International Journal of Scientific & Engineering Research Volume 12, Issue 4, April-2021
ISSN 2229-5518

A Survey on Sports Data Analysis through Video

Aditya Sankolli, Apurva Shrivastava, Savani Nimonkar, Amritanshu Ray, Prof. Snehal Umare

Abstract— Sports analytics is a field that applies statistical procedures through algorithmic processing to analyze various components of the sports industry, such as individual player performance, team performance, business performance, recruitment, and more. Its combination with video analysis has introduced vast applications of the same and in-depth analysis of the sport. However, only in the recent decade that advanced data mining and machine learning techniques have been utilized for facilitating the operations of sports domain. The major research gap in the previous surveys is that they have confined themselves to broadcast and professional sports videos. We wish to throw light upon this gap, and specifically bridge this gap by stating research works as well as implemented solutions which work equally good for any sort of sports video, be it a simple backyard video, or a professional sports video. We believe that our findings can advance the field of research on sports video analysis for all kind of videos.

Index Terms— Sports Analytics, Computer Vision, Video Processing, Image Processing, Video Summarization, Object Analysis, Event Detection, Content Exctraction.

1 Introduction

The stereotypical way in which sports analytics was carried out, which is still in use in different formats, is that a statistical sheet of numerical statistics and events was generated manually. This sheet was then decoded by someone or a "middleman" who had in-depth knowledge of numerical analysis patterns, as well as the sport being played. As technology has advanced over the years, data-collection has become more precise, in-depth and much easier than before. In the last two decades, the advancements in data-collection have influenced the field of sports analytics, providing advanced statistics and sport-specific technologies which allow analysis through videos and provide accurate game-like simulations prior to play.
Research interest in sports video analysis has substantially increased over time, because of the rapid growth of video transmission over the internet, due to availability of 'smart devices' and the ease with which a video can be shot and uploaded on social media.

1.1 Scope and Organization of Paper

Combining sports analytics with video analysis is something that has been addressed before, and is taking shape over the last decade. In this paper, we reviewed various methods of sports video analysis and provided equivalent and feasible solutions even for non-professional or non-broadcast videos. These techniques involve understanding and arranging the video content on the basis on intrinsic and semantic concepts.
We focus on sports data analysis from videos applied in all possible formats of videos, and the organization of our paper goes in the following manner:-

1. We survey all the previous surveys and taxonomies which were conducted on a similar topic and pry on some similar grounds when it comes to video analysis.
2. We review all the designs and technologies mentioned in the previous papers with respect to a definite format, categorizing the high level semantic knowledge into three groups – object, event, and content-oriented groups.
3. We then discuss the prominent challenges identified in the literature and suggest the potential directions for future research in sports video analysis.

The remainder of this paper contains a conclusion, which summarizes everything, and in brief, talks about the research gap, after which, we acknowledge all the professors, faculty members and our esteemed University for all the help that they've provided. Lastly, we have mentioned all the references on which this paper is based on.

2 Surveys and Taxonomies

Since 2000, sports video analysis has immensely drawn research attention, leading to a rapid increase in research implementation work and surveys in this field. In [1], a preliminary survey of sports video analysis was conducted to examine various research topics such as tactic summarization, highlight extraction, computer-assisted referral, and content insertion. This survey listed out all the comprehensive methodologies with an interactive user input, for strategic and tactical summarization, as well as decision making systems that assisted the umpires or referees before judging any controversial event.
In [2], the authors highlight how detailed game logs obtained through tracking technologies in addition to physiological training data collected through sensor technologies have become available for research. At the same time, the sheer amount of data can become another issue arising from the opposite end of the spectrum. Hence, this paper discusses how

- *Aditya Sankolli, Amritanshu Ray, Apurva Shrivastava, Savani Nimonkar are currently pursuing bachelor's degree program in computer science and engineering at MIT World Peace University, Pune, India.*
- *The research was conducted under the guidance of Prof. Snehal Umare of MIT World Peace University, Pune, India.*
- *PH: +91 8408885356. E-mail: adityasankolli@gmail.com*

ISSN 2229-5518

big data and modern machine learning technologies may help to address these issues and aid in developing a theoretical model for tactical decision making in team sports. The paper shows usage of a combination of historical data (such as W-D-L, form, etc), internal parameters (such as tactics, technique, psychology) and external parameters such as home vs way, rank, refrees). Using spatial tracking data, Kihwan et al [3] applied a temporal kernel method to predict the location of the ball on the pitch.

