View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 2 (2010) 3397–3403
Procedia Engineering
www.elsevier.com/locate/procedia

8th Conference of the International Sports Engineering Association (ISEA)

Automated Viewer-centric Personalized Sports Broadcast

Michael Chuang*, Priya Narasimhan

Electrical and Computer Engineering, Carnegie Mellon Univerisity, Pittsburgh, PA, 15213, USA

Received 31 January 2010; revised 7 March 2010; accepted 21 March 2010

Abstract

Mobile handheld devices and online media are currently reaching a stage where it is possible to watch streaming video/TV on them, such as a live sports game broadcast. However, current sports broadcast methods are designed to replicate the same feed to many different TV viewers, which makes it difficult to personalize the feed for each individual viewer. Clearly, since TV viewers have different tastes, this generic broadcast is not likely to satisfy the preferences (e.g., favorite player, specific camera angle) of every viewer. Advertisers already target their audiences with commercials during sports broadcasts in a personalized way (e.g., targeting commercials to specific demographics that are the expected viewers of a game), and it seems reasonable to expect the next step to be the personalization of the sports broadcast feed itself in a viewer-centric way. Adding human directors to create different personalized feeds in the broadcast-production trucks will not scale due to the associated costs. We propose instead a solution that uses a computer program to target the sports broadcast to suit the preferences of each viewer. In this paper, we explore designing an automated computer-driven broadcast director that provides different, personalized automated broadcasts, depending on the viewer's preferences as well as the specific actions unfolding in the game. By using a combination of video analysis to detect the relative player locations and "size" of player templates to predict the next few seconds of game action, along with viewer-specified preferences (e.g., favorite player, favorite camera angle), we automatically select the most interesting camera angle to display to that viewer in real-time. We present the preliminary experimental results of our prototype system using the different video-camera feeds of both an ice-hockey and an American football broadcast.

© 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.

1. Introduction

Every sports broadcast network aims to attract and retain as many fans as possible during a game broadcast because from a commercial standpoint, for every fan watching the game, broadcast networks receive higher ratings, with accompanying increased revenue for the advertisements that they show to these fans. Inevitably, there are diverse kinds of fans watching any given sporting event. There are always the die-hard fans who will never miss a second of the game and who are well-versed in every aspect of the sport. There are also the new fans who are perhaps discovering the sport/game for the first time, and who likely do not comprehend all of the nuances and rules of the sport.

* Corresponding author. *E-mail address*: mchuang@andrew.cmu.edu.

1877-7058 © 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2010.04.164

Figure 1. YinzCam's mobile live streaming of a hockey game from multiple camera angles (iPhone, Blackberry Bold, and Android G1)

Broadcast networks are fully aware that die-hard fans will watch the game, regardless of whatever else the network chooses to show. To appeal to a broader viewer audience, to encourage new fans to engage in the sport and also to ensure that these new fans keep coming back, sports networks try to create a generic broadcast, trying to strike a balance between sports-centric content (to please the die-hard fans and to keep the focus on the sport) as well as non-sports content, e.g., views of celebrities in the audience, personal stories of the players (to attract and retain new fans). Ultimately, this balance might not please the die-hard fan or captivate the new one. Instead of such a generic broadcast, we believe that TV networks should adopt a more personalized, viewer-centric broadcast. Some level of broadcast personalization has started to appear, particularly within sports venues. For example, Carnegie Mellon University's YinzCam [1], shown in action at a 2009 Pittsburgh Penguins hockey game in Figure 1, allows fans inside a hockey arena to select the specific camera angle that they want to watch on their wifi-enabled smart-phones, and supports a variety of live video-camera angles, including those that allow fans to follow their favorite player on the ice, watch the bench, watch the goal-tender, etc.

Imagine if the sports broadcast network knew which team or which player a viewer favored more during a game, and then showed that viewer more camera angles or more broadcast content that highlighted that viewer's favorite team/player more, rather than staying neutral with the generic broadcast. Being offered content that is personalized and that focuses on a fan's desired perspectives, would substantially enhance a fan's viewing experience. This could equally be extended to a group of viewers, where the dominant preferences of the entire group could be taken into account, to personalize the broadcast for that group's viewing pleasure.

