1 score of m5 is 42.86%, while the F1 score of m6 is 60.38%. The features of m5 and m6 seem to be the key factor. Without depth information from images, the arm angle is hard to measure correctly. Even if the arm keeps the same angle during the batting motion, the angle measurement might still change because of different facing directions in a 2D video. This makes angle measurement imprecise. Although the detection rates of m5 and m6 are lower, the overall F1 score is still 79.25%, which suffices to show the effectiveness of our proposed system in detecting common mistakes.

Next, we evaluate motion mistake detection results of each player (see Figure 11). Players p1, p2, and p3 have played baseball/softball but without much experience. They are classified as mid-level players. The other players are new to the sport and are classified as novice players. The experiment is to verify how well the proposed system works for different types of players.

From Figure 11, we can see that the proposed system is able to detect motion mistakes of various players. However, the F1 score of mid-level players seems to be lower than the others. By closely examining the detection rate of common mistakes for each type of players, shown in Table III, we can see that the precision rate of mid-level players is only 52.67%, which is lower than the novice players. The F1 score of novice players is also higher. It seems that our proposed system works better on the novice players rather than mid-level ones. This is expected, because mid-level players have already had some experiences and their motions should be more like experts'. Novice players usually perform more obvious mistakes in a bat swing motion.

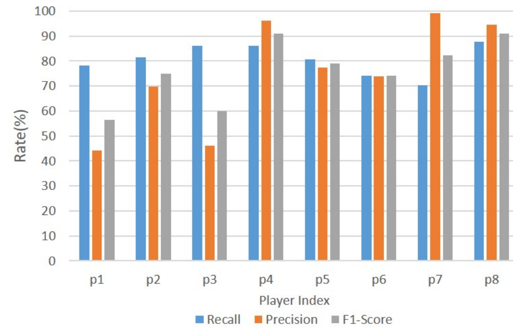


Figure 11. Detection rate of motion mistakes for each player.

TABLE III. DETECTION RATE OF MOTION MISTAKES FOR DIFFERENT TYPES OF PLAYERS

Player Type	Recall(%)	Precision(%)	F1-Score(%)
Mid-level	85.07	52.67	65.02
Novice	77.96	85.94	81.74

Next, we evaluate the effects of using the impact point to divide a batting action into stages. Recall that our tool first detects the impact point with the wearable sensor, based on which the whole sequence of batting motions is divided into stages (see Figure 3). Table IV shows the overall performance results with and without impact point detection. We can see that the overall F1 score increases from 71.43% to 79.25% with impact point detection. This clearly shows its effectiveness in segmenting the batting motions.

Finally, we evaluate the effectiveness of reverse alignment of sensor data with video frames before the impact point. The intuitive way for alignment is sequential. Our tool reverses the alignment direction before the impact stage. From Table V, we can see that the overall F1 score increases from 73.20% to 79.25% using reverse alignment. Other metrics also show improvements, which suffice to demonstrate the effectiveness of reverse alignment on system performance.

To summarize these experiments, we can see that our proposed system can effectively detect common mistakes in baseball/softball batting actions and provide proper guidance for the players to improve. The system is especially suitable for novice players. Note that we did not compare our tool with

TABLE IV. EVALATION EXPERIMENTS FOR IMPACT POINT DETECTION

Method	Recall(%)	Precision(%)	F1-Score(%)
w/o impact segmentation	75.64	67.69	71.43
w/ impact segmentation	79.97	78.60	79.25

TABLE V. EVALUATION EXPERIMENTS FOR REVERSE ALIGNMENT

Method	Recall(%)	Precision(%)	F1-Score(%)
w/o reverse alignment	74.45	72.05	73.20
w/ reverse alignment	79.97	78.60	79.25

professional systems because they serve different purposes as ours and thus have different requirements in terms of error detection and guidance provided. Due to time and budget limitation, we could not experiment with more subjects in this paper and we only had one coach to label the batting videos. In the future, we hope that this discrepancy can be remedied. Generalization of the proposed tool requires further studies. After all, baseball/softball batting is a very specific motion that the player is fixed in location. It is thus easy to capture the motion with one camera.

V. CONCLUSION

In this paper, we introduce a diagnosis and guidance system for baseball and softball batting motions. Our tool extracts skeletal information from player's motion video and segments the motion data of sensor and skeleton into five swing stages. The segmented data is evaluated by a statistic evaluation model. By evaluating the results of each stage, we then detect common mistakes in batting and provide proper guidance to the player. The experiments show that our proposed system has about 80% accuracy in detecting common batting mistakes and can provide proper guidance to players. In the future, we would like to cooperate with multiple coaches and use majority voting in labeling the batting videos to provide a more robust motion judgment. Batting action evaluation based on machine learning techniques may be exploited to see whether they can provide judgments and guidance that are comparable to those provided by professional coaches. To overcome the problem of angle measurement, two or three cameras might be needed to provide the depth information for measuring the motions with a 3D view. To provide a more robust guidance for players of various types, more advanced algorithms of computer vision would be very helpful. Furthermore, generalization of the proposed tool could be explored.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 107-2218-E-007-035 and by Information and Communications Research Laboratories of Industrial Technology Research Institute, Taiwan.