In [4], the major focus is on usage of different types of AI in a variety of sports, and the reasoning behind it. The purpose of the analysis is sports performance analysis and injury risk estimation. Method used was systematic searching through the PubMed, Scopus, and Web of Science online databases for locating articles reporting AI techniques or methods applied to team sports athletes. Artificial neural networks, decision tree classifier, support vector machine, and Markov process were the most frequently used artificial intelligence techniques according to this survey. Team sports like soccer, handball, volleyball and basketball were the dynamic sports considered with more applications of artificial intelligence. For injury risk assessment, the artificial neural network, decision tree classifier, and support vector machine have been used in soccer, basketball, American football, Australian football, and handball.

Moving on from all the spatiotemporal approaches, H. Shih in [4] mentions all the methods which go across this topic from a content-based point of view. This includes all "context-oriented" or "content-aware" techniques of analyzing a broadcast sports video. The author has introduced some great concepts such as content-pyramid, which we have included in the coming segments. His survey has thrown light upon deriving content from sports videos and basing all possible methods and techniques around deriving this content. It focuses methods to extract information from professional level broadcast videos, and performing all sorts of "content-aware" analysis on those videos through all the audio-visual features and cues.

3 Survey of designs and technologies

Considering the working of a video based sports analytics system, it can be divided into three major components – object related processes, event related processes and content related processes.

3.1 Object Related Processes.

When the input is just the video, we first need to extract data from the video in order to go ahead with performing analysis. So, the first step is to identify and detect objects. After this, various types of information can be extracted by tracking these objects, classifying different types of objects, comparing the position of objects related to each other, comparing types of interactions of objects with each other, etc. This step will create a sort of data which can then be subjected to further analysis.

The idea of representing video information in terms of its content is one of the most important multimedia applications that have been developed through computer vision and other technologies within the last two decades. In particular, object-based representation consists of decomposing the video content into a collection of meaningful objects. [5]

As suggested, an object is a physical entity which is captured in the video, which is 'meaningful'. The word 'meaningful', is contextual, which basically means that the entity must have some importance with respect to the video, or in this scenario, with respect to the sport being played. for eg) In a Soccer video, objects would be players, ball, goal, etc

3.1.1 Object Identification

Object Identification is a computer vision technique that allows us to detect objects/physical entities from the video based on specific features given to the algorithm. These features can be as simple as colour and shape, or high level features on which a model can be trained upon using deep learning and other techniques that complement computer vision.

In [6], a combination of HMM (Hidden Markov Model) and GMM (Gaussian Mixture Model) is used to correctly detect players from a given broadcast video, or a simple sports video from a pan-tilt camera. Combination of HMM's along with Kalman filter is a good solution for identifying and tracking objects from videos.

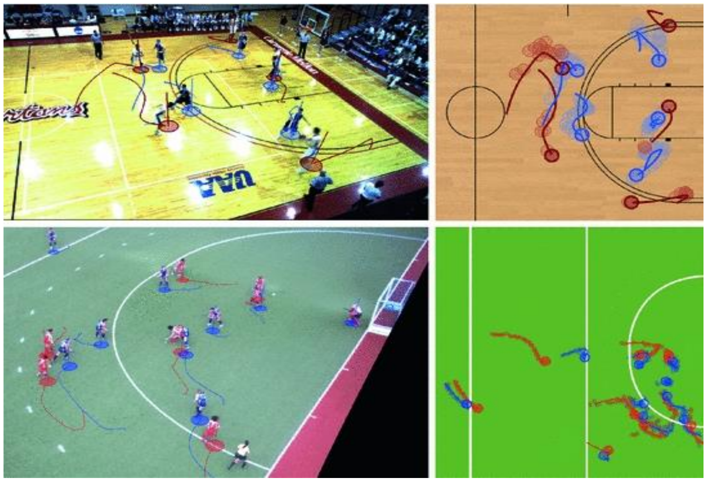
In [7], an efficient five-step approach is used for all object related processes, out of which, the first three steps focus on object detection. These are very generic when it comes to any 'field' game or sport, namely – court detection, individual detection, and colour classification. Morphological operations in computer vision are used to find lines of court, which is really helpful for mapping the court on a 2-dimensional graph, as well as for filtering out the noise, which in this case is people not involved in the game being played. Individual detection uses models to detect human entities in the video, and based on the colour of their jerseys, teams are separated in the third step.

