MI

MOTION IMAGING JOURNAL

COVERING EMERGING TECHNOLOGIES FOR THE GLOBAL MEDIA COMMUNITY

SMPTE

TECHNICAL PAPER

Sports Production Through AI-Powered Sports Action Tracking and PTZ Cameras

By David Edwards and Siddhi Imming

Tradicionalmente, as transmissões esportivas (NBA/NFL/Copa do Mundo) dependem de valores de produção significativos para atender às expectativas do público. No entanto, produzir conteúdos para esportes menos populares ou com orçamentos reduzidos permanece um desafio. É ai que a tecnologia de IA, em conjunto com câmeras PTZ (Pan-Tilt-Zoom), está sendo utilizada para oferecer produções esportivas de qualidade com custos reduzidos. O artigo desta edição descreve como um sistema híbrido, que combina uma câmera panorâmica para monitorar toda a cena e uma câmera PTZ de alta qualidade, pode ser usado para capturar a ação de forma precisa e imersiva. O sistema utiliza algoritmos de IA para rastrear movimentos em tempo real e prever a posição da ação esportiva, ajustando automaticamente o foco das câmeras para fornecer uma cobertura otimizada. Desafios como latência no processamento de dados, diferenças de perspectiva e limitações de resolução, não foram esquecidos e são abordados, com soluções que incluem o uso de sensores de alta resolução e algoritmos preditivos. Realmente o avanço destas tecnologias promete tornar as transmissões esportivas mais acessíveis, mantendo um padrão de qualidade elevado. Boa Leitura!

Tom Jones Moreira

Abstract

Consumer expectation when it comes to the viewing experience is often based on watching tier 1 sports productions. In response, sports programmers are searching for solutions that deliver an engaging experience at a cost that is in line with the available budget. Artificial intelligence (AI)-driven sports production solutions have been available for some time. But to date, the ability to accurately track sporting action combined with visual clarity has been a challenge. This article describes how a panoramic camera system can be implemented to monitor the entire field of play, to feed an AI system that identifies the sporting action. Furthermore, how the AI engine can drive a high-quality pan-tilt-zoom (PTZ) camera to follow the action to deliver the program feed is also described. The article also discusses how the AI engine can accurately point the PTZ camera—taking into account video and AI processing delays, how those processing latencies can be measured, to forward-predict directional vectors and correctly point the program feed camera irrespective of distance from the camera and the variation in angular velocity realized.

Keywords

Artificial intelligence, image quality, pan-tilt-zoom camera, sports action tracking

Introduction

Live sports is a major draw for linear broadcasting with such events holding places in the top ten most-watched programs each year across the globe.[i]

Sports programming delivers high-end production values, achieving a visceral experience that plays to the tribal needs of watching your team win. There are many sporting events that are yet untapped and there is high demand among the public to consume this content.

With expectations of viewing experience set through watching tier 1 sports productions, sports programmers are searching for solutions that deliver an engaging experience for a cost that reflects the budgets available to produce new sports programming.

Automated action tracking in sporting events is a complex task. Simple logical rules cannot accommodate the complexities that arise in following the action played out by sentient humans competing and making rapid decisions to achieve a win. AI technology is, however, well placed to achieve good accuracy with this task.

Artificial intelligence (AI)-driven sports production solutions have been on the market to offer automated production of sports games for some time. But to date, their impact has been limited. The ability to accurately track the sporting action combined with visual clarity has not yet matched the classic production values to which viewers have become accustomed.

To deliver a similar viewing quality experience, the desired sporting action needs to fill the majority of a high-quality, full frame rate camera sensor. Pan-tilt-zoom (PTZ) lenses provide a good solution to deliver visual quality and to automate control of where to point the camera.

With a PTZ camera closely following the sporting action, challenges arise in delivering accurate tracking of sporting action across the entire field of play. As human camera operators do, an automated system must monitor the events beyond the field of view of a single, closely zoomed camera lens.

With data-processing being required to identify the position of the sporting action within a 3D space, the processing period creates inevitable delays. Technical and predictive measures may be utilized to ensure that the PTZ camera is pointed at the correct area.

[i]For example, U.K. top ten TV shows 2022—TV since 1981 | Barb, U.S. top 10 TV shows 2022—Most watched TV shows in the U.S. 2022 | Statista.

