



US 20040017389A1

(19) **United States**

(12) **Patent Application Publication**

Pan et al.

(10) **Pub. No.: US 2004/0017389 A1**

(43) **Pub. Date: Jan. 29, 2004**

(54) **SUMMARIZATION OF SOCCER VIDEO CONTENT**

Related U.S. Application Data

(76) Inventors: **Hao Pan**, Vancouver, WA (US); **Baoxin Li**, Vancouver, WA (US)

(60) Provisional application No. 60/398,971, filed on Jul. 25, 2002.

Correspondence Address:
Kevin L. Russell
Suite 1600
601 SW Second Ave.
Portland, OR 97204-3157 (US)

Publication Classification

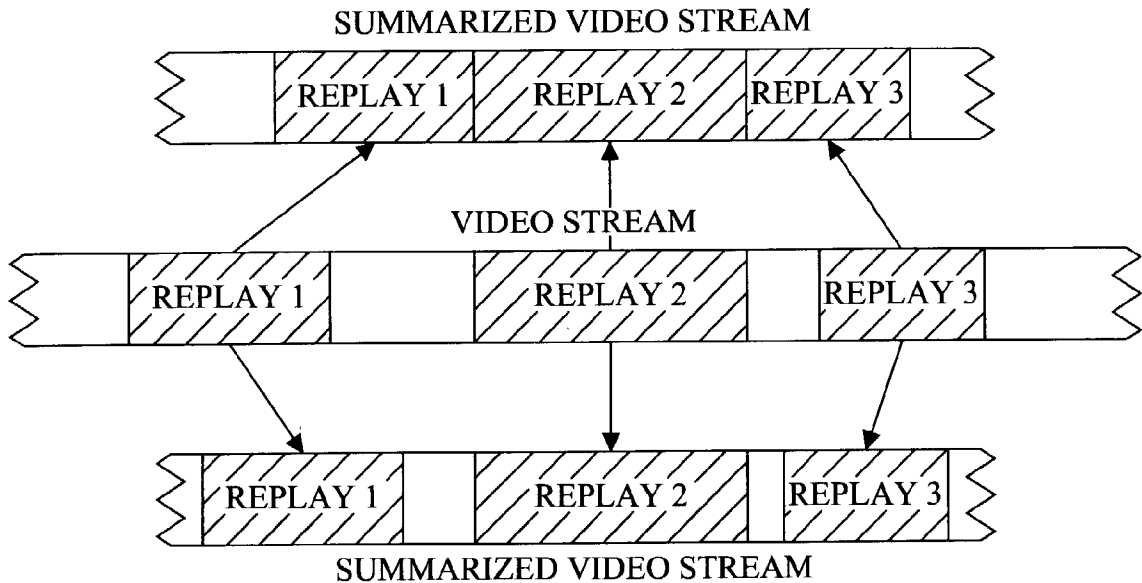
(51) **Int. Cl.⁷ G09G 5/00**
(52) **U.S. Cl. 345/723**

(21) Appl. No.: **10/260,110**

(57) **ABSTRACT**

(22) Filed: **Sep. 27, 2002**

Summarization of video content including soccer.



