AICO, Artificial Intelligent COach

Vicent Sanz Marco
Osaka University
1-32 Toyonaka city,
560-0043, Osaka,
vicent.sanzmarco@aist.go.jp

Norimichi Ukita
Toyota Technological Institute
2-12-1 Nagoya,
468-8511
Japan
ukita@toyota-ti.ac.jp

Natsuko Kobayashi
Osaka University
1-32 Toyonaka city,
560-0043, Osaka,
kobayashi-nat@cmc.osaka-u.ac.jp

Morito Matsuoka
Osaka University
1-32 Toyonaka city,
560-0043, Osaka,
matsuoka@cmc.osaka-u.ac.jp

ABSTRACT

Choosing effective strategies before playing against an opponent team is a laborious task and one of the main challenges that American football coaches have to cope with. For this reason, we have developed an artificial intelligent American football coach (AICO), a novel system which helps coaches to decide the best defensive strategies to be used against an opponent. Similar to coaches who prepare a winning game plan based on their vast experience and previously obtained opponents' statistics, AICO uses power of machine learning and video analysis. Tracking every player of the last recorded matches of the opponent team, AICO learns the strategies used by them and then calculates how successfully their own defensive strategies will perform against them. We have used 7350 videos in our experiments obtaining that AICO can recognize the opponent's strategies with about 93% accuracy and provides the successful rate of each strategy to be used against them with 94% accuracy.

Keywords

Computer Vision, image and Video Processing, Pattern Recognition, Sport, AI

1 INTRODUCTION

Nowadays, the advancement of neural network applications has been a useful tool to develop automatic analysis systems to help with sport analysis, being an active research topic in the last years [Jia16, Fre19]. In this research, we have created an artificial American football coach to help the coaches to determine the best strategies to be used against opponent teams. In order to assist coaches with this laborious task, we have created AICO, a novel video analysis recognition system which works together with machine learning.

Recognition and tracking players in the field is made by video analysis, which determines the strategies used by both teams, offensive and defensive ones. To the best of our knowledge, other research related to American football is only focused on professional teams, not being easily accessible by other kind of teams, such as small teams or amateurs. There are studies to detect players in the field [Rie13, Dir18], track player's movement [Yam13], recognize offensive strategies [Atm13, Sid09].However, all these investigations required professional resources, such as, broadcast videos obtained from TV [Ste17] or the utilization of specialized and expensive devices to track players [Bur17].

For this reason, we have developed AICO, an inexpensive and versatile system that can be used everywhere. AICO is an artificial recognition system which can be used by any type of American football team, since it does not require any previous device installation on players, stadium or field.

American football coaches spend a considerable time studying American football videos from opponent teams, trying to discern strengths and weakness and use them for their own benefit. AICO is a novel automatic system developed to help coaches in that matter. AICO gathers the opponents movements and strategies in order to discover the best strategies to be used against that team, lightening the amount of work made by the coaches involved in the analysis of the opponent videos.

On the other hand, since during our research we could not find any available American football video dataset containing match plays, we have created an American football video dataset containing about 7350 plays of different teams, making possible the use of this dataset in further video recognition comparisons in future projects.

2 AMERICAN FOOTBALL BACKGROUND

In American football, there are two teams with 11 players who compete in four quarters of 15 minutes. In every play, one of the teams is defined as the offensive team, the one in possession of the ball, while the other team is defined as the defensive team.

According to American football regulations [Goo18], the offensive team needs to move the ball forward at least 10 yards. This is why the field has clearly marked yardage lines on it. The offensive team has 4 attempts, called *downs* in American football, to either score or gain 10 or more yards. If the ball is moved that far, the count resets, and the team earns another set of four downs to try go a further 10 yards. If the offensive team does not reach 10 or more yards in the 4 downs, the defensive team gains the possession of the ball and changes its role to offensive team. In this paper, a *play* is defined as a down of the offensive team.

To know if a strategy was successful or not, we are using the American football analytic made by Football Outsider [Out19] which focuses on advanced statistical analysis of the NFL and is run by professional sport journalists. An offensive strategy is defined as *successful* by Football Outsiders if it gains at least 40% of the yards-to-go on the first down, 60% of yards-to-go on the second down and 100% of yards-to-go on third or fourth down. Otherwise, the strategy is defined as *unsuccessful*.

American football coaches have a *playbook* containing all the offensive and defensive strategies their team can play during a match. The offensive strategies are separated in *passing strategies* and *running strategies*. In *passing strategies*, the quarterback attempts to pass the ball to another player. On other hand, in *running strategies*, the quarterback runs with the ball or gives the ball to a closest player that makes the run. A team has numerous passing and running strategies in their playbook. The defensive strategies are divided depending on the amount of players used in the front line. For our experiments, the playbook used has about 100 offensive strategies (60 passing and 40 running) and 100 defensive strategies.

To determine the offensive strategies to be used against an opponent team, coaches choose the best strategies players performed in the previous games or training sessions. However, for selecting defensive strategies is a different story.

