International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 8 (2018) pp. 6158-6164

Integration of IOT and Data Analytics: Performance Analysis System for Outdoor Sports using Electromyography and Image Data

Dr. B. Swaminathan, R. Rahul, S. Ragul and M. Ramavel

Computer Science and Engineering department, Rajalakshmi Engineering College, Rajalakshmi Nagar, Thandalam, Chennai, Tamil Nadu, India.

Abstract

The performance of sportsmen can be improved using statistical and inferential analysis. People depend on technology for making their lives better. Sport is one industry that has been enjoying all the benefits from technologies for the past few years. A notable fact here is that only historic data of the matches are taken into account for better insights about the game and players. This scenario can be countered by applying machine learning algorithms to real time data that get generated in the course of a game using sensors and IOE for communication of those data with the processing devices. Such an idea is discussed here where data analytics is made on data from a muscle sensor in correlation with processed images captured while shots were played. The results provide the behavior of muscle fibers of a player during the course of the game with respect to the player positions and angle of shots. Hence these data will provide coaches and players with wealthy and crucial knowledge that otherwise goes unnoticed. Also the performance enhancement can be made and game play shall be adjusted in the middle of the game with real time feedbacks.

Keywords: Xbee Bluetooth Module, EMG, IoE, Multiple Linear Regression, Haar Cascade, Real Time Analytics, OpenCV, MyoWare Muscle Sensor

INTRODUCTION

The modern world is gradually moving towards automation and wireless technologies. Though it is difficult to incorporate them in every industry, there is no denying the fact that better results will be yielded in everything, since there is no room for human error. The concept of introducing sensors and IOE to sport industry, has been discussed in this paper. Coaches, as humans, do not possess the ability to evaluate intricate details of efforts made by their player students to understand their performance. The contribution only to factors that can be comprehended is carried out. With emerging technology in wearable sensors and wireless networks, the sources of information about the performance of players shall be improved and widened. Compact and wearable sensors can perform the job of collecting unrecognized data from players and wireless networks provide the environment to transmit to processing devices for analysis. This idea can change the whole perspective of how a sport can be interpreted for gaining meaningful knowledge out of each match. Also, the results of analyses can be delivered to players during the course of a match, in intervals, to adjust the strategies in order to win. This is because of the fact that the production of data is done in real time. The methodologies and technologies used to achieve this are discussed in the further sections.

DATA ANALYSIS

Real Time data analytics is done in order to improve the performance of the business, marketing, sports skills, etc. anywhere, anything and anytime. Real Time analytics is the ability of a business enterprise to utilize all available enterprise data at the time of need. A critical feature of real-time analytics is that the available systems and setup should quickly generate reports and visualization of knowledge gained out of the data received, ideally within a minute after the data got generated. A big advantage of real-time analytics is that the data is fresh and versatile. Organizations can reap a lot of profits from doing real-time analytics purely because of their close relevance to market realities.

As per Gartner's definition, big data has 3 dimensions. To develop a commercial end to end application for a business enterprise, all the incoming sensor data needs to be captured and stored. The voluminous data needs to be formally organized and synthesized to get the real meaning in real time. The demand is to provide an application that can capture a variety of data, including textual format. Analyzing data using a batch process is not a feasible solution for Real time data. This works well when the incoming data rate request is slower than the batch processing rate. With the new sources of data such as social media, mobile devices and sensors, the batch process is not effective.

Real-time refers the ability to process data at the same moment as it arrives, rather than processing it in batches later. So, the main significance is that the data are not stored and processed in the future, but done immediately. It shall be defined as the ability to make better decisions and carry out meaningful actions at the right time. It is about combining and analyzing

data, so that right actions shall be executed at the right time and the right place. The essence is to generate value from disparate data. A live video camera is used to make comparative analysis of the data from the sensor with respect to the shots played which are caught on camera. This paves way for gaining information about the player's performance at different stages of the game, the contraction and relaxation of muscle fibers while shots are played at certain angles and position in Tennis, Cricket and other outdoor sports.

A. Classification, Clustering

Classification and clustering techniques are the most prominent algorithms used for analytics and reports for visualization. There is no intention to make predictions for future in this paper, which leaves out the regression techniques.