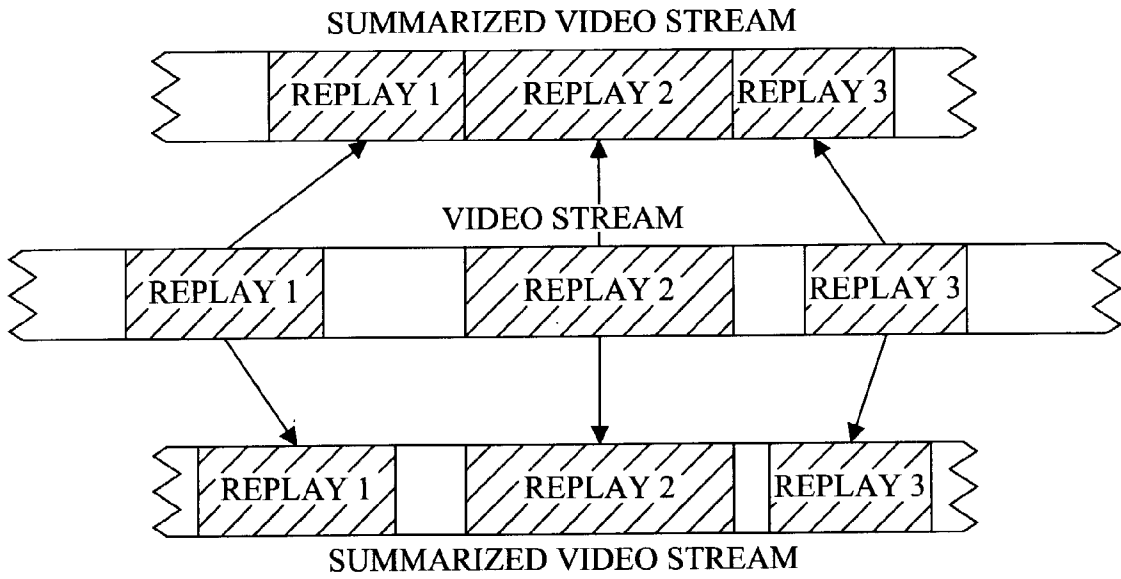


FIG. 1

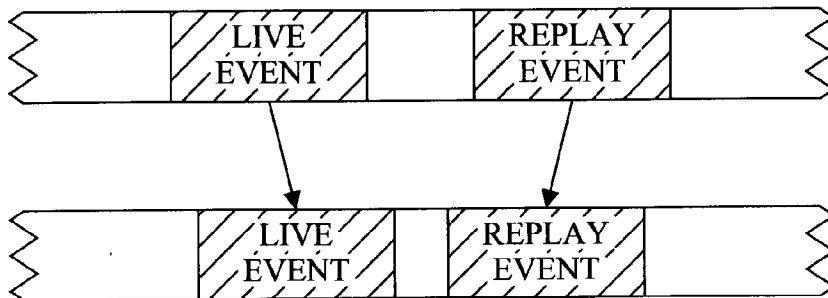


FIG. 2

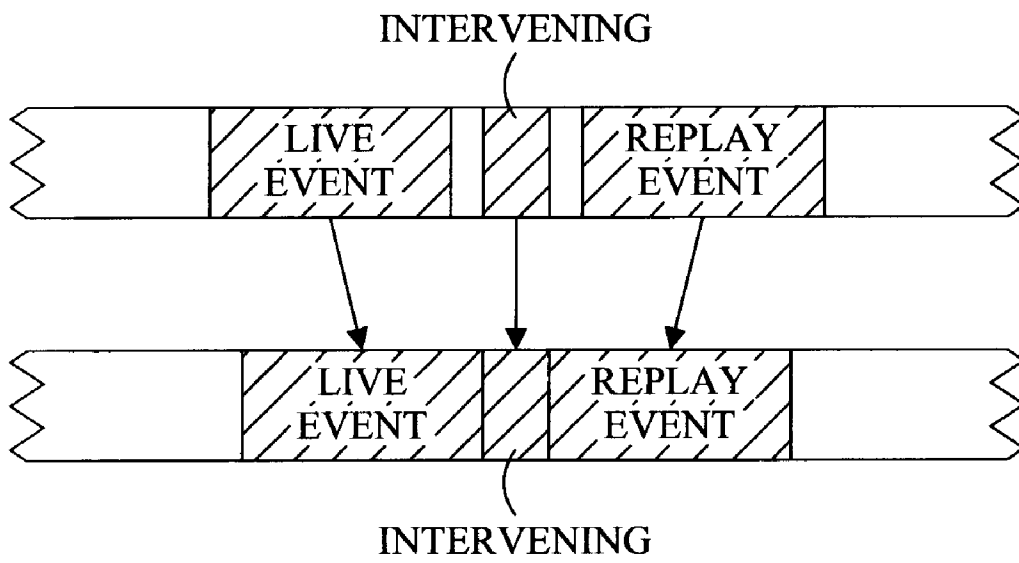


FIG. 3

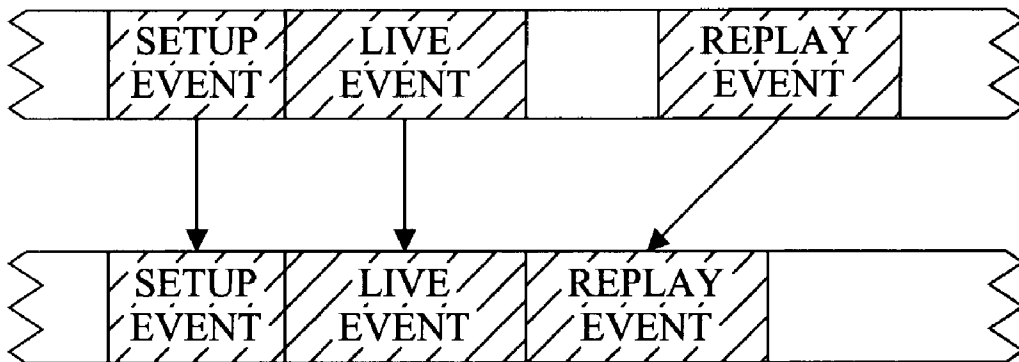


FIG. 4

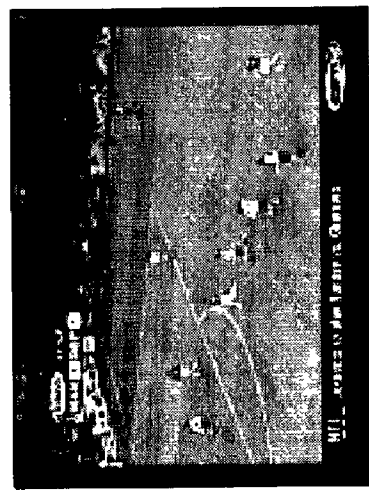


FIG. 5A

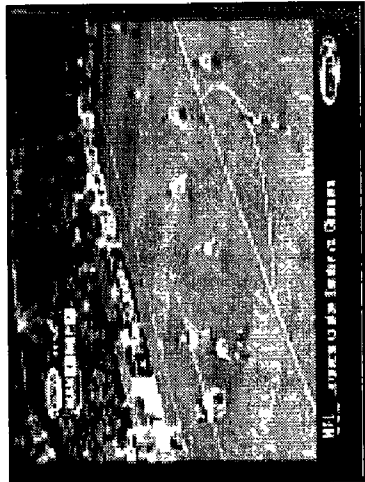


FIG. 5B



FIG. 5C

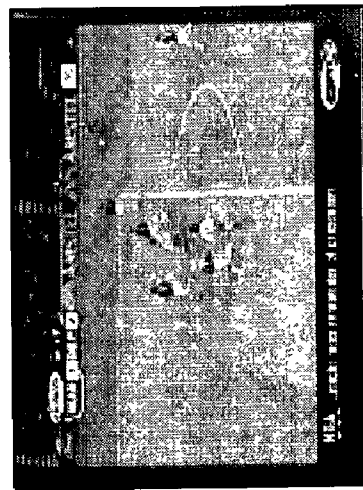


FIG. 5D

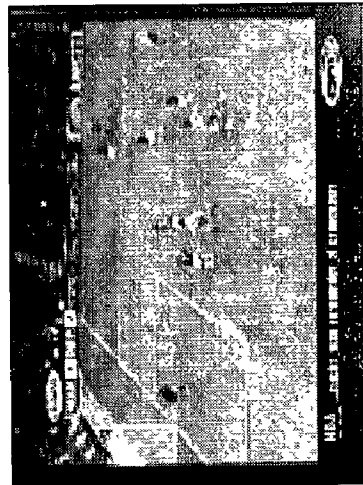


FIG. 5E

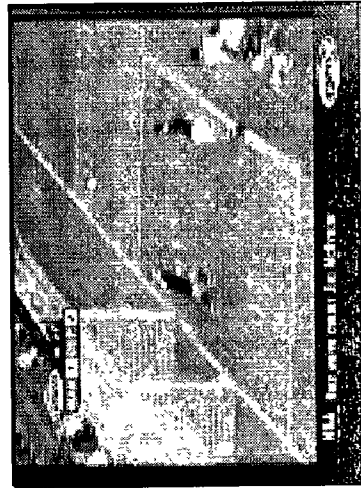


FIG. 5F

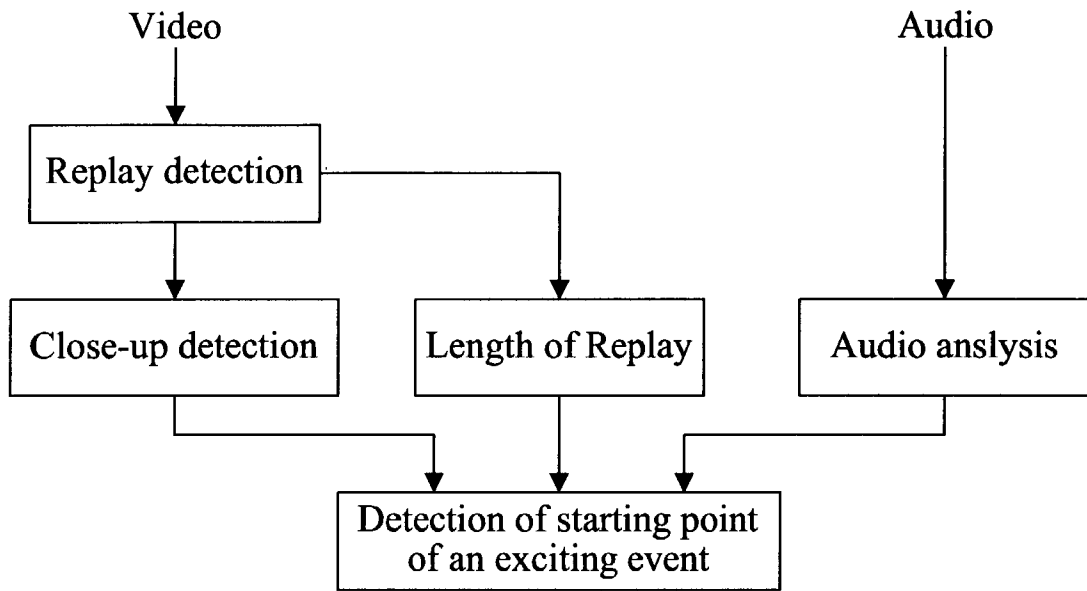


FIG. 6

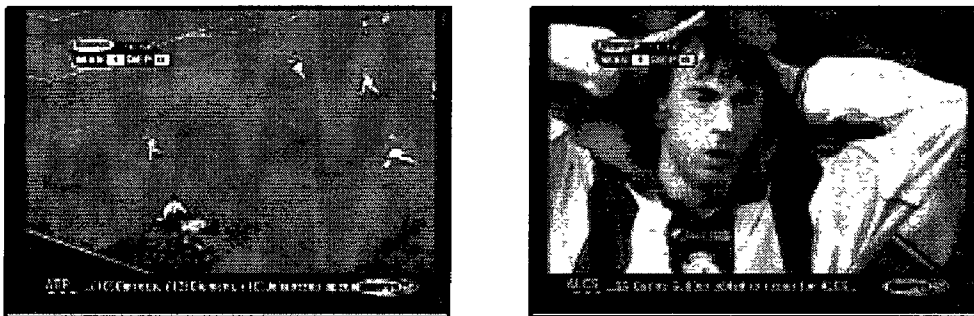


FIG. 7

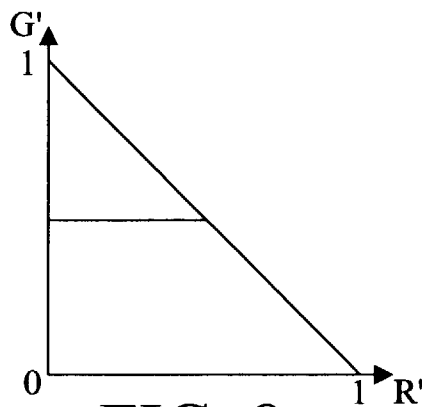


FIG. 8

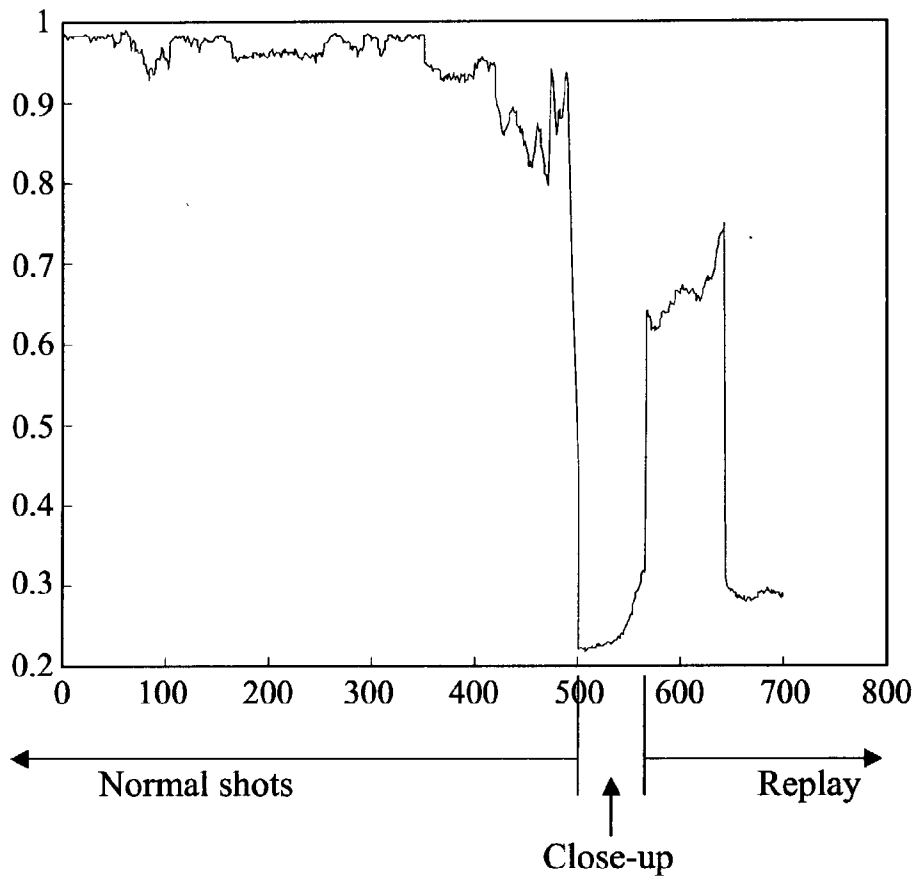


FIG. 9

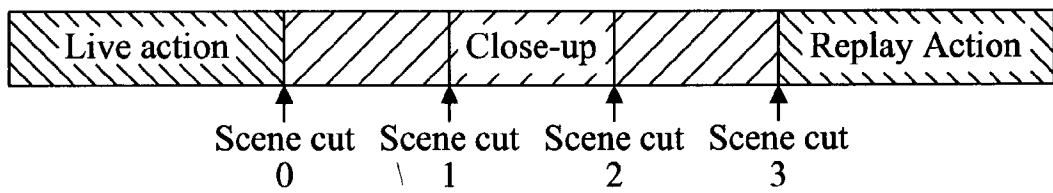