Unfortunately, current sports broadcast methods are designed to replicate the identical video feed to a large number of many viewers at home, which makes it difficult to personalize the feed to suit each individual viewer. Broadcast networks rely on the judgement, expertise and instincts of a human-in-the-loop, the broadcast director, who dictates which specific video feed or camera angle is selected to be sent to the viewers, and at what time. Using this practice, it would simply not be cost-effective to create the many different personalized broadcasts that would be required to satisfy the viewers at home, because many more broadcast directors would need to be hired. Suppose that we could replace the human broadcast director with a computerized director—we would then be able to automatically personalize and target each camera feed to satisfy the different viewers with their respective team/player preferences. This would allow networks the ability to satisfy more fans' preferences, effectively leading to more viewers watching the broadcast, translating into higher ratings and increased advertising revenue (the latter being two of the most important metrics of success for broadcast networks).

Our work in this paper focuses on replacing a human sports director by a computerized director providing a proof-of-concept of how a sports broadcast can be personalized to provide fans with an enhanced experience. In particular, we focus on solving the problem of automatic camera-selection, i.e., given the availability of many video-camera angles during a game, which camera angle should the broadcast network show to each viewer? In our discussions with several human sports directors, the consensus appears to be that they decide the specific camera angle to show to viewers by selecting the most ``interesting'' camera angle at that point in time in the game. The term ``interesting'' is clearly subjective. How does the human sports director know what is interesting to a given fan/viewer? Each fan inevitably has his/her own definition or opinion of what he/she might consider interesting to watch during a game. The research question that we sought to answer with our work was: *given certain information (e.g., fans' preferences, players' statistics), can we automate the task of a human sports director in determining the most interesting camera angle to show to a given fan/viewer?*

The contribution of this paper is an approach that automatically predicts which game-centric objects (e.g., players) will show up on camera, in particular, the object's size and its location in the field of view/play. From that data, given a viewer's stated preferences and player/game statistics, we automatically decide which camera-angle to show to that viewer, in real-time, even as the game is underway. To the best of our knowledge, no other related work has used such a unique combination of data to produce the effect of an automated, viewer-centric sports broadcast. Our implementation, as described in this paper, is completely automated, works in low-latency or in real-time, supports a fan's customized viewing experience given the fan's player, camera, and graphical preferences, and demonstrates our ability to handle multiple camera angles. We validate our approach using two data sets, which we had access to, from real games: (1) footage from 4 different camera angles of a 2009 National Hockey League (NHL) game between the Pittsburgh Penguins and the Washington Capitals, and (2) footage from 8 different camera angles of the 2007 Fiesta Bowl between Boise State and Oklahoma.

The remainder of the paper is organized as follows. In Section 2, we present an overview of our automated sports broadcast approach. In Section 3, we discuss our experiments and results. We present related work in Section 4, and the conclusion and future directions in Section 5.

2. Approach

2.1. Data Gathering

We require several key data points for our automated, real-time broadcast director approach. Our assumptions arise from the need for these data points. Assumption (i), i.e., player tracking, allows us to obtain a player's (x, y, z) coordinates and allow us to determine where each object (player, puck, ball) is relative to the world (the football field, the ice rink). Assumption (ii), i.e., calibrated cameras, enables us to know what each camera is focused on or pointing at, relative to the world. Assumption (iii), i.e., viewer's preference, enables us to determine how which players to focus on and which camera angle to select, in order to personalize the broadcast for that viewer. In the context of this paper, we allow the viewer to weight game objects and camera angles on a scale of 1-100 (1 being least in significance, 100 being the most) to signify the relative degree of importance of the object/camera to the viewer. Assumption (iv), i.e., game statistics, provide us with the ability to favor "star players" (i.e., players with superior statistics), who are likely to be the most followed/viewed, in most cases. In the context of this paper, we use player ratings similar to those used in video-games for statistics; an example for hockey is shown in Table 1.