Classification is the process of predicting the target class of the dataset by carrying out analysis on the training dataset. In our application, professional sportsmen get to wear the product and generate top performance data sets, which can then be used to classify an ordinary player across various quality levels based on his/her performance. The various classification algorithms are,

Clustering is the process of grouping similar dataset objects into clusters, in our case similar game strategies. It is an unsupervised learning algorithm capable of providing numerous clusters, each describing very minute deviations in the behavior of datasets, which is very effective, when a proper clustering algorithm is chosen and trained. Clustering algorithms can be classified into two main categories, Linear clustering algorithms and Non-linear clustering algorithms,

Linear clustering algorithm is classified into k-means clustering algorithm in that bringing all the elements in to a cluster, and find out nearest object to mean. Hierarchical clustering algorithm, Gaussian(EM) clustering algorithm and Quality threshold clustering algorithm are other linear clustering algorithms to name a few.

B. Muscle Arm

The activity of muscles is an important aspect when it comes to sports. This factor often go unnoticed and neglected. There are sensors incorporated in sport equipment such as cricket bats to find out the pressure exerted on the ball after a shot. But the data doesn't convey necessary insights as it directly depends on the swing speed and speed of the ball. So it cannot be accounted for the determination of performance of players. Contraction and relaxation are the two processes carried out by muscle fibers in arms such as the Biceps and the Triceps. In cricket, while playing certain shots or catching balls, we need to use soft hands which is a result of muscle relaxation. In tennis, it can be seen while playing volleys and drop shots. In cases where we need to find gaps to score shots and exert force, our muscles contract. This behavior conveys crucial knowledge about the performance and efforts of players with respect to other data such as the position of bat and height of shots.

C. EMG Electrode pads

The data, electrical signals, produced by skeletal muscles are in the form of analog data, termed as "Electromyography". And are primarily used in the medicinal field as a diagnostic procedure to assess the health of muscles and the nerve cells that control them (motor neurons). Motor neurons transmit electrical signals that cause muscles to contract. The surface EMG electrodes are used to record this data which are minute and can be attached on the skin surface without pain. More than one electrodes are deployed in general, because they produce potential difference (voltage difference). The output data can be read in Volts.

SPORTS SYSTEMS IN EXISTENCE

The world of sport is constantly changing over the years and technology has a major part in it. Technology can be used to make people train effectively for improving their performance. It helps in analyzing sport performances and enabling coaches to improve the quality of feedback to players/athletes by a great extent. Accuracy in calculating delicate measurements and microscopic details shall be improved greatly. Referees, umpires and sport officials can make better decisions on rule infringements. But the implementation is monotonous. Only historic data are provided as sources for analytics. IOE is still at its growing stage in any industry, if not just in sport. The existing systems involving analysis in cricket discusses the process of studying the performances of cricket players.

A. Game of Tennis

The paper "An automatic system for sports analytics in multi-camera tennis videos" by Rafael Martín Nieto and José María Martínez Sánchez (2013) of the "Universidad Autónoma de Madrid" presents an automatic system which could detect and track each one of the players on the court or field in single player sports. After that, the system could be able to extract statistics and performance of the players. This system was modular. The system was based on a camera detection and tracking system of players in the sports domain. The system would identify the position of players in individual sports where they have their own spaces in the field. The paper uses camera as a single source of data (image and video files) for make analysis.

B. Stroke recognition in tennis

Our paper proposes the concept of image processing in order to obtain screen shots of video frames when a shot is played in any sport. This mechanism has been detailed in a paper titled "Player Tracking and Stroke Recognition in Tennis Video" by Terence Bloom and Andrew P. Bradley. They aimed at tracking tennis players and recognize the strokes played by them. They eventually wanted automate the task of digital tennis footage annotation resulting in the metadata, which are the time codes and description of strokes, automatically getting appended to the video. The system developed solves the

problem of tennis player tracking and stroke recognition using image processing operations with the knowledge of the image capture conditions, the background scene, and the application domain.

C. Inference

Hence there have been a lot of enthusiastic experiments made in various sports using modern technology. The developments are still in their early stages. Predominantly only statistical data from completed matches are being considered for algorithms and methods to carry out data analytics. There is lack of application of sensors developed as wearable devices in sports. Notable systems such as pressure sensors incorporated in the cricket bat to calculate the bat swing speed, Hawk eye technologies to determine a point or score in more than one sport are in existence. The source of data has also been limited to a single one.

METHODOLOGY FOR WINNING STRATEGY AND TECHNOLOGIES

Winning a match in any sport involves effort, pressure management and hard work. But there is a void that needs to be filled here to complete the list, which is nothing but "Smart work". It basically is the process of obtaining better results by saving time and effort. The idea behind the same is to concentrate on crucial factors that matter the most for the outcome of a match. One such factor is the muscle activity. In the context of smart work, introduction of technology alone can help to achieve what is needed. Hence, in the further sections, discussion about the application of various technologies such as sensors and wireless networks to formulate winning strategies based on players' muscle activity on the field is done.