In sports, team coaches analyze the opponents teams in order to find the best strategy to play against them. American football is not an exception. Thanks to our collaboration with *Kosei Gakuen High School* American football (KSS Lotus) team, we verified that coaches visualize the last matches of the next rival, checking every offensive play used by the opponent team. Figure 1

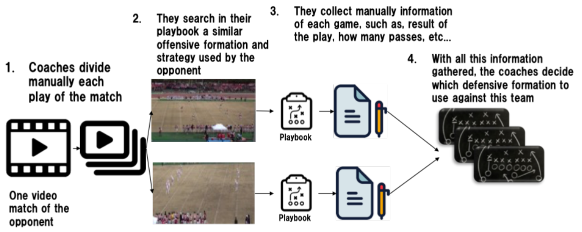

Figure 1: Flowchart showing how the coaches decide their defensive strategies against an opponent

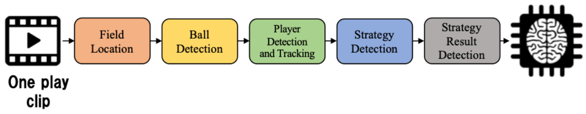

Figure 2: Preprocessing data to train AICO from a play clip

shows a flowchart describing how coaches gather manually the information about the opponent team. In every play, coaches compare their offensive strategies in the playbook to the opponent strategy used, trying to match it with the most similar one in their playbook. They also collect other relevant information, such as, how many passes failed or who was the most relevant player of the other team. After having collected all the information from the opponent team, coaches sit together and discuss the best defensive formations and strategies to be used against that team. This work is manually done by coaches, however, in this paper, we propose an alternative system, AICO. A novel video analysis system together with neural networks that can automatically gather the relevant information of the opponent team and determines the best defensive strategies to be used against that team.

3 ALGORITHM

One of the two main goals pursued by AICO is the creation of an automatic video analysis system. The flowchart showed in Figure 1 represents how coaches examine a specific opponent team using the opponent match recordings. This procedure made by the coaches is repeated for every opponent team. Similar to the way coaches get information about their opponents, the video analysis system implemented is responsible to obtain the data from the recordings of the last opponent team matches. A strategy recognition algorithm has been developed to analyze the last video matches of the opponent team and detect strategies used in every play. The algorithm detects the offensive strategies of the team chosen as opponent, and also calculates how successful were the defensive strategies of the other team against them. From now on, this strategy recognition process is going to be called *preprocessing step*, whose overview chart can be found in Figure 2.

The second main goal of AICO uses the results obtained in the *preprocessing step* to learn the strategies of a specific adversary and provide the most reliable defensive strategies to be used against this opponent team, calculating the percentage of success of every strategy provided by the coaches. As consequence, AICO is trained independently for every opponent, obtaining different AICOs for every team analyzed.

3.1 Field Location

AICO currently works only using one single-camera and there are no restrictions on the camera used to record the match as long as it is in a fixed position with a tripod and has at least a quality of 720p.

After receiving the full match of an opponent team, AICO uses an improved Chen's algorithm [Che14], with 92% of detection accuracy, in order to recognize different sequences of plays. We have modified Chen's algorithm to detect the plays in any kind field, due to the algorithm only work if the field has grass, but this is not always the case. In the labeling step, AICO classifies every play depending on the actions the teams are performing. AICO only keeps offensive and defensive plays (downs), which are labeled as *play clips*, discarding other kind of plays such as, field goal plays and extra point plays.

For each *play clip*, AICO performs Direct Linear Transform (DLT) [Har03] to detect what part of the field the camera is recording. DLT algorithm is used to resolve the homography matrix H between the first frame of the *play clip* and a digital American football field model. From now on, this digital American football field model will be referred to as the *football model*. Since the system works in homogeneous coordinates, a point (x,y) from the real field and a point (x',y') from the *football model* can be expressed as:

$$c \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

where c is any non-zero constant, and

$$H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \tag{2}$$

To resolve Equation 1, the first row is divided by the third row:

$$-h_1x - h_2y - h_3 + (h_7x + h_8y + h_9)x' = 0 \tag{3}$$

and the second row is divided by the third row:

$$-h_4x - h_5y - h_6 + (h_7x + h_8y + h_9)y' = 0 \tag{4}$$

As consequence, Equation (3) and (4) can be expressed in a matrix form:

$$A_i h = 0 \tag{5}$$

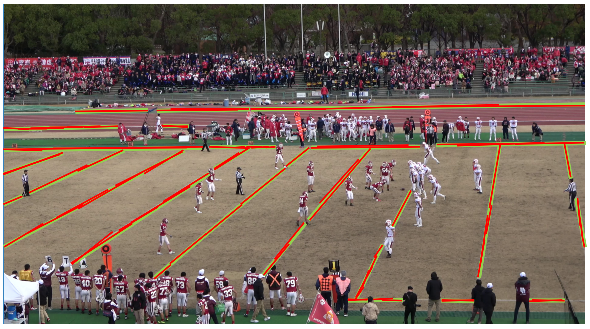

Figure 3: Hough transform lines are represented in red while RANSAC lines are in green

where $A_i = \begin{bmatrix} -x & -y & -1 & 0 & 0 & 0 & x'x & x'y & x' \\ 0 & 0 & 0 & -x & -y & -1 & y'x & y'y & y' \end{bmatrix}$ and $h = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 \end{bmatrix}^T$.