FIG. 10

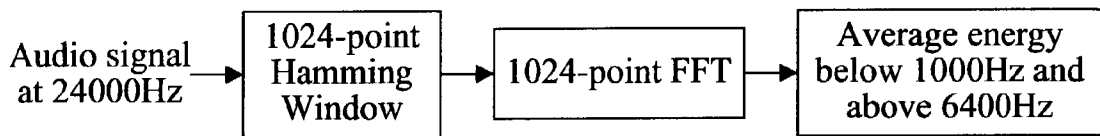


FIG. 11

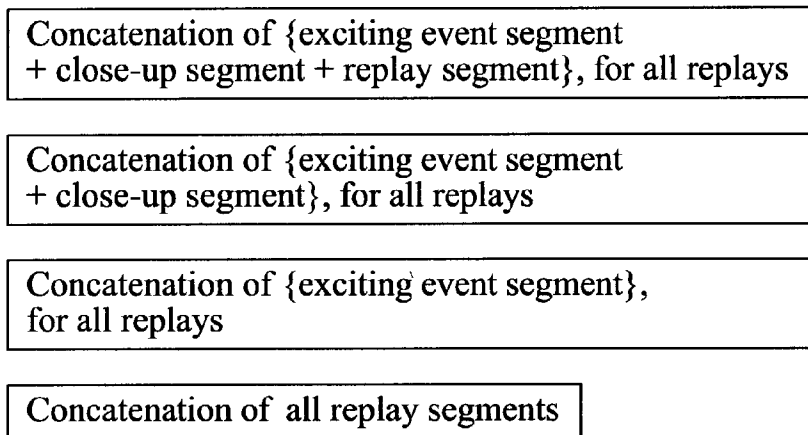


FIG. 12

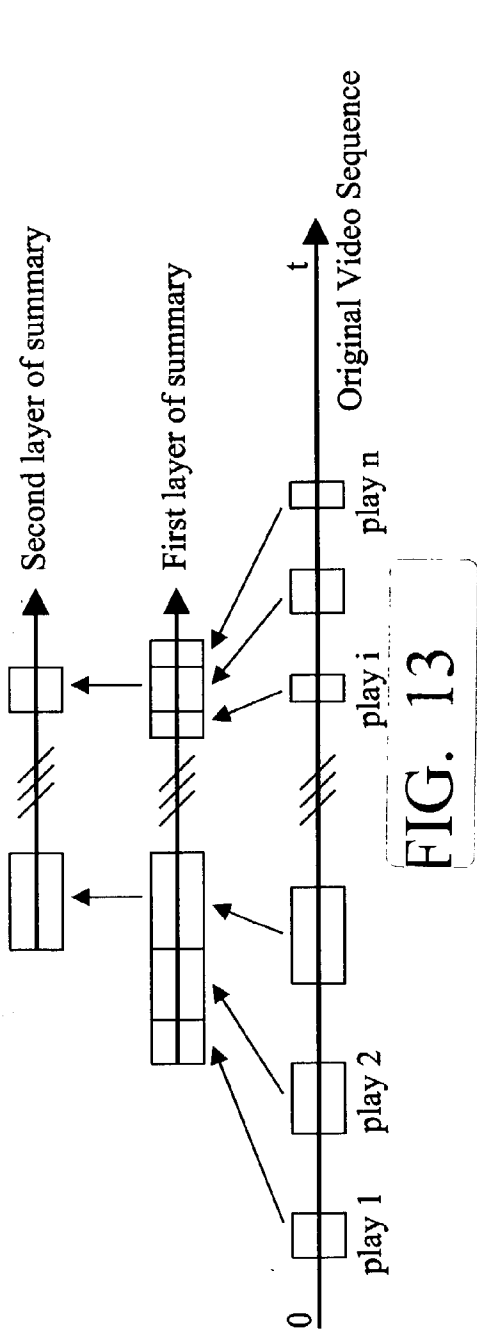


FIG. 13

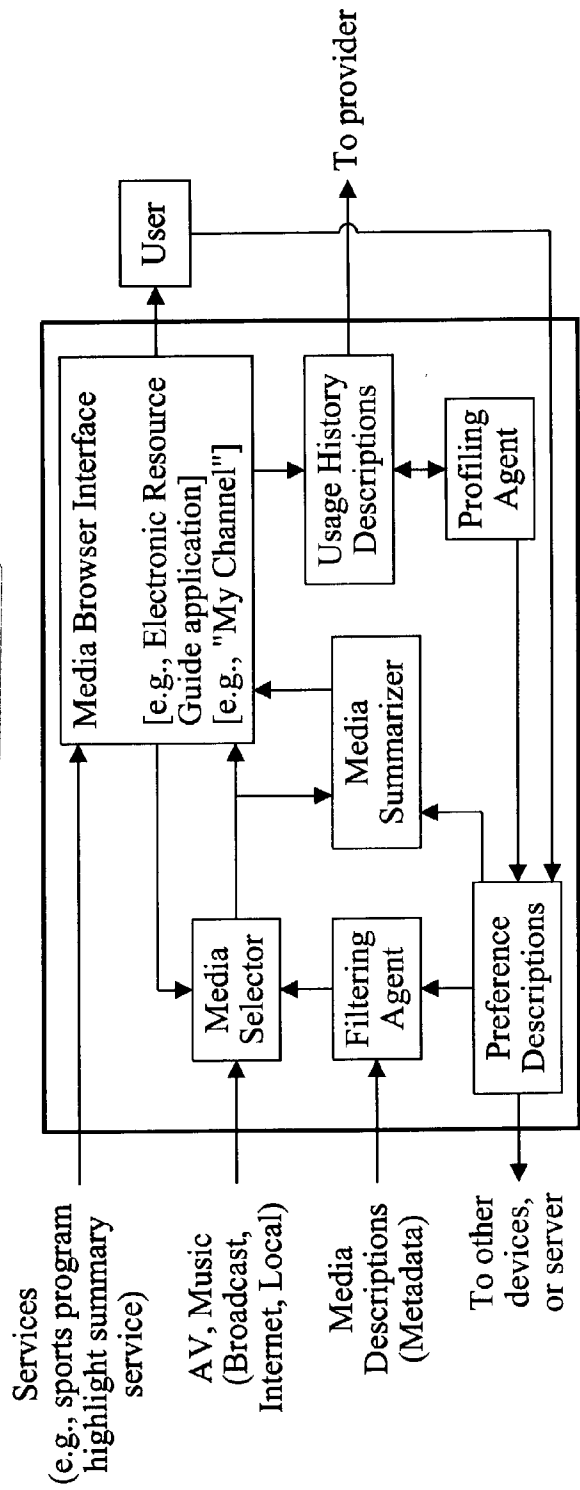


FIG. 14

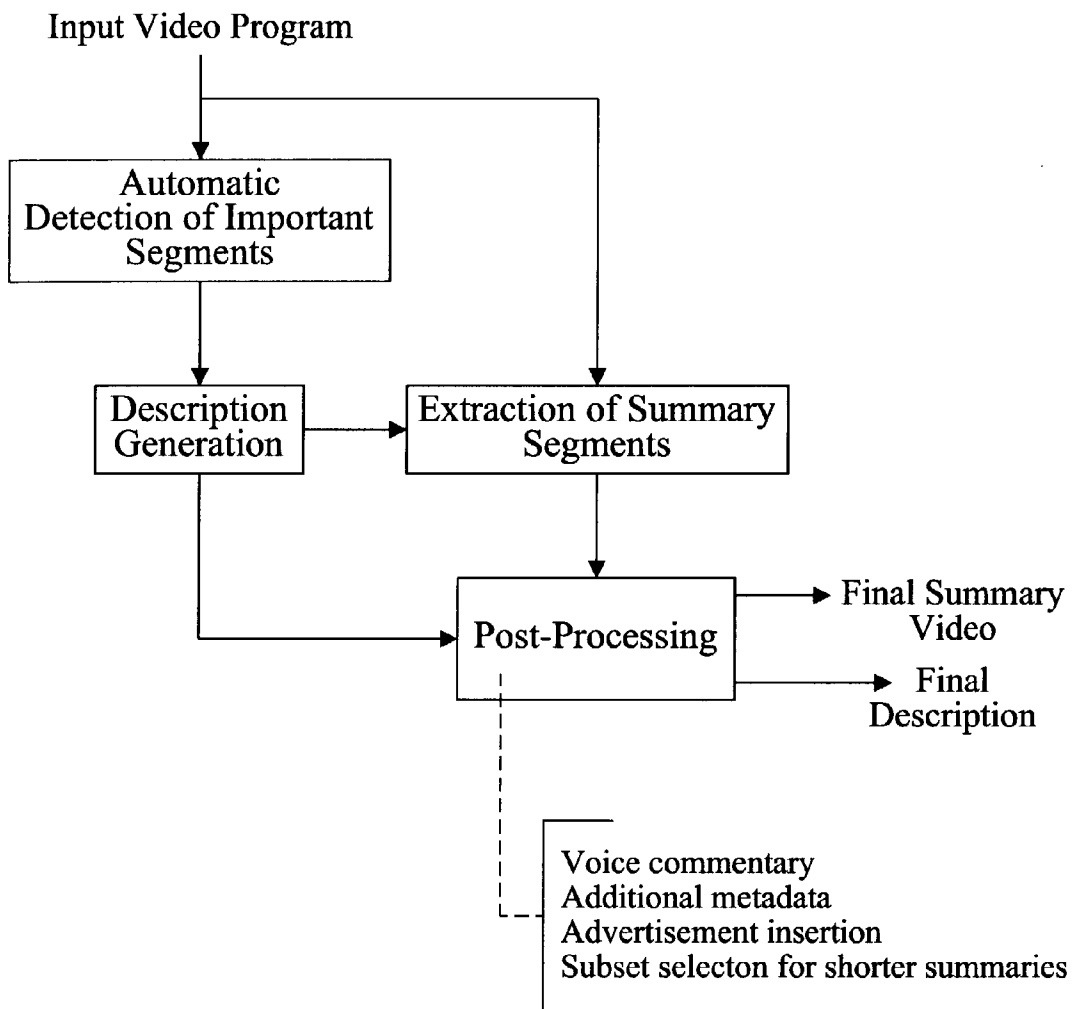


FIG. 15

SUMMARIZATION OF SOCCER VIDEO CONTENT

[0001] This application claims the benefit of U.S. Patent Serial No. 60/398,971 filed Jul. 25, 2002.

BACKGROUND OF THE INVENTION

[0002] The present invention primarily relates to summarization of video content including soccer

[0003] The amount of video content is expanding at an ever increasing rate, some of which includes sporting events. Simultaneously, the available time for viewers to consume or otherwise