Table 1. Sample Player Statistics

ID	Team	Name	Rating
0	REF	Puck	99
4	WSH	J. Erskine	84
8	WSH	A. Ovechkin	91
17	PIT	P. Sykora	86
21	WSH	B. Laich	73
24	PIT	M. Cooke	84
29	PIT	M.A. Fleury	84
39	WSH	D. Steckel	84
55	WSH	J. Schultz	55
60	WSH	J. Theodore	81
71	PIT	M. Malkin	85
81	PIT	M. Satan	81
87	PIT	S. Crosby	93

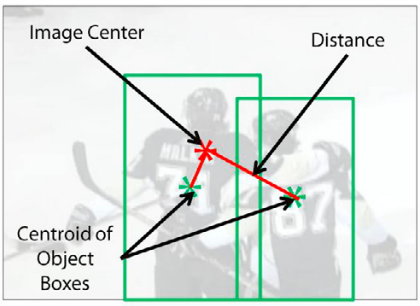

Figure 2. Play Prediction frame for a hockey game.

2.2. Play Prediction

Given the tracked player's coordinates and the camera information, we are able to predict the future positions (or "boxes") of a specific game-object that will show up on camera, as well as where the specific object will be. We show an example in Figure 2. The green boxes highlight the two players that are predicted to show up, and the lines show the distance between the centroid of the box to the image center.

2.3. Automated Director

After obtaining the boxes for each frame, we now need to decide which camera angle is the most "interesting" to show to the specific viewer. Typically this decision is made in real-time, in a viewer-agnostic manner, by a human director. However, as a part of our approach, we also seek to explore and articulate a good metric to replicate a human director's decision, to potentially imitate their actions if they were actually trying to personalize the broadcast for an individual viewer. We calculate each of the camera's score by taking sum of each (ObjectBoxArea/CentroidDistanceFromImageCenter) multiplied by (ObjectUserRating + ObjectStatRating). We then multiply each resulting camera's score by the corresponding CameraUserRating to get the new camera value. Of all the new camera values thus obtained, the maximum camera value is the one that we choose to show to the viewer since it is deemed to be the most interesting (in accordance with our metric) to the viewer.

3. Experimental Validation

To test our approach, we utilized data from real hockey and football environments. Through Fox Sports Network, we were able to obtain three highlight clips during the 2007 Fiesta Bowl game between Boise State University and Oklahoma University (hook-and-ladder play, Boise State scoring an overtime TD, and statue-of-liberty play). In Figure 3, we show the 8 camera angles of one highlight clip where Boise State scores a touchdown on a hook-and-ladder play. Unfortunately, due to the absence of player-tracking infrastructure inside the Fiesta Bowl, we were forced to manually annotate/identify all of the players' location relative to the football field.

Figure 3. Experimental dataset: 8 different camera angles for the 2007 Fiesta Bowl between Boise State University and Oklahoma University.

Figure 4. Experimental dataset: 4 different camera angles for a 2009 NHL game between the Pittsburgh Penguins & the Washington Capitals