REFERENCES

[1] D. Dharmayanti, M. Iqbal, A. Suhendra, and A. B. Mutiara, "Velocity and acceleration analysis from kinematics linear punch using optical motion capture," in Proc. Second International Conf. on Informatics and Computing (ICIC), Nov. 2017, pp. 1–6.

[2] B. Dowling and G. S. Fleisig, "Kinematic comparison of baseball batting off of a tee among various competition levels," Sports Biomechanics, vol. 15, no. 3, Sept. 2016, pp. 255–269.

[3] D. Hirayama, K. Yoshizawa, H. Sogo, and T. Henmi, "Quantitative comparison of technical differences in baseball batting motion by motion analysis," in Proc. International Conf. on Advanced Mechatronic Systems (ICAMechS), Nov. 2016, pp. 115–120.

[4] K. Kolykhalova, A. Camurri, G. Volpe, M. Sanguineti, E. Puppo, and R. Niewiadomski, "A multimodal dataset for the analysis of movement qualities in karate martial art," in Proc. 7th International Conf. on Intelligent Technologies for Interactive Entertainment (INTETAIN), June 2015, pp. 74–78.

[5] http://www.zepp.com/en-us/baseball/. [Last retrieved: Jan. 2019].

[6] https://buy.garmin.com/en-US/US/p/579018. [Last retrieved: Jan. 2019].

[7] https://smashprosports.com/. [Last retrieved: Jan. 2019].

[8] D. Leightley, J. S. McPhee, and M. H. Yap, "Automated analysis and quantification of human mobility using a depth sensor," IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 4, July 2017, pp. 939–948.

[9] F. Patrona, A. Chatzitofis, D. Zarpalas, and P. Daras, "Motion analysis: Action detection, recognition and evaluation based on motion capture data," Pattern Recognition, vol. 76, April 2018, pp. 612–622.

[10] P. Parmar and B. T. Morris, "Learning to score olympic events," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), July 2017, pp. 76–84.

[11] S. Qiao, Y. Wang, and J. Li, "Real-time human gesture grading based on openpose," in Proc. 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Oct. 2017, pp. 1–6.

[12] A. Ahmadi, E. Mitchell, F. Destelle, M. Gowing, N. O'Connor, C. Richter, and K. Moran, "Automatic activity classification and movement assessment during a sports training session using wearable inertial sensors," in Proc. 11th International Conf. on Wearable and Implantable Body Sensor Networks, June 2014, pp. 98–103.

[13] M. Sharma, R. Srivastava, A. Anand, D. Prakash, and L. Kaligounder, "Wearable motion sensor based phasic analysis of tennis serve for performance feedback," in Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP), March 2017, pp. 5945–5949.

[14] H. Ghasemzadeh and R. Jafari, "Coordination analysis of human movements with body sensor networks: A signal processing model to evaluate baseball swings," IEEE Sensors Journal, vol. 11, no. 3, March 2011, pp. 603–610.

[15] D. Y. Kwon and M. Gross, "Combining body sensors and visual sensors for motion training," in Proc. of ACM SIGCHI International Conf. on Advances in Computer Entertainment Technology (ACE), 2005, pp. 94–101.

[16] H. Nakata, A. Miura, M. Yoshie, and K. Kudo, "Electromyographic activity of lower limbs to stop baseball batting." Journal of strength and conditioning research, vol. 26, no. 6, 2012, pp. 1461–1468.

[17] R. Gray, "A model of motor inhibition for a complex skill: Baseball batting," Journal of experimental psychology: Applied, vol. 15, no. 2, 2009, pp. 91–105.

[18] E. S. Chang, M. E. Bishop, D. K. Baker, and R. V. West, "Interval throwing and hitting programs in baseball: Biomechanics and rehabilitation." American journal of orthopedics, vol. 45, no. 3, 2016, pp. 157–162.

[19] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1302–1310.

[20] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4645–4653.

[21] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4724–4732.

[22] D. Yang, J. Tang, Y. Huang, C. Xu, J. Li, L. Hu, J. Zhang, G. Shen, M. Liang, and H. Liu, "Tennismaster: An imu-based online serve performance evaluation system," in Proc. of 8th ACM International Conf. on Augmented Human (AH), 2017, pp. 1–8.

[23] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, Feb. 1989, pp. 257–286.

[24] L. R. Rabiner and B. H. Juang, "An introduction to hidden markov models," IEEE ASSP Magazine, vol. 3, 1986, pp. 4–16.

[25] K. Liu, C. Chen, R. Jafari, and N. Kehtarnavaz, "Fusion of inertial and depth sensor data for robust hand gesture recognition," IEEE Sensors Journal, vol. 14, no. 6, June 2014, pp. 1898–1903.

[26] A. Saggese, N. Strisciuglio, M. Vento, and N. Petkov, "Action recognition by learning pose representations," in Workshop of Recognition and Action for Scene Understanding (CAIP Conf. 2017), 2017.