A. Gathering the real-time data

The real time data that we talk about in this paper are those of the contractions and relaxations of the muscle fibers on our body, especially forearms and biceps which indicate the force that a competitor is trying to generate on a shot in Tennis or Cricket or Badminton, etc. A muscle sensor is used to serve this purpose. The muscles are the primary factors to control things and people are accustomed to doing it while pushing buttons, pulling levers, moving joysticks. The MyoWare Muscle Sensor Development Kit including electromyography (EMG) sensing pads provide a portable, wireless way of generating analog signals denoting the muscle activity. The MyoWare board, included in this kit, acts by measuring the filtered and rectified electrical activity of a muscle and outputs 0-Vs Volts depending on the amount of activity in the selected muscle. The shields are to be attached to the muscle sensor, including a few electrodes, reading the voltage out by flexing our muscles. The sensor is the latest revision of the Muscle Sensor of old, now with a new wearable design that allows us to attach biomedical sensor pads directly to the board itself getting rid of pesky cables. This new board also includes a slew of other new features including, single-supply voltage of +3.1V to +5V, RAW EMG output, polarity protected power pins, indicator LEDs, and an On/Off switch. Additionally, there are a few shields (Cable , Power , and Proto) that can attach to the Muscle Sensor to help increase its versatility and functionality. Measuring muscle activity by detecting its electric potential, referred to as electromyography (EMG), has traditionally been used for medical research. However, with the advent of ever shrinking yet more powerful microcontrollers and integrated circuits, EMG circuits and sensors have found their way into all kinds of control systems.

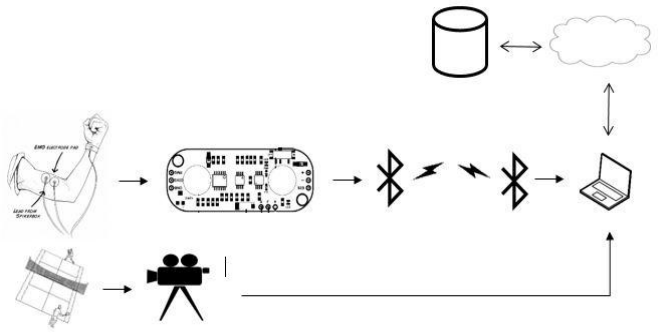

Figure 1. Architecture depicting the flow of data from Myoware sensor and Video camera via Bluetooth to the processing system for performing analysis

B. Transmitting the data across wireless network

It is necessary to think about the data generated by the device in three stages. Stage one is the initial creation, which takes place within the device, and then sent over to processing devices via wireless networks. Stage two is about the central system collecting and organizing that data. Stage three is the usage of the obtained data for the future processes. For smart devices and sensors, each event creates data. But they are very much redundant that need to preprocessed and cleaned. This information can then be sent over the network back to the central application. At this point, the data standards and ways of sending it over to the network need to be addressed. The analog data from myoware sensor is acquired for test data which are later converted into digital data for processing. It is done with the help of Xbee Bluetooth module. The XBee RF Modules interfaces a host device via a logic-level asynchronous serial port. The module is capable of communicating with any logic and voltage compatible UART through its serial port or a level translator to any serial device. Data is acquired by the module UART via the DI pin (pin 3) as an asynchronous serial signal. The signal will be at idle when there is no data being transmitted. The data bytes consist of a start bit (low), 8 data bits (least significant bit first) and a stop bit (high). On the other hand, a Rewy wireless video camera is used to stream the ongoing game to a processing device which in turn captures screenshots of players while striking the ball in tennis.