Since each correspondent point in the first frame (x,y) and its relative (x',y') in the *football model* provides 2 equations, 4 correspondent points are enough to calculate H [Ela08]. AICO requests the user to provide 2 matching reference points for the first frame and other 2 for the *football model*.

AICO uses the homography matrix H to localize the field boundaries. Once it is obtained, AICO will only focus on the players located inside the field, not considering anything or anyone out of it.

When all the *play clips* sent are obtained from the same point of view, the homography calibration only needs to be made once and can be used by all of them. Only when a *play clip* is sent to AICO from a different perspective, it will be necessary to introduce the 4 matching reference points to calculate H again.

To resolve this issue, AICO has an automatic algorithm to detect if the *play clip* has a different perspective from the previous one. This algorithm uses Hough transform [Bal87] to detect the lines in the first frame. After these lines are obtained, the algorithm uses RANSAC [Chu03] to join lines that are detecting the same line. AICO stores these RANSAC lines together with the 4 matching reference points.

Figure 3 shows an example of the lines obtained from Hough transform (red lines) and the RANSAC lines (green lines) obtained. AICO will store these RANSAC lines or used them to compare with the RANSAC lines of the previous *play clip*.

When a new *play clip* is analyzed, AICO compares all RANSAC lines of the first frame of this *play clip* to the all RANSAC lines obtained from the previous *play clip*. When the location of the RANSAC lines of both play clips coincide, with a difference of less than 10%, AICO assumes that the location of the camera is the same in both *play clips*. As consequence, the 4 reference points of the previous *play clip* are used for this new *play clip*, and the user does not need to introduce them again.

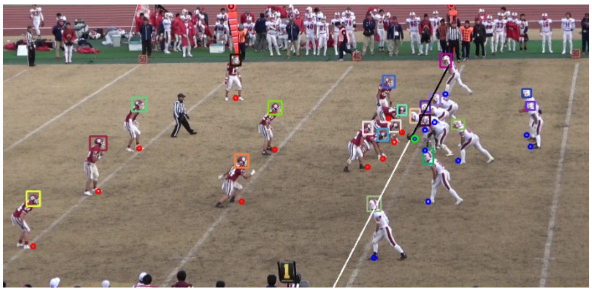

Figure 4: Example of ball line detection and player detection

3.2 Ball Detection

The American football regulation establishes that, at the beginning of every play, the ball has to be on the ground, and only the *Center* player from the offensive team is allowed to touch the ball with one hand. This *Center* player is the one who passes the ball to the quarterback at the beginning of each play. Team strategy performances start right after the ball starts moving, so this is the crucial moment for tracking players and determining the strategies of both teams.

As consequence, to determine when the play starts, it is necessary to track the ball and detect when the ball is moved for first time. When the ball is moved, AICO starts tracking players in order to determine the offensive and defensive strategies used in that *play clip*.

AICO uses the Convolutional Neural Network (CNN) Inception_v2 [Jof15] in the first frame of each *play clip* to locate the ball. This CNN has been trained to detect only American football balls.

Once the ball is detected, it is tracked via Sparse Collaborative appearance Model (SCM) [Zho14]. This tracker algorithm has been demonstrated to be one of the best state-of-the-art tracker models [Wy015]. AICO uses Equation (1), where (x,y) are the ball tracker coordinates, and H is the homography matrix calculated in the field location step, to locate the ball in the *football model*.

When the *Center* player passes the ball to the quarterback, the ball is not only moved in the real world, but it is also moved in the *football model*. We have defined that if the ball is moved 1 yard or more in the *football model*, then the tracker following the ball is deleted and the frame is frozen. This frame is used to detect all players in the field in the player detection step.

Additionally, AICO draws an artificial line that crosses the initial ball position and goes parallel to the closest yard line in the field. This line, called *ball line*, is used by the player detection step to determine if a player is part of the offensive or defensive team. Figure 4 shows an example of the ball line detection. AICO only uses one line, however it has been represented by two lines (black and white) in the figure to make it easy to see the ball position.

3.3 Player Detection and Tracking

After the ball has been moved to the quarterback, AICO freezes the frame and uses the CNN Inception_v4 [Sze17] to detect every player in the field. This neural network has been trained to detect only American football players. So referees are not included in the detection.

Inception_v4 frames all the players detected in the field. Once players are detected, AICO treats each player's frame separately. We have seen in our experiments that tracking the whole body of the players results in lost tracking of the player assigned. However, if the helmet is tracked instead of the body, the tracking accuracy obtained is very high (around 95%). For this reason, another CNN (Inception_v2) is performed on each player's frame. This inception_v2 has been trained to detect the players' helmet in each player's frame.

The helmet detected of the player is assigned to the Siamrpn++ [Li19] tracker. As consequence, each player has each on personal Siamrpn++ tracker.

However, the position calculated in the *football model* using the helmet position does not allow AICO to localize the correct position of each player in the *football model*. To obtain the correct position of the player in the *football model*, it is necessary to use their feet. Thus, for every helmet tracker, AICO also established a *location point*. This *location point* is the lower point of the line that goes from the middle of the helmet tracker to the bottom of the player's frame. The *location point* follows the helmet tracker, so, wherever the helmet tracker moves, the location point is always defined with the tracker. As consequence, the correct position of the player in the *football model* is determined using the *location point* together with the homography matrix H calculated in the field location step (Equation (1)).