view all of the desirable video content is decreasing. With the increased amount of video content coupled with the decreasing time available to view the video content, it becomes increasingly problematic for viewers to view all of the potentially desirable content in its entirety. Accordingly, viewers are increasingly selective regarding the video content that they select to view. To accommodate viewer demands, techniques have been developed to provide a summarization of the video representative in some manner of the entire video. Video summarization likewise facilitates additional features including browsing, filtering, indexing, retrieval, etc. The typical purpose for creating a video summarization is to obtain a compact representation of the original video for subsequent viewing.

[0004] There are two major approaches to video summarization. The first approach for video summarization is key frame detection. Key frame detection includes mechanisms that process low level characteristics of the video, such as its color distribution, to determine those particular isolated frames that are most representative of particular portions of the video. For example, a key frame summarization of a video may contain only a few isolated key frames which potentially highlight the most important events in the video. Thus some limited information about the video can be inferred from the selection of key frames. Key frame techniques are especially suitable for indexing video content but are not especially suitable for summarizing sporting content

[0005] The second approach for video summarization is directed at detecting events that are important for the particular video content. Such techniques normally include a definition and model of anticipated events of particular importance for a particular type of content. The video summarization may consist of many video segments, each of which is a continuous portion in the original video, allowing some detailed information from the video to be viewed by the user in a time effective manner. Such techniques are especially suitable for the efficient consumption of the content of a video by browsing only its summary. Such approaches facilitate what is sometimes referred to as "semantic summaries"

[0006] There are several proposed techniques for the summarization of a soccer game based on fully parsing the structure of the whole soccer game. For example, Gong et al. propose to classify the frames in a soccer game into various play categories, such as a shot at left goal, top-left corner kick, play in right penalty area, in midfield, etc., based on a model with four components, the soccer field, the ball, the players, and the motion vectors. See, Y. Gong, L. T. Sin, C. H. Chuan, H.-J. Zhang, M. Sakauchi, "Automatic parsing of TV soccer programs," IEEE conference on Multimedia systems and computing, pp 167-174, 1995.