Fig.1. Individual Detection. [7]

3.1.2 Object Tracking

Object tracking is basically taking the initial set of detected objects, and tracking their movement throughout the video creating a sort of trajectory which can be used for further analysis. The idea is to successfully locate and track the same object throughout all the frames of the video. The major limitation in tracking objects in sports analytics is the camera motion. In [8], the authors applied a region-based detection algo-

ISSN 2229-5518



rithm for eliminating fast camera motion effects in goal scenes and tracking soccer players. When the region-based algorithm is used to track players, either the template matching method or split-and-merge approach is applied for occlusion reasoning between players.

Fig.2. Tracjectory Tracking [10]

Tracking objects is difficult generally when the objects interact with each other too much, or when objects overlap and are not completely seen. In this respect, in [9] the authors have used the ZXY data directly in combination with video to follow individual players and parts of the team directly through the help of a particular camera calibration and body sensors for all individuals. The sensor data can then be used as a benchmark to compare the accuracy of tracking algorithms applied just on the videos.

In [10], a hierarchical approach is adopted to track all the trajectories, and divided into three groups according to hierarchy, that are low-level, mid-level and high-level trajectories. All these trajectories are extracted using different algorithms and then compared to provide accurate results. Low-level trajectories are considered to be a result of noisy detections, the mid-level trajectories are obtained via Hungarian algorithm, and the high-level trajectories are a result of cost-flow networks. The final result of tracking these trajectories can be seen in the image above.

3.1.3 Object Naming/Labelling

Object Naming or Object Labelling is the process of creating a unique id for every object that is detected in the frame, and maintaining this id throughout all the frames of video. Depending on the quality of the video, and professionalism of the sport, there are two ways to go across this process. The similar part in both things is that the inanimate objects in the video can be simply named by their common names, without getting into any complicated indexing or reference. That is, objects like ball, goal, basket, etc can be named as they are.

The first way is for videos that are shot through any simple device by anyone, which can be a simple backyard video or a high school field game. In this situation, teams can be separated through colours, and unique ids can be created according to team colour and a simple natural number index (e.g. White Team Player 1, Red Team Player 2, etc). The second way is for professional level broadcast videos. Frameworks are used in such situations which integrate two procedures, 1) time-stamped character annotation by aligning subtitles and transcripts and 2) face tracking and speaker detection.

In [11], the use of Masked R-CNN is encouraged for detecting players and assigning them indexes according to team colours, which works like a charm after pre-processing the video through algorithm modules like court detection and background removal.

A bidirectional model to simultaneously assign names to the trajectories in the video which were mentioned in the text was applied and presented in [12]. It's comparatively easier when it comes to videos from news or movies than sports videos, due to stable face images as well as subtitles can be easily obtained.

3.1.4 Action Recognition

Humans can easily recognize and identify actions in a video but automating this procedure and making a machine or a system do it is challenging. Human action recognition in video is of interest in recent times for applications such as automated surveillance, elderly behavior monitoring, human-computer interaction, content-based video retrieval, and video summarization.[13]

Action recognition has been studied in-depth over the recent years, but only a few of the studies focus on the sports genre. Action recognition in sports basically refers to actions of the individuals, the actions concerning the interaction between two individuals, two objects, or an individual and an object, or the interaction between a group of individuals with each other. Considering a generic domain not focusing on just sports videos, but in general all videos, Stephen in [14] focuses on use of AlphaPose to detect a human body within an image and provide a full description of a human pose. Alpha Pose is the "first real-time multi-person system to jointly detect human body, hand, and facial key points on single images using 130 key points. [15]

Considering the sports domain, we can further simplify action recognition in the following manner – posture and movement, feature reperesentation, classifier learning and use of efficient datasets. In [4], various sources of good datasets are mentioned and a good amount of good datasets for action recognition are also mentioned. There are in total seven datasets in various formats mentioned in this survey, from various sources, and in various formats. For e.g) Complete frame images, vectors, mini video clippings, actual video data, etc.

Feature Representation and classifier learning are the two most critical steps in the learning phase when it comes to action recognition. The idea is to extract posture and movements as feature vectors which can be done through deep learning algorithms (Body Pose Estimation, etc), and once the feature vectors are extracted, a classification framework can be used to recognize the type of action, using the available datasets.