Figure 5 shows an example of the player detection procedure. In this example, Inception_v4 detects the player and creates the green frame. Afterwards, Inception_v2 recognizes the helmet and defines the blue frame. It is possible to create a vertical line (yellow line) between the middle width point of the helmet frame and the player's frame bottom. The intersection between this yellow line and the player's frame bottom is represented as a red dot. This red dot is the *location point* of that player.

In addition, AICO can automatically determine if the player is in the offensive or defensive team. Once every player in the field is located at the beginning of the play, AICO searches the *Center* player. This player is the closest player to the ball and it is always part of the offensive team. In Figure 4, the *location point* of the *Center* player is marked by a green circle.

After having located the *Center* player, AICO uses the *ball line* to determine if the offensive team is on the

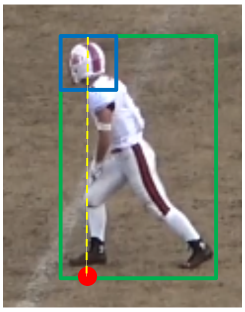

Figure 5: Player detection method

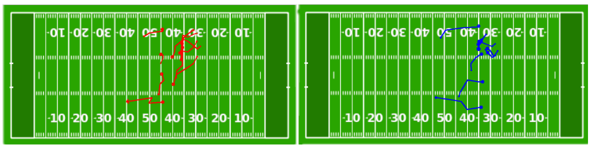

Figure 6: Tracking obtained after player detection step is finished. Defensive tracking on the left and offensive tracking on the right

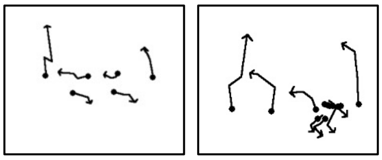

Figure 7: Converted images of the offensive (right) and defensive (left) tracking used to determine the strategies of both teams

right or left side of the ball. Consequently, using the *ball line* and the location player, AICO can define if players are part of the defensive or offensive team.

Once all players are detected and tracked, AICO creates two images: one has the tracking of the offensive players and the other contains the tracking of the defensive players. Figure 6 shows an example of these two images. As it is shown in Figure 7, these two images are sent to an application that uses OpenCV [Kae16] to clean, rotate and convert them in images that can be used to determine the strategy used by both teams in the strategy detection step.

Defensive strategies in coach playbooks do not include players in the front line, since their main responsibility is to stop the other team offensive front line and execute always the same movement. Therefore, the OpenCV application removes the defensive front line players in the defensive converted image.

3.4 Strategy Detection

AICO compares the offensive and defensive strategies obtained in player detection step to the coach playbook, looking for the most similar strategies contained in the playbook. To compare the images AICO uses OpenCV.

For this comparison, the strategies in the playbook are converted in images and then compared to the converted images obtained in the player detection step. AICO uses Optical Character Recognition(OCR) to detect the text contained in the playbook images.

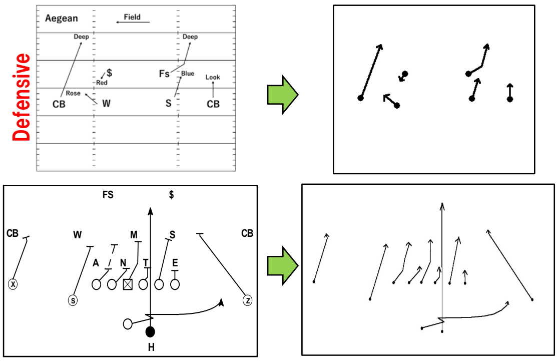

Figure 8: Example of strategy image conversion

Figure 8 shows the strategy conversion using the OpenCV application (right) from the playbook (left). All strategies in the playbook use the same tags to define the players position, such as, *CB* for cornerback player or *H* for running back player. These tags are used by the OpenCV application to find the initial position of each player, as well as they are used to localize the closest line that determines where the movement line of the player starts. Then, player tags are replaced by dots and the remaining text is removed from the final image converted. Additionally, the player movement lines are converted in similar lines that the ones AICO uses in the converted tracking images of the players generated in the player detection step. In the Figure 8, the defensive strategy called *Aegan* has only 6 players, since the 5 players from the defensive front line are not included because they are not relevant to the strategy. The defensive front line players are not relevant due to their only objective is to go forward and tackle the quarterback independently of the defensive strategy used.

The converted strategies are stored in a database that can be used by AICO in the future, this conversion process is executed only once per strategy. In this database is store the original strategy, the name of the strategy and the converted strategy. The coaches can use several playbooks and increase the strategies that AICO contains in its database. As consequence, AICO will have more strategies to compare with, achieving a more precise detection of the strategy used by the teams. For our experiments, AICO uses a strategy database of around 100 offensive and 100 defensive strategies.

AICO uses OpenCV to search the most similar strategies from the playbook that matches the converted tracking result images obtained in the player detection step. *SIFT* algorithm [Low04] is used to compare the tracking image of the offensive and defensive teams against all the offensive and defensive strategies contained in the playbook, respectively.