C. IOE Implementation

Across the IOE, devices create data that are sent to the main application, consumed and used. Depending on the device, the network and power consumption restraints, data can be sent in real time, or in batches any time. However, the real value is derived from the order of creation of data points. Time-series data shall be created as events take place around the device and are then sent. This use of real-time information provides a complete record for each device. Alternatively, it can be collated as data is sent across in batches and the historical record of data will also be there but just aren't available in real time. This is common with devices where battery life is a key requirement to meet the need for data to be delivered in real time. Either way, the fundamental requirement is that each transaction on each device must be placed in at the right time-stamp for sorting and alignment. If this is needed to be achieved in real time with hundreds of thousands or potentially millions of devices, then write-speed at the database level is an essential consideration. The data from the muscle sensor in our application are then associated with the angles and positions, with respect to the body, wrists, elbows, etc. of the shots played which are caught on the live recording video camera. The heights at which the shots are played from ground level also considered as a data source. The processing of images are made on the processed information from the video using image processing techniques with python OpenCV and stroke recognition process from an existing paper discussed above. The correlation of image data (time stamps of strokes) and sensor data are done in order to refine the sensor values which are obtained in abundance. The point of contact of the ball with the bat/racket is the point of coincidence that contributes to the quality of the shot. This is where the correlation is made to gain performance information neglecting data at other instances.

D. Image Processing

The images are processed for finding the height at which the shots are played, angles and how much a player stretches to reach the ball. A trained cascade of images are used to recognise the necessary sources such as tennis ball, shoe and body to carry out processing. Object Detection using Haar feature-based cascade classifiers is effective for object detection. It is a machine learning based approach. Here, a cascade function is trained from positive and negative images. It is then used to detect objects in other images. In Haar-cascade, the system is provided with several positive images (Tennis balls and sports shoes in our case) and negative images (images that do not contain anything of our interest like chair, table, wall, ground, human etc.), and the feature selection is done with the classifier training using Adaboost and Integral images. Classification learning requires a set of positive and negative images for training to differentiate them and a set of features are selected for training the classifier, which is done using Ada-boost. The 'Boosting' process works with the learning of single simple classifier, then rewriting the weights of data where errors were made with higher weights. Afterwards a second simple classifier is trained on the weighted classifier, and data are reweighted by the combination of 1st and 2nd classifier, so on till the final classifier is learned. Therefore, the final classifier is the combination of all n-classifiers obtained previously. To train the cascade, the no of stages to undergo and memory requirements for processing are exclusively provided. At the end, an XML file is generated containing the trained images with features for matching. The objects are then recognized from screenshots and the values for shot height, amount of stretch are calculated. The calculation is done by finding the difference of the coordinates of two distinct objects.

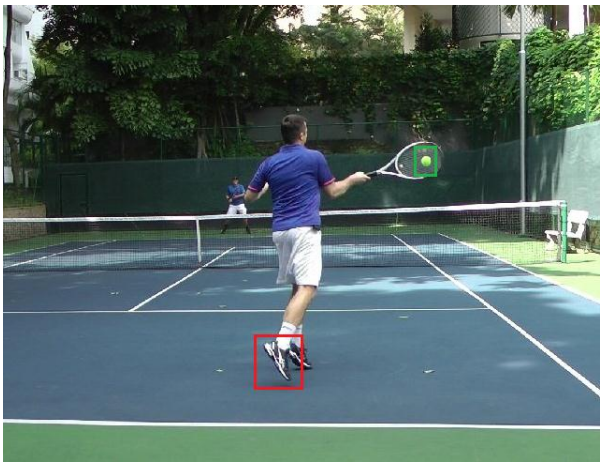

Figure 2. The shoe and Tennis ball are recognized from the screen shots using trained cascades.

The data from sensor is preprocessed by filtering out necessary data, i.e., only the data generated at the time of playing a shot. This is achieved by comparison of timestamps. We obtain data whose timestamp matches with that of the screenshots. Then, we eliminate the outliers and noise. At the end of this module, we have 3 attribute performance estimation data, from two different sources, that is ready for analysis.

E. Multivariate Regression Learning

Once the data are obtained from sensors to a processing device, they are refined with respect to the shots played, as discussed above. The shots are captured using a video camera. This process is known as data preprocessing. Now we have sensor data and image data in hand which are ready to be used in a data model or analytical model. The models are generated by machine learning/ deep learning algorithms to carry out analytics.

MULTIPLE LINEAR REGRESSION (MLR) is used to determine a mathematical relationship among the data from sensor and the camera. In examines how multiple independent variables, in our case 3, are related to one dependent variable. Once we determine each of the independent factors to predict the dependent variable, the information on the multiple variable create an accurate prediction on what levels they affect the outcome variable. The model creates a relationship in the form of a straight line (linear) to approximate all the individual data points in the best way.

The model for multiple linear regression is:

$$y = B_0 + B_1x_1 + B_2x_2 + B_3x_3 + e \quad \text{..... (1)}$$

Where y = Muscle power (EMG Signal – Volts) : dependent variable.

x_1 = Height of the Shot (Cm) : independent variable 1.

x_2 = Distance from the body i.e., the reach (Cm) : independent variable 2.

x_3 = Backhand or Forehand (0 or 1).

e = random error in prediction, that is variance that cannot be accurately predicted by the model. Also known as residuals.