In [16], sports videos are classified with an integrated combination of deep learning models and transfer learning. The basic idea is to categorize and classify the type of sport being observed by the algorithm/system as the input, based on rele-

vant and specific sports actions which contain spatial and motion features.
In [11], the author suggests use of R-CNN with Body Pose Estimation and other deep learning algorithms for a desired specific purpose of actions. The idea is to create a predefined set of actions, and then compare it to the video at hand and extract all the actions through pattern matching algorithms.

3.2 Event Related Processes

After detection of objects and their relation with each other, specific events can be defined based on the apparent position of objects, action recognition, play breaks, scene change detection, etc. After defining these events, these can be extracted for summarization and other related applications
In the context of sports analytics through video, an event is a user-defined or classifier-defined situation which occurs in the video, which is basically an important terminology or incident, with respect to the sport being played in the video.
For eg) In a Soccer Video, goals, shots, substitutions, saves, great passing moves, fast counter attacks, etc. are some examples of possible events.

3.2.1 Shot Transition Detection

An event is made up of a series of video scenes or shots. The shot boundary detection technique aims to divide a long video sequence into video segments. There can be various possibilities of a frame switch especially when it comes to broadcast videos. In case of simple videos shot on an unprofessional level, we cannot categorize it much further, but can still be subjected to great analysis. For Professional level videos, the possibilities of frame switch can be: (i) A camera switch in a multi-camera system broadcast, (ii) Video paused due to end of phase and then resumed again (during half time or breaks, and the final video being one continuous video), (iii) Infomercial break, (iv) Technical issue, etc.
The only possibility for normal videos is that the video was paused because of the end of a phase, and then resumed again, as this domain contains all possible devices that can record videos. So, in most situations for normal videos, the frame switch is considered to be an end of phase, assuming that the video was paused. Thus, a shot transition or a scene change is always considered to be informative, and it's necessary to detect these when extracting events.
In [17], the complete focus is on broadcast videos. Thus, the camera switch in a multi-camera set up is a possibility of a frame switch. The author has mentioned two such scenarios which can be considered as a scene change detection considering the above possibility. First one being a camera switch from a wide angle camera to a close up view of some part of the field where the action is going. This can be detected by a sudden change in the number of entities/objects in the frame, thus, a frame switch or scene change. Second one was slightly specific to the application, which is detection of animation. As we all know, that when there are slow motion highlights, expert reviews, or other related things, there's some sort of animation, like the broadcast channel logo popping up, or a transition animation, which can be detected, and is an efficient indicator for detecting scene change or frame switch.
In [18], a combination of deep learning models is used for tracking ball in a soccer video. The method to detect frame switch from this, would be a discontinued trajectory of the ball, due to pause and resume of camera, or change in camera angle. All the player tracking algorithms depend on these discontinuous trajectories. A frame switch detected, can be very important when it comes to sports analytics, because it provides some or the other kind of information. For eg) end of phase, highlight, technical issue, timeout, review, halt in the game (because of ball out of play, spectator invasion, etc.), substitution, etc. This information can be used to further recognize and extract important events from the sports video.

3.2.2 Play and Break Analysis

To design a generic model for analysis all types of sports videos is difficult, because the 'event' is different for every sport. Thus, the concept of play-and-break analysis was presented, which is a more compromised approach. Event being a domain dependent concept, it needs to be defined separately for every single sport. Play and break analysis divides these events into “plays”. “Plays” is a concept which leans purely towards sports. A set of events which have some sort of tactical or strategic importance as well as those which can be used to reflect the performance of a team or individual, can be termed as a “play”. Basically, those events a manager or coach would like to watch over and over again in order to perform better analysis and come up with better strategies, as compared to those which don't hold much importance from an analytical point of view, can be considered as “plays”. The definition is very subjective, and can be molded around, based on the purpose and application. Moving over without any priority or hierarchy, we can classify play as the part of video without stoppages. That is, when the sport is actually being played. And the other phases, like timeouts, or when the ball is out of play, is considered the break phase. The basic idea in this segment is to focus more on the play phase, and categorize it according to intended purpose.
Data analyzed from this sort of analysis, processes the video, and creates a list of plays in chronological order. This can then be subjected to further analysis, such as categorizing the plays, prioritizing the plays based on excitement levels, team advantage, individual skills, or other momentum building/shifting plays.