The offensive and defensive strategies with the highest successful comparison percentage from the list are identified as the strategies performed by the offensive and defensive teams. This percentage is determined by the confidence match of the converted tracking image and the strategy image from the playbook. However, if the highest percentage is lower than 65%, then AICO assumes that the playbook does not contain the strategy tracked. In this case, AICO sends an email to the coach to inform that a new strategy has been detected. At this point, the coaches can create a new strategy and send this new strategy to AICO. Then AICO will compare and match this new strategy in the play.

The output of this strategy detection step are two strategy images from the playbook, offensive and defensive, which match the converted tracking images obtained in the player detection step.

3.5 Strategy Result Detection

At this point, AICO can detect and track players, and determine the strategy used by both teams in a play. However, AICO does not know yet if the strategies used by each team were successful or not. For this purpose, AICO needs to count the downs of the offensive team and calculate how many yards the offensive team has gain/lose in each down.

Regarding the count of downs, AICO needs to identify if the team continue being the same in the next play, if that is the case, the down counter of that team is increased. If it is not the same team, the possession of the ball has changed, so the down counter needs to be set to 1 and the ball position is stored as *initial ball* position. This *initial ball* position is used at the end of this step to determine which strategy is successful in each down.

To detect the team, AICO uses the helmet color of the *Center* player detected in the player detection step. AICO stores the color of the helmet and compares it to the helmet color of the *Center* player in the previous *clip play*. In case the color is the same, then the offensive team has not changed and the down counter is increased by 1. Otherwise, the offensive and defensive team has switched roles.

To determine the yards gain by a team, AICO checks the ball position of the *play clip* and compares how many yards has moved from the previous *play clip*. AICO uses the orientation of the offensive team to determine if the yards gained were positive or negative from the previous *play clip*.

AICO uses the yards calculation obtained by the offensive team, the down counters and the *initial ball* position to determine if the strategy used in the *play clip* was successful or not. Therefore, according to Football Outsiders, AICO defines an offensive strategy as

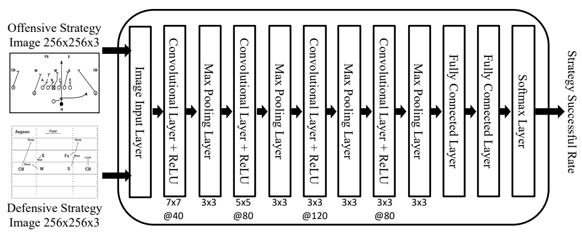

Figure 9: The CNN structure of AICO

successful if it gains at least 4 yards (40%) for the first down, 6 yards (60%) for the second down and 10 yards (100%) for third or fourth down compared to the *initial ball* position. Otherwise, the offensive strategy is defined as *unsuccessful*. Regarding the defensive strategy used in that *play clip*, it is consider *unsuccessful* when the offensive strategy is labeled as *successful* and vice versa.

After all the *play clips* of a whole match have been analyzed, AICO asks the coaches to give the names of both teams that participated in the match. AICO identifies both teams using the color of the player's helmets. Lists of all the offensive strategies used by each team are stored. There is one list per team. These offensive strategy lists are used later by the coaches to select the offensive strategy that they want AICO to compare against their defensive strategy.

3.6 AICO

AICO adopts the CNN structure shown in Figure 9 which has 12 deep neural network layers: one image input layer, four pairs of convolutional and max pooling layers, two fully connected layers and one softmax layer.

The filter size and the number of filters of each convolutional, ReLU and Max Pooling layers are set by parameter fine-tuning, shown below them in the figure. For example, the first convolutional layer has a 40 filter of $7x7$ size. Both fully connected (FC) layers multiply the input by a weight matrix and then adds a bias vector. The first FC layer uses a weight matrix of $50x1280$ numbers and a bias vector of 50 numbers. The second FC layer uses a weight matrix of $18x50$ and a bias vector of 18 numbers. The last layer is a softmax layer that uses the softmax function, also known as the multiclass generalization of logistic regression [Gho18]. Last but not least, AICO CNN structure output is the success rate of the strategy input against the other team.

The dataset utilized for training this CNN contains information gathered in the *preprocessing step* from several *play clips* of the target team, containing only offensive *play clips* of that team. Every dataset entrance contains the offensive strategy image used by the target team in a *play clip*, the defensive strategy image used against the target team in the same *play clip*, and a label confirming if the offensive strategy was *successful* or not.

Since AICO needs to be trained individually for each opponent team, there will be a specific AICO per every opponent team. For instance, if the coaches want information from 3 teams, there will be three different AICOs trained using different datasets related to each team.

After AICO has finished its training and now it can be used by the coaches. AICO provides a list of offensive strategies used by a team and the neural network trained for that specific team. This neural network has two inputs: an offensive strategy and a defensive strategy image with a size of 256*x*256 matrix. As consequence, both images will be resized before they are input in the CNN. The offensive strategy image is an image selected previously by the coaches from the offensive strategy list of the target team. The defensive strategy image is an image of a defensive strategy provided by the coaches.

In summary, AICO examines the defensive strategies received from the American football coaches against the offensive strategy of the target team, and returns the effectiveness of using this defensive strategy against that offensive strategy of the opponent team.

4 EXPERIMENTS

AICO performance has been tested using real-world American football videos. Since we could not find available public datasets, we requested to KSS Lotus team to provide us American football video matches from diverse teams.