B_0 = y - intercept at time zero.

B_1 = regression coefficient that measures a unit change in the dependent variable when xi1 changes.

B_2 = coefficient value that measures a unit change in the dependent variable when x_2 changes.

The least squares estimates, B_0, B_1, B_2, B_3 are calculated by statistical software. As multiple variables can be included in the regression model, each independent variable is differentiated with a number - 1, 2, 3, 4...p. The multiple/polynomial regression model allows an analyst to predict a result based on knowledge gathered from multiple explanatory variables. Still, the model is not always completely accurate as each data point can differ slightly from the predicted values by the model. E is the residual value which is the difference between the outcome variable and the predicted outcome variable. It is included for accounting slight variations.

The multiple regression model is based on the following assumptions:

1. The dependent and independent variables have a linear relationship.
2. The independent variables are not highly correlated with each other.
3. 'y' observations are selected randomly and independently from the population.
4. Residuals need to be distributed normally with a mean of 0 and variance σ.

The co-efficient of determination is used to calculate the explanation of variation in outcome with the variation in the independent variables. It is a statistical metric and denoted by R-squared or R2. R2 increases every time as more predictors are added to the MLR model though the predictors may not be related to outcome variables. Therefore, R2 by itself, cannot be used to identify whether the predictors need to be included in a model or not. R2 can only lie between 0 and 1, where 0 indicates that the outcome cannot be predicted with the help of any of the independent variables and 1 indicates that the outcome can be predicted without causing any error from the independent variables.

The mplot3d toolkit supports 3D plotting functions to matplotlib by supplying an axes object that can develop a 2D projection of a 3D environment. The interactive back ends also supply the capability to rotate and enlarge the 3D scene.

X – axis : Muscle Power (V)

Y – axis : Height of Shot (cm)

Z – axis : Distance from body/Reach (cm)

For a sample test data,

Height of Shots (CM)	Distance from Body/Reach (CM)	Muscle Power (Volts)
164	80	734
178	85	985
190	82	834
140	85	907
128	79	873

The trained MLR model predicts the Ideal EMG volts (Muscle Power) and makes comparison with the generated output power. This is visually shown in the 3D graph by Scattering the Obtained values based on the 3 parameters (Red Dots) and Plotting the Predicted outcome (Green Linear Lines). The variation of factors in the 3D graph with respect to the trained ideal values suggests the performance deviation of players and how much they need to improve their game in aspects of Height of shot, Reach from body and Muscle Power.

[Predicted EMG values are ~ 821,899,851,860,782 V Resp.]

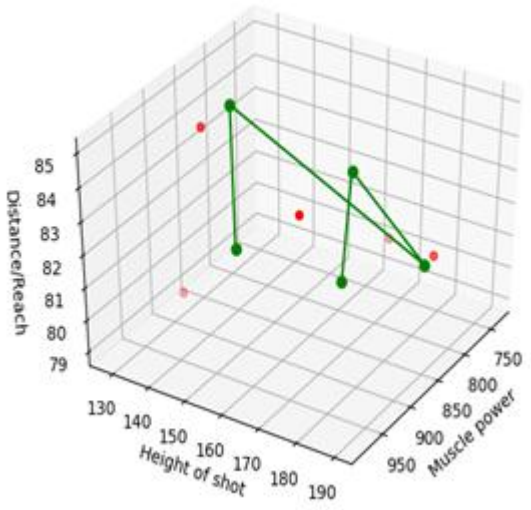

Figure 3. Performance review by Multivariate linear regression model using sensor and image data

F. Integration of Modules

The integration of all the components that are developed in various domains will be done. This paves way for real time analytics of continuous muscle activity of users using the data model created, for better intuitions, to assess their performance and providing revelations to them via the UI. The quality of their game shall be revealed immediately at the end, by taking into account, the image data during stroke play and image processing.

CONCLUSION

Thus the primary motive behind this paper is to enrich the sport industry with cutting edge technology applications using wearable sensor in integration with processing device and analytical tools and correlation of data from multiple sources. The idea can take real time analytics in sport and training methodologies to next level. A different perspective to matches, from technology point of view, can be obtained. Nevertheless, the industry is open to all new advancements that can make the experience more exciting for both the players as well as the spectators.