3.2.3 Key-frame Determination

After having a good data based on ‘plays’ and events, The next step would be to index these plays and classify them into various categories, and prioritize them according to importance and consequences with respect to the performance (of the team or individual) and the final result. Key-frame determination is a crucial operation performed after the retrieval of “plays” from the previous step. Key frame extraction is a powerful tool that implements video content by selecting a set of summary key frames to represent video sequences. A key-frame can be identified by the timestamps of the start of a play to the end of a play, also taking into consideration, overlapping of plays, that is a play can be started before the previous play ends.

PlaySight is a known solution for such type of analysis and determination. It determines key-frames from sports videos using a predefined set of rules and datasets. Thus, using a rule-based-inference in combination with multiple deep learning and video processing modules, indexing and analysis of keyframes is performed. [19]

In [20], multiple methods and applications of key-frame determination are mentioned and applied. Its major focus is to represent adequately and fast in the form of important key-frames.

3.2.4 Highlight Detection

"Highlights" can be defined as the most important plays or events, which are directly related to the performance of teams and individuals, and the final result. One of the most important application in the sports industry these days, is to convert a full game video, into smaller snippets of "highlights" which can be published for business aspects, improving fan experience, as well as can be analysed strategically from a coaching and managing point of view. For eg) Algorithms are developed to convert a 90 minutes soccer game, into a 2-3 minute highlight video, which consists of all the outstanding and deciding moments from the match.

Highlight extraction requires specific features and analyzers. A video consists of audio and visual features. Several previously proposed frameworks have applied audio features instead of classifiers, such as SVMs, HMMs, Bayesian belief networks (BBNs), DBNs, maximum entropy models, and multilevel semantic networks, which are an extension of the BBN. [4]

In [21], a method for detecting and extracting replays and slow motion replays from broadcast videos is mentioned. This is a type of highlight extraction, as broadcast videos always replay all the snippets which are important and decisive when it comes to the flow of the match, performance of the teams, and final result. The authors have used this method in one of their case studies concerning NFL videos, which is a dataset of American football videos. This was achieved by the additional replay-detection module, which converted the video into three layers, which are the original video (plays and replays), just the plays, and just the replays (highlights), which were then processed for highlight detection, using the third highlight layer.

In [22], the main focus is on normal videos shot from simple portable and wearable devices. Using an integrated framework of deep learning modules and audio-visual features, a separation process is performed, which divides the video into two parts, highlight part and the non-highlight part. The paper categorizes 15 unique sports categories based on their domain knowledge, although the highlight detection using this method is confined first-person videos using wearable devices like GoPros, etc.

In [23], the authors emphasized the framework of mining event patterns. They applied temporal pattern analysis for semantic event mining and highlight detection, by using a group of audiovisual features and domain knowledge of the sport.

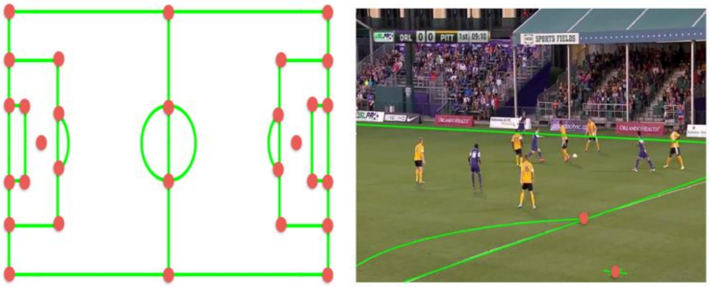
3.3 Content related Processes: -

All previous surveys focused on a definite type of methodology to extract content from sports videos, but they were only confined to broadcast videos as they could use all the audio visual cues as well as a sort of text analytics through the captions provided by the broadcast channels and the commentary. In latest and recent survey papers about this topic, this sort of methodology or approach was called the content-aware approach. Even though this approach did introduce many methods concerning sports video analysis, most of them still focussed on content-extraction from professional level broadcast videos. To find methods that work equally well for normal and professional videos or finding methods which can be applied on normal videos that are equivalent to methods applied on broadcast videos is the greatest challenge in this domain. Considering the number of videos in general, obviously normal videos are going to be in much higher numbers, as compared to actual professional level broadcast videos, and according to us, all these videos can provide quality content for improving team and individual performance.

A comparatively costly solution to this problem is provided in [9] which encourage the use of body sensors for every player on the soccer field, and a particular camera angle and calibration. This allows mapping all the action on the field on a 2d graph, and quality content can be derived from the body sensors, like distance traveled, endurance, velocities over time, positioning, and many more. These types of stats which are extracted can be used to provide great content, as they can be used to analyze the strategies and tactics used by the teams, as well as provide individual ratings to players based on their previous plays, or universal constants.