KSS Lotus team has supplied videos from 5 different teams, which are considered as opponents, playing against other teams. There is a total of 10 matches per opponent team. These 10 matches are used as a ground truth due to we know in advance the strategies used in each play by each team. A total of 5 additional matches where KSS Lotus team played against that 5 teams, one match per team, together with the respective strategies, has been provided as well.

The 50 matches of the opponent teams together with the strategies used by both teams in each *play clip* were used to train AICO. The matches of KSS Lotus team are used to verify AICO's strategy prediction accuracy. All of these 55 matches has been recorded using a Sony FDR-AX60 video camera with a quality of 720p.

KSS Lotus coaches segmented the 55 match videos into *play clips*, and we grouped them in a dataset. In total, the dataset contains about 7350 *play clips*, where 6700 are *play clips* from the 5 opponent teams and 650 are the *play clips* where KSS Lotus team is playing against one of the 5 opponent teams. Based on

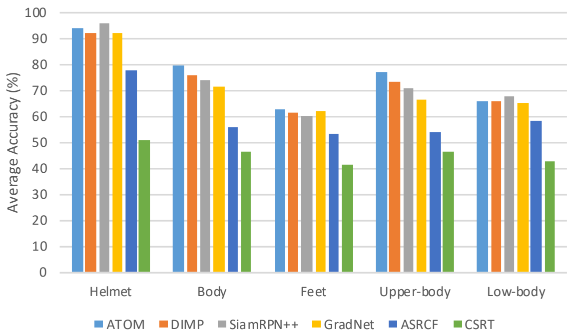

Figure 10: Tracking accuracy comparison using different parts of the player's body

the video matches used in the experiments, an American football team performs around 65 offensive and 70 defensive plays per match. This number can vary if one of the teams is stronger than the other team in the match. As a result, the dataset created contains about 650 offensive and 700 defensive plays of each opponent team, no counting the match against KSS Lotus team. The 6700 *play clips* are used to examine AICO performance on all the experiments together with the playbook strategies, however, the 650 *play clips* where KSS Lotus team played as well as the playbook are used to verify the defensive strategy selection effectiveness. This dataset containing all the *play clips* will be made publicly available in future projects.

4.1 Tracking Accuracy

Due to the movement of the players, a tracker can jump from one player to another, since one player is overlapped by another player in the video. As a consequence, one player may have two trackers, having lost the other player his tracker. It makes difficult to determine the strategy used by the teams since the tracker is not following the correct player. For this reason, it is necessary to achieve a high tracking accuracy.

In this experiment, we have used 100 *play clips* where we manually defined the position of each player as a ground truth. After having these ground truth videos, we have tested the following trackers from the *seventh visual object tracking vot2019 challenge* [Kri19]: ATOM, DIMP, SiamRPN++, GradNet and ASRCF. Each of this trackers has been trained to track American football players using the datasets created for the players detection. We have tested as well the CSRT tracker [Luk17] provided by OpenCV. Additionally, these trackers has been tested using different parts of the players body: helmet, body, feet, upper-body and low-body.

Figure 10 shows the tracking accuracy of the trackers tested using different parts of the player's body. This tracking accuracy is obtained by comparing the ground truth movement of a player together with his tracked movement. As we can see in the results, the player's helmet is the part of the body that obtained the highest tracking accuracy, and the best tracker is SiamRPN++ achieving around 96.075% accuracy. For this reason, AICO tracks the helmet of each player using SiamRPN++.

4.2 Neural Networks

In this research, the neural networks utilized are required to achieve a high accuracy performance. We compared the performance of the following CNN classifiers: Inception [Iof15], Resnet [Kai16] and Mobilenet [How17]. Regarding Inception models, they have been tested using the four versions currently available (*Inception_v1*, *Inception_v2*, *Inception_v3*, *Inception_v4*), and *Inception-ResNet-v2* which is a hybrid inception module using Resnet and has a similar computational cost than *Inception_v4*. Additionally, the following deep residual networks(ResNets) are also tested: *Resnet_v1_50*, *Resnet_v1_101*, *Resnet_v1_152*, *Resnet_v2_50*, *Resnet_v2_101*, *Resnet_v2_152* and *Resnet_v2_200*. Last but not least, diverse configurations of Mobilenet with 224 as input resolution have been added to the experiments: *Mobilenet_v1*, *Mobilenet_v1_075*, *Mobilenet_v1_050* and *Mobilenet_v1_025*. All these models are built upon TensorFlow GPU.

To the best of our knowledge, there are not any datasets available to detect American football balls, player's helmet or the players in the field. For this reason, we have created 3 different datasets using images and videos provided by KSS Lotus team from season 2013 to 2018. These datasets are composed of 2000, 15000 and 20000 images of balls, helmet and players, respectively, and they will be publicly available for everyone in the future. For the experiments, these datasets have been used to train each CNN classifier.

Figure 11 shows the average accuracy of detecting the ball, the helmet and the player using different neural networks. Inception_v4 achieves the highest accuracy (92.41%) on detecting the players in the field compared to the other CNN classifiers. Regarding the ball and helmet detection, the best option for both is Inception_v2 since it always detects them and has the fastest inference time, compared to other CNN that achieves the same accuracy. Since the ball is always detected, AICO can always define the *goal line* that is utilized to determine if the offensive strategy was *successful* or not.