AI Action Tracking—More Than Just Follow the Ball

Automated action tracking in sporting events is a complex task. Simple logical rules cannot accommodate the complexities that arise in following the action played out by sentient humans competing and making rapid decisions to achieve a win. AI technology is, however, well placed to achieve good accuracy with this task.

Different sports have many different properties and no single coded or learned behavior can be directly transferred from one sport to another. However, there are similarities within different types of sports—ball games, track races, and so on.

Within ball-based sports games, the best productions are associated with more than just simple follow-the-ball action tracking. The play is complex. The ball may go out of play in one location and may come back into play at a different location. The ball may get obscured from the camera view, while the action proceeds. Different on-field players exhibit different behavior—goal keepers and referees.

The gameplay can often involve a cluster of players in the region around the ball for many sections of play. For other segments in a match, if the ball approaches the goalkeeper, it may be best to ensure that both the ball and goalkeeper remain in shot. Consequently, there can be advantages for the AI director engine to have the ability to recognize different types of players and players from different teams.

For track sports, it is important for the AI engine to distinguish between different on-track participants. A bike race or a horse race may involve assistants who participate in the start or end of the race. The AI engine may need instruction to ignore those people. In short, the AI engine will provide a better, more immersive production if it is context-aware of the start and end of the race (**Fig. 1**).

When is a Ball not a Ball? When it is a Head

Achieving accuracy in tracking the sporting action is of key importance in televising an event. Missing a goal in a soccer game or the winner crossing the line in a track event would be bad. Care needs to be taken to ensure that it can differentiate between key action elements and false objects.

In soccer, some on-screen objects can be mistaken for the ball—the penalty spot, for example, or the head of the coach standing on the sideline. For track racing sports, the sporting action takes place on the track. By including positional rules into the AI director, false positives of sporting action can be reduced and accuracy of action improved (**Fig. 2**).

Quality Problem

Lens Quality Matters

For many years, the general public has been subjected to marketing messages from smartphone and camera manufacturers claiming that more pixels deliver higher quality images. For AI sports systems, as with consumer imaging products, this can be true. With digital camera systems, there must be at least one pixel spacing between two objects for the objects to be resolved as separate elements. For this reason, existing AI sports production systems have moved to high-resolution cameras.

Should the pixel resolution of the image sensor greatly exceed the resolution of the lens, further increases in the number of image pixels can deliver few additional resolution benefits. Driving to ever higher pixel resolution is not the only answer.

There are limits to image quality gains that greater pixel increases can deliver to the benefit of image clarity. The push for higher quality pixel resolution images for AI sports production systems can create its own issues. More image data requires more computational power—potentially increasing image-processing latencies or system costs (**Fig. 3**).

Lens quality can be vital in delivering high-quality images. Lower-quality lenses are likely to deliver softer images containing greater levels of distortion and diffraction effects, which can limit the usability of closely digitally zoomed (cropped) sections of the image.

The Rayleigh criterion, which deals with diffraction effects, expressed with the equation

$$\theta = 1.22 \frac{\lambda}{D} \tag{1}$$

FIGURE 1. In ball sports, machine-learning AI algorithms can detect the area of sporting play as a group of players near the ball.

FIGURE 2. In ball sports, additional work is needed to ensure the AI system can differentiate a ball from a head, or, for example, a penalty spot.

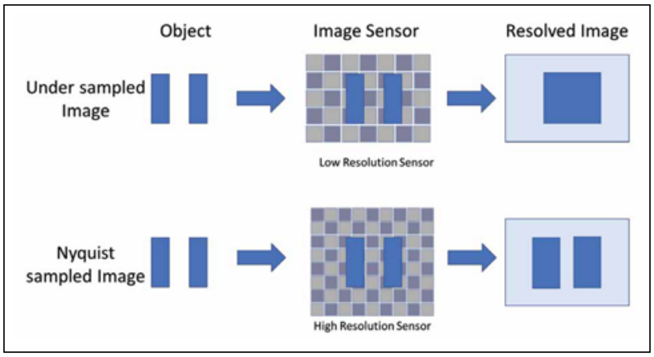

FIGURE 3. Image quality can be improved with higher-resolution sensors.

FIGURE 4. Observing the difference in image clarity between a low-quality and a high-quality lens.

tells us that two objects can be better resolved in relation to their angle of separation θ if the lens aperture or lens diameter is larger.