In [24], a multi-camera set up solution for converting simple practice session and training videos into professional level videos is provided and applied. It's a connected multi-camera technology which has smart trackers, that captures all action in the "SmartField", to create a pro-level broadcasting experience without the need of an operator. They also have their own activity cloud which can save all the videos recorded through this system which eventually improves further analysis and accuracies. The interface is also user-friendly and interactive, through which we can rectify mistakes by the algorithm if any. Thus, any normal video can be converted to pro-level broadcast video by using this system, and then subjected to analysis as per a broadcast level video.

In [25], a method to figure out what part of the field or court is being captured by the camera, which can be used to map the action on the field on a 2d graph for extracting content is applied. This can work for any sort of video, but works much better if the camera angle is such that the camera is set up on the longer side of the field. By simply extracting lines and points detected from the video through morphological operations, what part of the pitch is being filmed is calculated through combination of methods called ground truth annotation. This can lead to analytical content extraction when it comes to strategies and tactics. Although we can only watch some part of the pitch, so, to map every single player on the 2d graph is not possible, we believe that there is enough content to be extracted even from the parts which are being recorded, that is the part where the action is taking place.



(a) (b)

A concept called "Logical Story Unit" was introduced in [26] which can be defined as semantically related consecutive series of image frames that depicts and conveys a high-level concept such as event, topic, object, location, and action, which constitutes a story in a video. Especially, an event can be defined as an incident or situation, which occurs in a particular place during a particular interval of time, for example – home-run in a baseball game, actor's entrance on stage, car explosion on a highway, etc. Under these definitions, video scene and event detection is to find all video intervals corresponding to a specific event from a given video. Thus, given a context "story" or a predefined "event", it can be used to extract quality content from any video, and can be used for sports video of any level if and only if we have enough video data from which a database can be made, and events can be predefined from this video data. The applications are vast.

4 Challenges and Future Directions

Over recent years, there have been many research achievements taking into consideration the field of sports analytics, from a video point of view as well. These surveys and researches have highly influenced the direction of prospective research, and advancements on daily basis have created a sort of vast scope in this topic. We wish to highlight all the prominent challenges, which will surely be surpassed over the coming years. Hence, we wish to term them as potential future directions, more than "challenges" or "research gap". After going through all the literature, we have categorized the prominent challenges and future directions in the following manner:-

4.1 Creating a Generic Model

Considering all the techniques mentioned above and taking into consideration object, event and content related processes in sports video, it takes a lot of effort, space, and time to create a single working module and to implement it on a specific domain. As all the research papers surveyed and reviewed focus on only one aspect, it can give us a fair idea about how long it takes and how difficult it is to create and implement just a single module on a specified domain, be it just player tracking, highlight extraction, play-and-break analysis, etc. The idea of creating a generic model that will calculate everything and perform all known procedure and give an accurate and precise report of a sports video, is highly improbable currently. But we know that times are changing and technology is improving and being updated every single moment, and hence, the problems that currently arise, like high processing time if multiple modules run simultaneously on the video at the same time, huge space requirements for creating a database to improve accuracy, integration of all modules with each other, and many more obstacles, may actually be resolvable in the near future. We already have solutions that have integrated and interfaced a smaller number of such modules, but multiple modules to get a specific purpose out of a sports video. It seems quite improbable, but it's still possible even today with years and years of tedious work. But we'd positively like to believe that the technical advancements over the years and development of new technologies and platforms will be a great catalyst for this and will make things easier in the near future.