As a result, AICO uses two Inception_v2 to detect the helmet and the ball, and one Inception_v4 to detect the players in the field.

4.3 Strategy Selection Accuracy

In the experiments, AICO has used the KSS Lotus playbook to detect the most similar strategies performed by

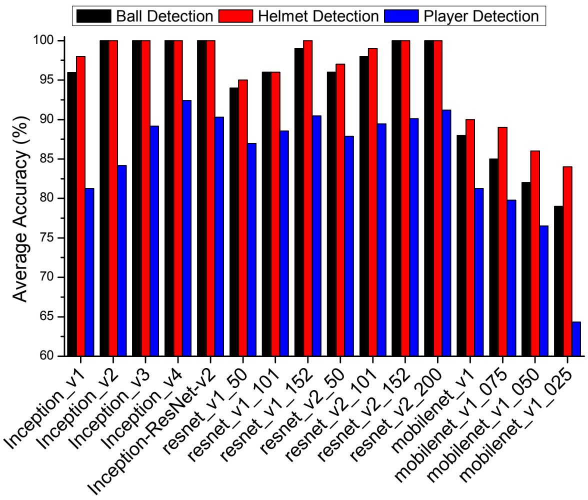

Figure 11: Average accuracy of detecting the ball, the helmet and the players using different CNN classifiers

each team in a *play clip*. This detection mechanism is similar as the coaches made in real life (Figure 1). After comparing the strategies selected by AICO to the strategies selected by KSS Lotus coaches in each *play clip*, we have verified that the 8% missing recognition of the players in the field made by Inception_v4 does not affect the detection of the correct strategy. As a result, about 95.65% and 90.58% of the times AICO recognizes the same defensive and offensive strategies as the coaches, respectively. Regarding the offensive strategies, 88.58% and 92.58% are achieved for running and passing offensive strategies, respectively.

4.4 Strategy Prediction Effectiveness

For this evaluation, the process is divided in three steps. First, AICO is trained using different amounts of an opponent match videos. In the second step, we examine the KSS Lotus team match against this opponent. In this analysis, we obtain the defensive strategies used by KSS Lotus team together with the successful rate of each defensive strategy used per offensive strategy used by the opponent team. For example, if KSS Lotus team performs the defensive strategy *A* 4 times against the offensive strategy *B* of the opponent team, being 3 times successfully defended, then strategy *A* has achieved 75% of successful rate.

In the third step, we input to AICO the strategy image of *A* together with the strategy image of *B*, and then AICO returns the successful rate of using *A* against *B*. For the experiment results, it is compared the successful rate of *A* against *B* obtained by AICO to the one obtained in the second step.

Figure 12 shows a successful rate example of using a some defensive strategy, such as *Zombie*, against to an offensive strategy like *Crunch_Read*. In blue it is represented the successful rate obtained by KSS in the

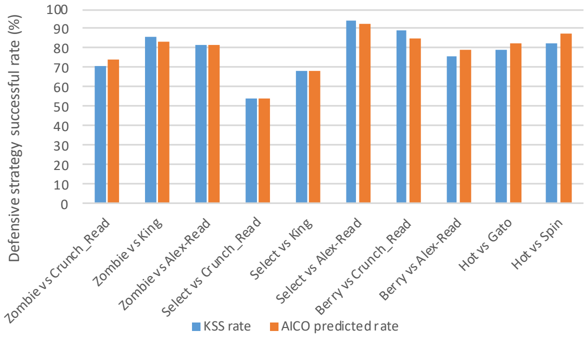

Figure 12: Example of some defensive strategy successful rate accuracy against offensive strategies

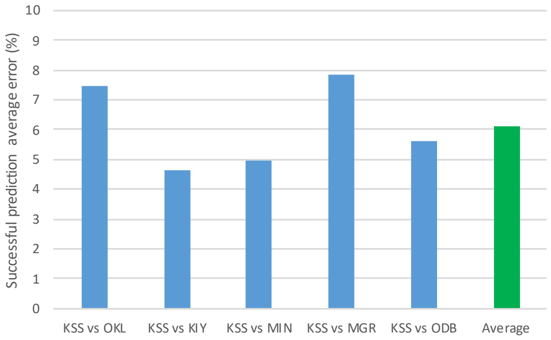

Figure 13: Successful prediction average error made by AICO using the 5 matches where KSS is playing

match, and in orange the successful rate predicted by AICO.

As we can see, AICO's successful rate matching up with that of KSS indicates the high performance level of AICO to predict how well a defensive strategy will perform in the real world.

Figure 13 shows the average error of predicting the successful rate using 5 matches of KSS playing against different opponents. AICO achieves 6.11% of error prediction, confirming that AICO can help the coaches to predict how successful will be a defensive strategy against to a particular offensive strategy

5 CONCLUSIONS

This paper presents a novel video recognition system working together with a neural network structure, AICO, which has been developed to analyze the opponent team strategies to successfully obtain the best defensive strategies to use against that adversary. Video analysis task made by American football coaches may become easier and more precise thanks to AICO.