A larger lens can also allow use of a larger image sensor. A larger image sensor with less densely packed pixels can deliver reduced camera noise and a sharper image, particularly in lower light conditions. With many sporting events scheduled for evening fixtures or inside arenas, it is not always possible to provide TV studio light levels. Low-light performance from the camera system can be important (**Fig. 4**).

Creating a Composite Image

Multilens panoramic cameras equipped with smaller-sized lenses are very capable of delivering individual images that are suitable for being digitally stitched together into a panoramic image.

The lenses can be physically spaced close to each other. With minimal parallax effects, it is possible to stitch together overlapping images to create a single multi-image that does not suffer from large parallax discontinuities—but they can still be present, especially for objects close to the camera (**Fig. 5**).

When fitting larger lenses and larger image sensors, the risk of parallax errors increases since the lenses must be physically distanced with greater separation. As a result, creating stitched images becomes more

FIGURE 5. Stitched multilens panoramic image prior to de-warping. The effect of blending can be observed in the highlighted area.

FIGURE 6. Vislink MVP XCAM multilens panoramic camera.

FIGURE 7. Parallax error—overhanging roof of the stand close to the camera does not give quite the same perspective between two camera lenses in this composite/stitched image.

challenging when seeking to use larger lenses in search for image quality gains (**Figs. 6** and 7). Therefore, different approach is required.

In Search of Better Quality With a PTZ Lens

One can learn from experience gained from years of manned sports productions that make use of broadcast/studio cameras equipped with large image sensors and high-quality optics to deliver sharp, high-contrast images with good color saturation.

The AI-driven solution must find a middle ground between high-end studio cameras and multilens panoramic cameras if it is to produce entertaining sports content without having the expense profile of a complete tier 1 production.

A PTZ lens is an appealing compromise in delivering a larger, higher-quality image sensor capturing content in broadcast native 1080p/50 or 60 frames/sec video formats in combination with a multi-element zoom lens, which together have the potential to provide sharp images even in lower lighting conditions.

However, a single PTZ lens cannot accomplish the task of observing the entire field of play—to identify the progress of the sporting action, and simultaneously provide a close-up shot of the action as it unfurls. An overall view of the entire field of play is also required (**Fig. 8**).

Integrating a PTZ Camera

To deliver on the ability of a PTZ camera to produce AI-controlled high-quality, close action of sports events, the AI director needs to have visibility of the entire field of play.

A hybrid system can be used where a wide-angle/panoramic view is deployed to find the action, while a high-quality PTZ camera can be used to capture the action for the program feed.

Using this hybrid, dual-camera architecture, the AI director, which identifies the sporting action, can operate through the panoramic image and steer the PTZ camera to produce the program feed (**Figs. 9** and **10**).

There are, however, issues that arise from identifying sporting action through one camera and capturing the program feed through a different camera. Data-processing latencies become a critical issue. Time is needed for the image processing to stitch together the multicamera panoramic image to create a composite picture. Time is needed for the AI director to identify the sporting action, calculate the position where the action is taking place, and command the PTZ camera to that location. By this time, the sum of the time taken to process all these actions often means that action has shifted to a different position on the field of play.

FIGURE 8. PTZ lens used by Vislink as part of the AI action tracking system.

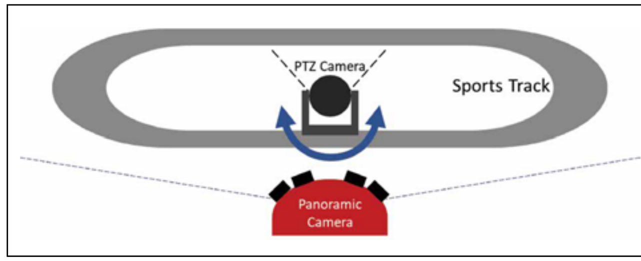

FIGURE 9. Schematic of combining a panoramic camera and PTZ camera.

FIGURE 10. Installation of the combined PTZ camera (top) and panoramic camera (bottom).

Managing Delay in AI-Controlled PTZ Camera Positioning

To ensure that the PTZ camera is pointing at the correct location, to observe the desired object, one needs to relate object positions between the panoramic camera and the PTZ camera.