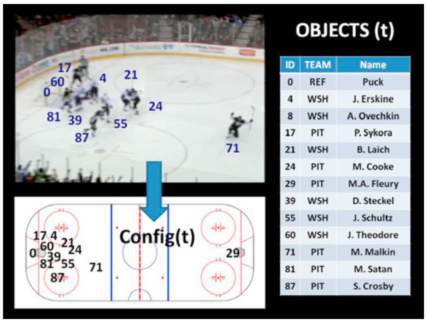

Figure 5. Sample Hockey Configuration

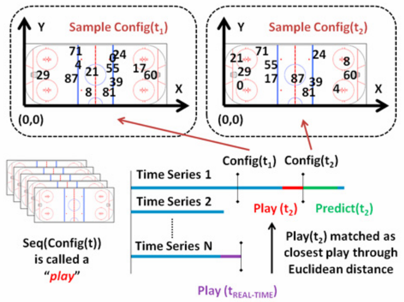

Figure 6. Hockey Play Prediction

Through our ongoing collaboration with the Pittsburgh Penguins for YinzCam, we obtained one 13-second highlight clip where Malkin scores a goal in a 2009 NHL game between the Pittsburgh Penguins and the Washington Capitals. This footage comprises 4 clean video-camera angles of the action taking place. In Figure 4, we show the 4 camera angles (Tight Shot, Wide Shot, Goalie Cam, ISO Cam). Again, because there existed no player-tracking infrastructure inside Mellon Arena (where the game took place), we resorted to the manual annotation of all of the players' locations relative to the hockey rink. A sample hockey configuration is shown in Figure 5.

Given a play (which is a sequence of configurations), we search through the time-series database to discover the nearest matching play. We use Euclidean distance as our play distance function. We highlight the steps in Figure 6. We simplified our experiment by searching in the respective environment (e.g. finding out where 3 seconds of a hockey play was within the 13 seconds of the available hockey game clip, or finding out where 3 seconds of a football play was within the 15 seconds of the available football-game clip). A perfect play match with Euclidean distance 0 would always be found since we used configuration sequences that were a portion of whole highlight clip. This is a current limitation of our experimental validation. After determining the best match, we figure out which camera angle we are using and determine what boxes will show up in next few seconds.

Given the predicted boxes for each camera angle and user preference parameters, we then just plug in the input values for the formula and output the most interesting frame to the specific viewer's device. In the process of developing and experimenting with our first prototype system for an automated sports director, one of the lessons that we learned is the approach is much easier to implement for hockey, as compared to football, because of the large number of objects on the field in the case of football.

4. Related Work

Not much research has been done in the field of generating automated sports broadcasts. Most of the work in this area in the past has been focused on creating tools to assist a human director during a sports broadcast, but not actually replace the human director, as we are attempting to do here. In the context of hockey, several research efforts have previously tried to track hockey players, the hockey puck, and attempted to recognize various hockey actions during a game [9, 2, 5, 8]. In the FoxTrax puck [2], the authors tracked a hockey puck by embedding sensors into it and then overlaying specific graphics on the TV broadcast, depending on the speed of the puck in real time. In [5] and [9], the authors present approaches to recognizing the actions of hockey players within a broadcast sequence. They match a given player template with a known action-template database to recognize the players' action.

However, only in [8] do the authors present an approach for an automatic sports director for hockey. This research is perhaps the closest to our own work because the authors use stationary cameras to track the hockey players over the course of a game and then generate an automatic broadcast from the resulting data gathered. However, our approach differs from this work in that our focus is on generating a real-time automatic broadcast system as opposed to an after-the-fact broadcast, and also in that the our camera-selection decision is viewer-centric as opposed to being play-centric. In addition, in our experimental validation, we have had the opportunity to obtain 4-8 camera angles and 2 real-world data-sets from two different sports (hockey, football), compared to the two camera angles that this related work has been validated with. Lastly, we extend this related work even further by allowing viewers to weight how important each player and camera perspective is to the overall broadcast, and we then leverage this information (the viewer's preferences), along with the player/game statistics to decide which camera to show to that viewer.

In the context of American football, several research efforts have focused on football play-recognition [4] and play formation recognition [3]. In [4], the authors tracked the player X, Y locations on a football field for several plays and tried to predict the football play. This was one of earliest well-known efforts where computers attempted to analyze football game video. In [3], the authors focused more on football-formation recognition rather than football-play recognition. Their football database is a lot to be more comprehensive than the earlier work in [4]. We extended each of these related efforts by additionally taking into account player statistics and viewer preferences to dictate which camera angle to show to the viewer. There has not been any work on automatic director programs of this nature for American football, although perhaps it is likely to be the sport that stands to benefit the most from such an approach.