4.2 Real-Time Application

The whole purpose of any sports analytics related application is that it should be able to help the team or individual to improve, and a collective improvement of everything including business and fan experience when it comes to big professional sports teams. All the systems previously worked on, when it comes to sports analytics through video, are those which take inputs from previously shot videos and perform analysis on those videos. The biggest drawback of this is that it is very difficult to create a system that will analyze these things in real-time, and perform analysis instantly. With the possibility of great decision making systems due to advancements in AI and ML, there lingers a possibility of having such systems assist coaches/managers/advisors/teachers real-time. Calculating counter measures, performing analysis and formulating the best way/strategy for an individual to get a particular outcome is something many researchers have focused on, but doing this in real-time is still a comparatively unexplored area. All these things are done manually, because in the end what any system needs to perform analysis is numbers, but while doing these things real-time, there are many factors which cannot be calculated by numbers. Examples of these factors are morale, atmosphere, momentum, weather conditions, motivation, player's mental/emotional/physical condition, internal conflicts, are the things which one can't calculate through just the video. So, the final decisions are always left to the in-charges, as we cannot take the above mentioned things into consideration, but, someone physically present there can account for these things and has a fair idea of what can be done. Hence, the idea is not to completely get rid of the manual input, but some real-time decision making system that is meant to make things easier for the person in-charge and "assist" that person, and not replace him. Even today, all these things are done by the "coaching staff" when it comes to big clubs, and in all unprofessional situations, only done by the person in charge or the players on bench. The major issue is processing time, as these algorithms take up a lot of time to perform analysis, but we believe that this issue will surely be resolved in the near future.

4.3 Sports Video Analysis on Media Clouds

Previously, video recording was a very specific task which needed a camera which was not so easily available. But recently, almost every person uses at least one device which has a camera and an active internet connection, such as camera handset, personal video recorder, smart phone, tablet, laptop, etc. Consequently, any user of these devices is a potential content producer. Users are interested in sharing their videos through social networks, and this has become a cultural phenomenon. A record of professional broadcast videos is always stored in the archive clouds of those particular leagues. But we've established before that there are clearly more unprofessional sports videos, than professional videos. A simple logic behind this assumption would be the fact that there are many people who "like" or are "associated" to a sport, than the people who actually make it to a professional level. So, these unprofessional videos can be great sources for analyzing a sport. Even in terms of sports, many new things and techniques are introduced, and one may not always have enough video data to catch and detect all these new things from a video. This issue can be solved by analyzing all the videos that are present on media clouds. Analyzing these and storing these into databases for better and robust analysis is a potential direction in this domain. This will keep all the machine learning and deep learning algorithms up to date, as all the latest trends will also be uploaded on the media clouds. The accuracy will also be improved as for machine learning and AI purposes, more training data will be available. Even for extracting robust visual features, and for precise human action recognition, having a reservoir cloud of huge ever-increasing data, will boost the accuracy of further analysis.

Conclusion

We conducted a survey and we went across research papers related to our topic, that is "Sports Analytics through Video". We also went over some systems and methodologies which are currently used in major sports events. An overview of these methodologies was categorized into three groups, based on all the techniques that were previously implemented. These three classes were object, event and content based processes. We compared the stereotypical approaches that analyzed broadcast sports videos, to unprofessional sports videos, and gave equivalent solutions for efficient and precise analysis of both. We believe that our survey can advance the field of research on sports video analytics.

Acknowledgment

We express a deep sense of gratitude towards our esteemed guide Prof. Snehal Umare for giving us this splendid opportunity to select and survey this topic and also for providing facilities for successful completion.

We thank Dr. Mangesh Bedekar, Head, Department of Computer Engineering, for opening the doors of the department towards the realization of the topic, all the staff members, for their indispensable support, priceless suggestion and for most valuable time lent as and when required. With respect and gratitude, we would like to thank all the people, who have helped us directly or indirectly, and our esteemed University, Dr. Vishwanath Karad's MIT World Peace University, Survey No, 124, Jijau Masaheb Marg, Kothrud, Pune, Maharashtra 411038.

References

[1] J. R. Wang and N. Parameswaran, "Survey of Sports Video Analysis: Research Issues and Applications," in Proc. Pan-Sydney Area Workshop Vis. Inform. Process., 2003, pp. 87–90.

[2] Rein, R., Memmert, D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus* **5**, 1410 (2016). https://doi.org/10.1186/s40064-016-3108-2

[3] Kihwan K, Grundmann M, Shamir A, Matthews I, Hodgins J, Essa I (2010) Motion fields to predict play evolution in dynamic sport scenes. Paper presented at the IEEE CVPR. 13–18 June

[4] H. Shih, "A Survey of Content-Aware Video Analysis for Sports," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 5, pp. 1212-1231, May 2018, doi: 10.1109/TCSVT.2017.2655624.

[5] Andrea Cavallaro and Touradj Ebrahimi. "Object based video: extraction tools, evaluation metrics and applications". Signal Processing Institute, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland.D.S.

[6] Lu WL, Ting JA, Little JJ, Murphy KP. Learning to track and identify players from broadcast sports videos. IEEE Trans Pattern Anal Mach Intell. 2013 Jul;35(7):1704-16. doi: 10.1109/TPAMI.2012.242. PMID: 23681997.