REFERENCES

[1] Arun Kejriwal, Sanjeev Kulkarni, Karthik Ramasamy. "Real time analytics : Algorithms and systems". Proceedings of the VLDB endowment Vol. 8 no. 12, 2015.

[2] Dann Jud, Rebecca Wu, Dr. Camillo Jose Taylor. "Sports analytics and Image processing". MIT Sloan sports analytics conference, Spring, 2014.

[3] Gurpreet Singh, Opinder Singh, Dr. Jatinder Singh from Khalsa college, Punjab. Wireless "Sensor Network using Bluetooth". International journal of computer science and technology, Vol.3, Issue 2, April-June 2012

[4] Pal Halvorsen, Simen Saegrov, Asgeir Mortensen : University of Oslo/Simula Research Lab, University of Tromso. "Bagadus : An integrated system for Arena sports analytics". MMSys'13, February 26 – March 1, 2013, Oslo, Norway.

[5] https://cdn.sparkfun.com/datasheets/Sensors/Biometric/MyowareUserManualAT-04-001.pdf. Myoware muscle sensor datasheet.

[6] Sneha Soni, Shailendra Shrivastava. "Classification of Indian Stock Market Data Using Machine Learning Algorithms" : (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 2942-2946.

[7] Alok Kulkarni, Sampada Sathe. "Healthcare applications of the Internet of Things: A Review" : (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6229-6232.

[8] Sricharan Shah, Partha Jyoti Hazarika and Jiten Hazarika. "A Study on Performance of Cricket Players using Factor Analysis Approach" : International Journal of Advanced Research in Computer Science, 8 (3), March-April 2017,656-660.

[9] David R. Loker, P.E. Collin G. Frampton, Titan J. McElhaney, Jonathan R. Mook, Anthony M. Sansone. "Remote Data Acquisition using Bluetooth" : Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition, Session 2147, 2005.

[10] Terence Bloom and Andrew P. Bradley. "Player Tracking and Stroke Recognition in Tennis Video": APRS Workshop on Digital Image Computing (WDIC'03), Brisbane, Australia, Feb 2003.

[11] M. Leo, N. Mosca, P. Spagnolo, P. L. Mazzeo, T. D'Orazio, A. Distante. A Visual Framework for Interaction Detection in Soccer Matches. In International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 24(4):499-530, 2010.

[12] P.S. Tsai, T. Meijome, P.G. Austin. Scout: a game speed analysis and tracking system. Machine vision and Application, 18(5):289–299, 2007

[13] Juan C. SanMiguel, José M.Martínez. A semantic-based probabilistic approach for real-time video event recognition, Computer Vision and Image Understanding, 116(9): 937-952, 2012.

[14] R. Hartley , A. Zisserman. A Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.

[15] Valter Di Salvo, Adam Collins, Barry McNeill, and Marco Cardinale. Validation of Prozone: A new video-based performance analysis system. International Journal of Performance Analysis in Sport (serial online), 6(1):108–119, June 2006.

[16] Sachiko Iwase and Hideo Saito. Parallel tracking of all soccer players by integrating detected positions in multiple view images. In Proceedings of the International Conference on Pattern Recognition (ICPR) , pages 751 – 754, August 2004.

[17] Jinman Kang, Isaac Cohen, and Gerard Medioni. Soccer player tracking across uncalibrated camera streams. In Proceedings of the Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), pages 172–179, 2003.

[18] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision . Cambridge University Press, second edition, 2004.

[19] Ming Xu, James Orwell, and Graetne Jones. Tracking football players with multiple cameras. In Proceedings

of the International Conference on Image Processing(ICIP) , pages 2909–2912, October 2004.

[20] Nils T. Siebel and Stephen J. Maybank. Fusion of multiple tracking algorithms for robust people tracking. In Proceedings of the European Conference on Computer Vision-Part IV (ECCV) , pages 373–387, 2002.

[21] Yingen Xiong and Kari Pulli. Fast panorama stitching for high-quality panoramic images on mobile phones. IEEE Transactions on Consumer Electronics, 56(2), May 2010.

[22] Christoph Fehn, Christian Weissig, Ingo Feldmann, Markus Muller, Peter Eisert, Peter Kauff, and Hans Bloss. Creation of high-resolution video panoramas of sport events. In Proceedings of the IEEE International Symposium on Multimedia (ISM) , pages 291 –298, December 2006.

[23] Hao Jiang, Sidney Fels, and James J. Little. A linear programming approach for multiple object tracking. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , June 2007.