Positioning of sporting objects detected by the panoramic camera and AI system can be referenced using the spherical coordinates system for a 3D space.

To perform positioning prediction and accommodate AI processing delays, the velocity or motion vector of the object needs to be calculated. Vislink has found that these calculations are best performed in the Cartesian geometry space. A translation from spherical to Cartesian coordinates is performed (**Fig. 11**).

Here,

$x = \rho \sin\varphi \cos\theta$

$y = \rho \sin\varphi \sin\theta$

$z = \rho \cos\varphi$.

There can be challenges in identifying the precise position of the sporting action. With a camera viewing across the sporting arena, there is a spatial variation in pixel resolution per meter between objects close to and far away from the camera. A reduction in the number of pixels per object can lead to inaccuracies or noise in position estimation (**Fig. 12**).

Objects, for example, bicycles on a velodrome racetrack, are viewed with different perspectives and can be interpreted to have different centers depending on their position as they traverse a racetrack. This difference in viewing perspectives can also introduce uncertainties in the object's center, as well as noise into the precise positioning identification and motion estimation (**Fig. 13**).

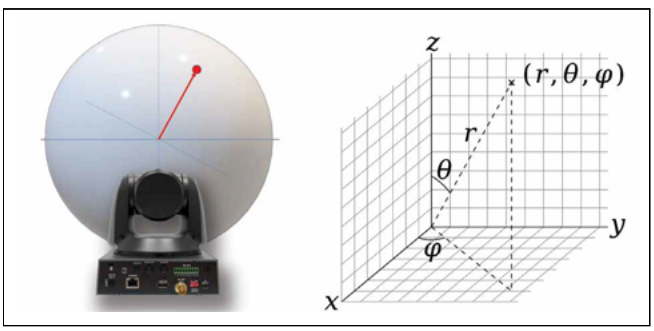

FIGURE 11. Object positioning using spherical coordinates.

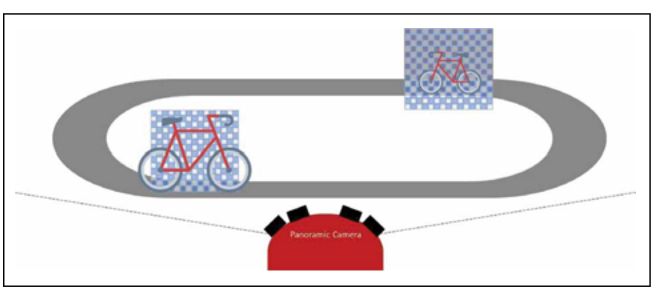

FIGURE 12. Reduced positional resolution for far-away from the camera.

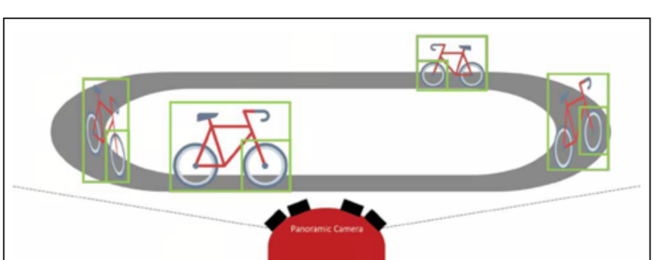

FIGURE 13. Resolving positional accuracy changes with different perspectives as an object moves across the sporting area.

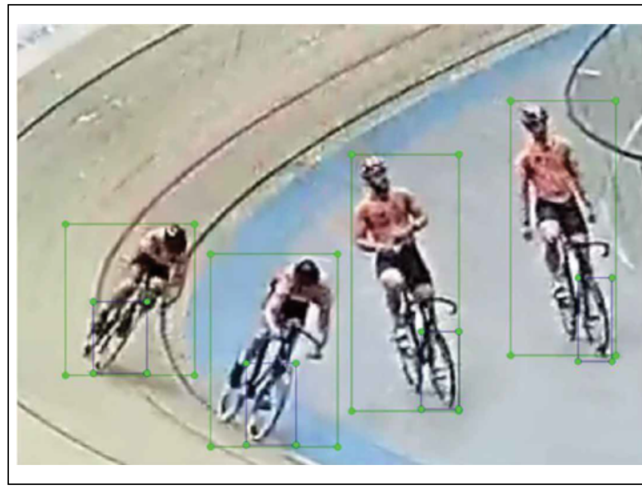

FIGURE 14. Examples of different object positional estimation in response to different perspectives.