[7] Cheshire, Evan, Cibele Halasz, and Jose Krause Perin. "Player tracking and analysis of basketball plays." *European Conference of Computer Vision*. 2013.

[8] H. Goto, Y. Hasegawa, and M. Tanaka, "Efficient Scheduling Focusing on the Duality of MPL Representation," *Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS '07)*, pp. 57-64, Apr. 2007, doi:10.1109/SCIS.2007.367670. (Conference proceedings)

[9] Pettersen, Svein & Johansen, Dag & Johansen, Håvard & Berg-Johansen, Vegard & Gaddam, Vamsidhar & Mortensen, Asgeir & Langseth, Ragnar & Griwodz, Carsten & Stensland, Håkon & Halvorsen, Pål. (2014). Soccer Video and Player Position Dataset. Proceedings of the 5th ACM Multimedia Systems Conference, MMSys 2014. 10.1145/2557642.2563677.

[10] J. Liu, P. Carr, R. T. Collins and Y. Liu, "Tracking Sports Players with Context-Conditioned Motion Models," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, 2013, pp. 1830-1837, doi: 10.1109/CVPR.2013.239.

[11] "A 2019 Guide to Object Detection" by Dereck Mwiti

[12] V. Ramanathan, A. Joulin, P. Liang, and L. Fei-Fei, "Linking people in videos with 'their' names using coreference resolution," in Proc. Eur. Conf. Comput. Vis., 2014, pp. 95–110.

[13] Shabani, A. H., D. A. Clausi, and J. S. Zelek, "Improved Spatio-temporal Salient Feature Detection for Action Recognition", *British Machine Vision Conference*, University of Dundee, Dundee, UK, August, 2011.

[14] "Sports Video Analysis using Machine Learning" by Stephen Janssen, published on December 16, 2019. (https://www.linkedin.com/pulse/sports-video-analysis-using-machine-learning-stephan-janssen)

[15] AlphaPose by Shanghai Jiao Tong University, Machine Vision and Intelligence Group(MVIG) official website: https://www.mvig.org/research/alphapose.html

[16] Russo, Ashraf & Kurniangoro, Laksono & Jo, Kang-Hyun. (2019). Classification of sports videos with combination of deep learning models and transfer learning. 1-5. 10.1109/ECACE.2019.8679371.

[17] Lu WL, Ting JA, Little JJ, Murphy KP. Learning to track and identify players from broadcast sports videos. IEEE Trans Pattern Anal Mach

Intell. 2013 Jul;35(7):1704-16. doi: 10.1109/TPAMI.2012.242. PMID: 23681997.

[18] Kamble, P. R., Keskar, A. G., and Bhurchandi, K. M., “A deep learning ball tracking system in soccer videos”, *Opto-Electronics Review*, vol. 27, no. 1, pp. 58–69, 2019. doi:10.1016/j.opelre.2019.02.003.

[19] PlaySight official website: https://playsight.com

[20] Yunyu Shi, Haisheng Yang, Ming Gong, Xiang Liu, Yongxiang Xia, "A Fast and Robust Key Frame Extraction Method for Video Copyright Protection", Journal of Electrical and Computer Engineering, vol. 2017, Article ID 1231794, 7 pages, 2017. https://doi.org/10.1155/2017/1231794

[21] “Event Detection and Summarization in Sports Video” by Baoxin Li and M. Ibrahim Sezan Sharp Laboratories of America 5750 N.W. Pacific Rim Blvd. Camas, Washington 98607, USA {bli,sezan}@sharplabs.com

[22] T. Yao, T. Mei and Y. Rui, "Highlight Detection with Pairwise Deep Ranking for First-Person Video Summarization," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 982-990, doi: 10.1109/CVPR.2016.112.

[23] M. Chen, S.-C. Chen, M.-L. Shyu, and K. Wickramaratna, “Semantic event detection via multimodal data mining,” IEEE Signal Process. Mag., vol. 23, no. 2, pp. 38–46, Mar. 2006.

[24] PlaySight official website: https://playsight.com

[25] Jianhui Chen, Fangrui Zhu, James J. Little. (26 Jan 2018) “A Two-point Method for PTZ Camera Calibration in Sports”. WACV 2018 accepted. Computer Vision and Pattern Recognition. Cornell University.