In addition, during this research, we have created a dataset that contains about 7350 *play clips* based on different American football teams which will become publicly available in future projects.

This video analysis recognition system can work with any type of recording, even if only a single camera has been used, and it can easily be used by any American football team at any field, since it does not require previous installation in the stadium, field or players.

AICO achieves a 93.12% of accuracy in strategy detection of both teams in a play compared to the coaches judgment. Furthermore, AICO obtains a successful rate prediction with around 94% of accuracy determining the successful rate of a defensive strategy against an offensive strategy of an opponent team. As a result, AICO becomes a reliable artificial coach advisor to help coaches in their opponent strategy analysis.

ACKNOWLEDGEMENT

Thank you very much to Fujitsu and Kosei Gakuen High School American football (KSS Lotus) for their support and help to develop AICO.

6 REFERENCES

[Atm13] I. Atmosukarto, B. Ghanem, S. Ahuja, K. Muthuswamy, and N. Ahuja. Automatic recognition of offensive team formation in american football plays. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 991–998, 2013.

[Bal87] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. In *Readings in computer vision*, pages 714–725. Elsevier, 1987.

[Bur17] M. Buchheit and B. M. Simpson. Player-tracking technology: half-full or half-empty glass? *International journal of sports physiology and performance*, 12(s2):S2–35, 2017.

[Che14] S. Chen, Z. Feng, Q. Lu, B. Mahasseni, T. Fiez, A. Fern, and S. Todorovic. Play type recognition in real-world football video. In *IEEE Winter Conference on Applications of Computer Vision*, pages 652–659. IEEE, 2014.

[Chu03] O. Chum, J. Matas, and J. Kittler. Locally optimized ransac. In *Joint Pattern Recognition Symposium*, pages 236–243. Springer, 2003.

[Dir18] C. Direkoglu, M. Sah, and N. E. OConnor. Player detection in field sports. *Machine Vision and Applications*, 29(2):187–206, 2018.

[Ela08] E. Dubrofsky and R. J. Woodham. Combining line and point correspondences for homography estimation. In *Advances in Visual Computing*, pages 202–213, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[Fre19] M. Frey, E. Murina, J. Rohrbach, M. Walser, P. Haas, and M. Dettling. Machine learning for position detection in football. In *2019*

6th Swiss Conference on Data Science (SDS), pages 111–112, June 2019.

[Gho18] J. Ghosh, Y. Li, R. Mitra, et al. On the use of cauchy prior distributions for bayesian logistic regression. *Bayesian Analysis*, 13(2):359–383, 2018.

[Goo18] R. Goodell. Official playing rules of the national football league. *National Football League: New York, NY, USA*, 2018.

[Har03] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003.

[How17] A. G. Howard, Z. Menglong, C. Bo, K. Dmitry, and W. Wang. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[Iof15] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML '15*, 2015.

[Jia16] H. Jiang, Y. Lu, and J. Xue. Automatic soccer video event detection based on a deep neural network combined cnn and rnn. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 490–494. IEEE, 2016.

[Jof15] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[Kae16] A. Kaehler and G. Bradski. *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. " O'Reilly Media, Inc.", 2016.

[Kai16] K. He, H. Kaiming, Z. Xiangyu, R. Shaoqing, and J. Sun. Identity mappings in deep residual networks. In *ECCV '16*, 2016.

[Kri19] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, R. Pflugfelder, J.-K. Kamarainen, L. Cehovin Zajc, O. Drbohlav, A. Lukezic, A. Berg, et al. The seventh visual object tracking vot2019 challenge results. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[Lee16] N. Lee and K. M. Kitani. Predicting wide receiver trajectories in american football. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, 2016.

[Li19] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4282–4291, 2019.

[Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004.

[Luk17] A. Lukezic, T. Vojir, L. Cehovin Zajc, J. Matas, and M. Kristan. Discriminative correlation filter with channel and spatial reliability. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6309–6318, 2017.

[Out19] Football outsiders, 2019. URL `https://www.footballoutsiders.com/`.

[Rie13] W. Riedel, D. Guillory, and T. Mwangi. Ee368 project: Football video registration and player detection. 2013.

[Sid09] B. Siddiquie, Y. Yacoob, and L. Davis. Recognizing plays in american football videos. *University of Maryland, Tech. Rep*, 111, 2009.

[Ste17] M. Stein, H. Janetzko, A. Lamprecht, T. Breitkreutz, P. Zimmermann, B. Goldlücke, T. Schreck, G. Andrienko, M. Grossniklaus, and D. A. Keim. Bring it to the pitch: Combining video and movement data to enhance team sport analysis. *IEEE transactions on visualization and computer graphics*, 24(1):13–22, 2017.

[Sze17] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[Wy015] Y. Wu, J. Lim, and M. Yang. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1834–1848, Sep. 2015.

[Yam13] T. Yamamoto, H. Kataoka, M. Hayashi, Y. Aoki, K. Oshima, and M. Tanabiki. Multiple players tracking and identification using group detection and player number recognition in sports video. In *IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society*, pages 2442–2446. IEEE, 2013.

[Zho14] W. Zhong, H. Lu, and M. Yang. Robust object tracking via sparse collaborative appearance model. *IEEE Transactions on Image Processing*, 23(5):2356–2368, May 2014.