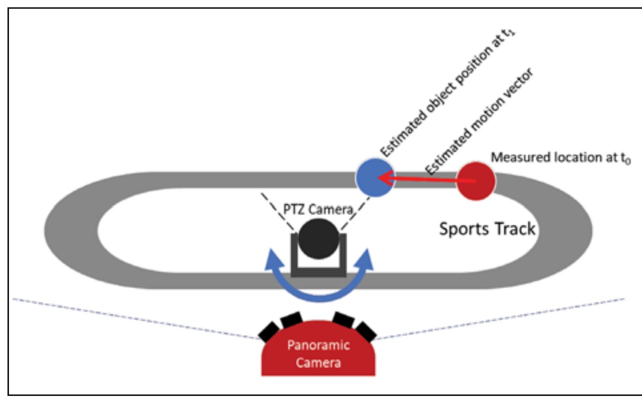

FIGURE 15. Calculating object position for a delay of $T_1–T_0$.

Using a predictive algorithm making frequent positional measurements, it is possible to create a continually updated, accurate estimation of the sporting object motion vectors, which accounts for the "noise" of variations in velocity and inaccuracies in positional measurement from the panoramic camera (**Fig. 14**).

With a knowledge of the inherent delay in the panoramic camera image processing and a noise-free AI-based object detection, one can calculate the position of the sporting object.

Having estimated the position of the sporting object, compensated for the data-processing delays, it becomes possible to map the predicted object location back into spherical coordinates needed for the camera pointing system (**Fig. 15**).

The combination of regularly measured object position—through the panoramic camera, plus a regularly refreshed estimated motion vector—ensures that PTZ camera pointing accuracy is maintained. As a control system, this acts as the measured (delayed) panoramic camera system acting as a feedback loop into the pointing algorithm and the motion vector estimation acting as a feed forward control.

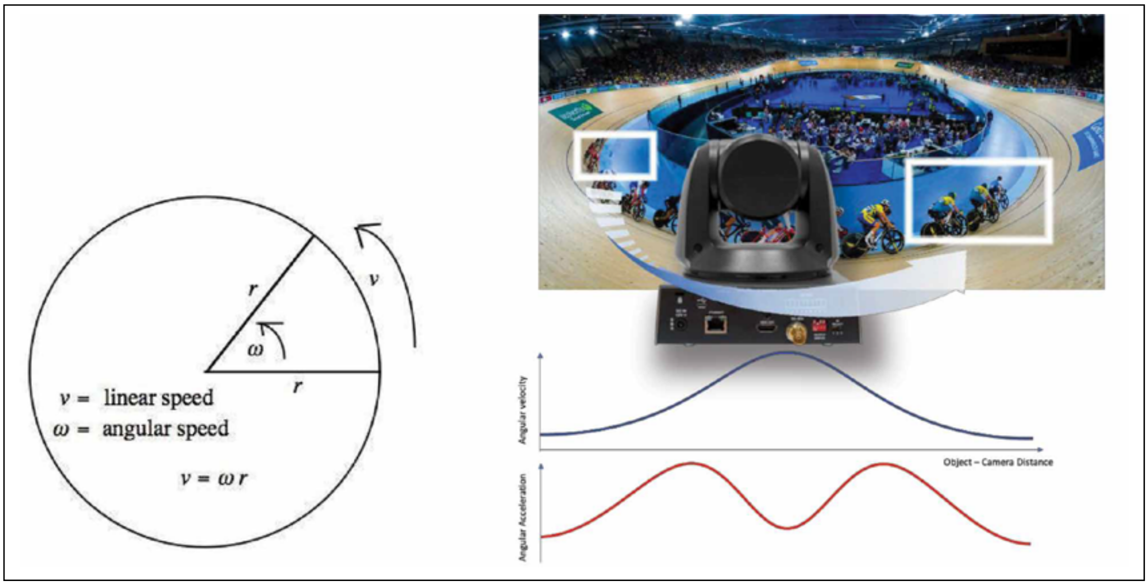

FIGURE 16. Angular velocity and acceleration.

Managing Angular Acceleration and Velocity

In tracking a moving object with a PTZ camera, it is important that the motorized control can keep pace with the action.