Other automatic sports broadcast approaches have been conducted in soccer [7]. However, in [7], they depend on higher-level, sport-specific semantic information, such as a shot or a foul recognized from audio and commentary, in order to determine which camera angle to show. Our automatic camera-selection approach is different because we do not use that type of higher-level information to choose a camera angle.

More recently, in the sports broadcast industry, SportVision has worked with Major League Baseball to track players during a game [6]. SportVision is a well-known company that overlays advertisement and game-time graphics (e.g., the yellow "1st down" line) on the field during a broadcast for sports networks. Even though we have tested our approach on hockey and football film, we believe that the same concepts can be readily applied to baseball, soccer, basketball, and many other sports. Effectively, we can envision our automated sports director approach building upon and leveraging SportVision's player-tracking data and calibrated-camera infrastructure.

5. Conclusions

This paper has articulated an approach for an automated sports broadcast director to determine the most interesting camera angle to show to a given viewer broadcast. Our implementation, as described in this paper, was completely automated, worked in real-time, supported a fan's customized viewing experience, given the fan's player, camera, and graphical preferences, and demonstrated the ability to handle multiple camera angles. We have validated our approach using two data sets taken from real games: (1) footage from 4 different camera angles of a 2009 National Hockey League (NHL) game between the Pittsburgh Penguins and the Washington Capitals, and (2) footage from 8 different camera angles of the 2007 Fiesta Bowl between Boise State and Oklahoma.

Effectively, given as input a player's (X, Y, Z) coordinates and camera properties from a given current play, our approach is able to predict the size of player objects and locations, for each object in the next play, for each camera angle. Using the resulting play-prediction information, statistics, and viewer-preference information, our computerized director program decides which camera angle to show to the viewer. Our future work entails how to factor objects being partially obscured by other objects into the formula (e.g. when a player blocks another player), supporting automated replays in the broadcast, and adding commentary. Since our dataset is fairly small (e.g. 4 plays–one hockey, three football), we also need to test our automated director formula on a larger dataset to evaluate the scalability of our approach. Using a part of the training set as our testing set is also not the best way method to test pattern (play) recognition, but we were constrained by what footage that we had access to, as well as the time that it took to manually annotate some of the data-gathering process (reverse camera-calibration and player-tracking). However, by creating a formula and demonstrating some preliminary results off real game footage from multiple cameras in two different sports, we believe that our approach provides some insight and will hopefully generate discussion on the right metrics to determine the most *interesting* camera angle to show to different fans.

Acknowledgements

We gratefully acknowledge the support and collaboration of the Pittsburgh Penguins in the YinzCam project, and Fox Sports Network for providing us with footage for a football game.

References

[1] http://www.yinzcam.com/.
[2] R. Cavallaro. The FoxTrax hockey puck tracking system. IEEE Computer Graphics and Applications, 17(2):6–12, 1997.
[3] R. Hess, A. Fern, and E. N. Mortensen. Mixture-of-parts pictorial structures for objects with variable part sets. ICCV, pages 1–8, 2007.
[4] S. Intille. Visual Recognition of Multi-Agent Action. PhD thesis, Massachusetts Institute of Technology, September 1999.
[5] W.-L. Lu, K. Okuma, and J. J. Little. Tracking and recognizing actions of multiple hockey players using the boosted particle filter. Image Vision Computing, 27(1-2):189–205, 2009.
[6] A. Schwarz. Digital eyes will chart baseball's unseen skills. New York Times, 2009. http://www.nytimes.com/2009/07/10/sports/baseball/10cameras.html.
[7] J. Wang, C. Xu, E. Chng, H. Lu, and Q. Tian. Automatic composition of broadcast sports video. ACM Journal of Multimedia Systems, 14(4):179–193, 2008.
[8] L. Wu. Multi-view hockey tracking with smoothing and camera selection. Master's thesis, The University Of British Columbia, April 2008.
[9] X.Wu. Template-based action recognition:classifying hockey players' movement. Master's thesis, The University Of British Columbia, May 2005.