Laws of angular velocity and acceleration result in objects moving close to the camera requiring faster tracking than objects further from the camera lens.

To accommodate high-speed tracking, the gearing mechanisms and speed of PTZ tracking motors need special attention. The ability to pan (and tilt) a camera at speed can create difficulties in enabling rapid pan acceleration and maintaining smooth and accurate tracking.

Camera positioning choice can partly alleviate the issue of finding the best balance of pan speed and acceleration. Positioning the camera further from the action can of course reduce the angular velocity that the camera needs to accommodate. For image quality reasons though, closer is often better. In practical installations however, their choice of camera position is rarely a free choice (**Fig. 16**).

Vislink made some changes to the pan mechanism for its PTZ AI solution to get the best balance of tracking speed, acceleration, and smoothness. By transitioning from stepper motor technology to control the pan and tilt of the camera to AC-powered servo motors, it becomes possible to get very fast, low inertia motorized positioning in combination with high positional accuracy (**Fig. 17**).

Should the angular velocity required to capture the sporting object still prove too high for the PTZ camera to perform accurate tracking, the option exists to perform a digital pan across the image plane of the panoramic camera image. One should note that in this situation, where the sporting object has maximum angular velocity, the sporting object is likely to be closest to the hybrid panoramic and PTZ camera system, where the panoramic camera is delivering greatest resolution and image clarity (**Fig. 18**).

FIGURE 17. Vislink pan-tilt camera mount utilizing high-speed, very low inertia ac servo motors.

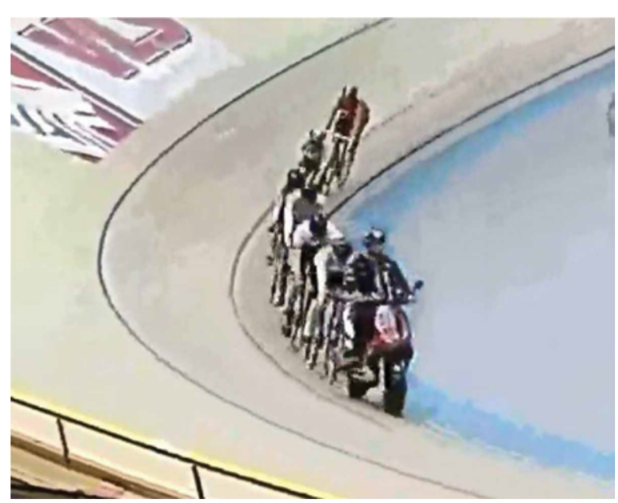

FIGURE 18. Closely tracked sporting action utilizing AI action tracking and program direction, originating from a panoramic camera with the program feed captured by a PTZ camera.

Conclusion

Through machine learning, with guidance from production rules, it is now possible to accurately identify sporting action and mimic the production style of human-directed sports events. However, to create compelling content, the AI-generated production not only has to achieve accurate action tracking, but also needs to deliver good-quality video imaging.

Good-quality lenses that allow the viewer to be presented with a high-resolution close-up of the sporting action are essential-to viewer enjoyment of the production. PTZ cameras are a good price/quality compromise for delivering high-quality images. However, since a PTZ is not capable of simultaneously delivering a close-up of the action and observing the entire area of play, a linkage needs to be made between the program feed PTZ camera and a panoramic view, which the AI engine can use to scan to determine the area where the sporting action is taking place.

With intrinsic delays in a panoramic image and computational delays through the AI action tracking algorithms, issues arise in realtime steering of the PTZ camera to the location of the sporting action. By linking together action tracking and motion vector estimation algorithms, it is now possible to anticipate the location of sporting action in the future.

Sporting action can move rapidly around the field of play. There are challenges in achieving motorization of a PTZ camera that can achieve rapid movement and acceleration of the camera while commanding accurate positional pointing. The use of AC-powered, very low inertia servo motors in place of conventional stepper motors is now being tested as a responsive and accurate camera pointing option.

About the Authors

David Edwards has been innovating in broadcast for more than 20 years, from MPEG-4 HD satellite systems to the revolution in IP architectures. He is now exploring how AI can automate video production.

Siddhi Imming is a software engineer experienced in using image recognition as part of control systems. He has been applying these skills to enhance AI sports production technology.

SMPTE