[0007] Yow et al propose to construct a panoramic view of selected events by recognizing the soccer field, tracking of the ball, and recognizing the camera movement. See, D. Yow, B.-L. Yeo, M. Yeung, B. Liu, "Analysis and presentation of soccer highlights from digital video," Second Asian conference on computer vision, 1995

[0008] Choi et al, propose to construct panoramic view of certain events by recognizing the soccer field by color information, tracking players by template matching and Kalman filtering, and processing occlusion by color histogram back-projection. The panoramic view is based on a model for the soccer field and all frames in the video are transformed into the views in the model. See, S. Choi, Y. Seo, H. Kim, K.-S. Hong, "Where are the ball and players? Soccer game analysis with color-based tracking and image mosaic", International conference on image analysis and Pro, Florence, Italy, pp. 196-203, September, 1997.

[0009] Leonardi et al propose to use both audio and visual features of a multimedia document and a Hidden Markov Model (HMM) as a bottom-up semantic framework and a finite-state machine as a top-down semantic framework. Broadcast soccer is one example of their general method. The video features are "lack of motion", "camera operations" and "presence of shot cuts." The audio features are Mel-Cepstrum coefficients and zero crossing rates and are categorized into music, silence, speech, and music. See, R. Leonardi, P. Migliorati, "Semantic indexing of multimedia documents," IEEE Multimedia, pp.44-51, April-June, 2002

[0010] Xie et al., propose using a Hidden Markov Model (HMM) as the framework to segment a soccer video into "play/break." The features used by Xie et al. are visual dominant-color ratio and motion intensity. See, L. Xie, S.-F. Chang, "Structure analysis of soccer video with hidden Markov models," ICASSP 2002. Xu et al propose a similar framework without using the hidden Markov models. See, P. Xu, S.-F. Chang, "Algorithms and system for high-level structure analysis and event detection in soccer video," ADVENT-Technical report #111, Columbia University, June 2001.

[0011] Xie, proposes using a framework that is entirely based on audio features. The approach segments the whole sound track into commentary and crowd noise and then picks up the excited/unusual parts in crowd noise. L. Xie, "Segmentation and event detection in soccer audio," EE 6820 Project, May 2001.

[0012] Babaguchi et al. propose a technique to link live and replay scenes in American football broadcast video. The replay scenes are detected by (1) a video caption detection or (2) the sandwiching of digital video effects on either side of the replay segment. Babaguchi et al. note that it is impossible to design a general detector applicable to any DVEs, because there are a considerable number of DVE patterns that appear in everyday broadcasts. The DVE effect taught by Babaguchi et al. is a gradual shot change operation related to spatial aspects. The linking of live and replay scenes is performed based upon the dominant color of the key frame and the ratio of the number of vertical lines to that of horizontal lines on the field, which is representative of the camera angle. This technique for linking using the technique taught by Babaguchi et al., namely using the ratio, is suitable for the intended application, namely, American Football. However, this technique is not applicable to soccer.

[0013] What is desired, therefore, is a video summarization technique suitable for video content that includes soccer.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0014] FIG. 1 illustrates a summarized video stream.
- [0015] FIG. 2 illustrates live and replay events.
- [0016] FIG. 3 illustrates live, replay, and intervening segments.
- [0017] FIG. 4 illustrates setup, live, and replay events.
- [0018] FIG. 5 illustrates a live goal
- [0019] FIG. 6 is a block diagram of a video summarization system.
- [0020] FIG. 7 shows far away and close up views.
- [0021] FIG. 8 is a normalized color space.
- [0022] FIG. 9 illustrates pixel ratios.
- [0023] FIG. 10 illustrates scene cuts.
- [0024] FIG. 11 illustrates audio energy.
- [0025] FIG. 12 shows replay based summaries.
- [0026] FIG. 13 shows different summary levels.
- [0027] FIG. 14 shows a browser.
- [0028] FIG. 15 shows video extraction.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0029] Soccer is perhaps the most popular sport in the world. A cumulative total of about 32 billion people watched the 64 games of the 2002 FIFA world cup. A typical soccer game lasts about 90 minutes, plus the half-time break and an occasional extension. Usually, not every moment in the game is eventful or important. After a soccer game is over, most audiences, especially those who have already watched the live game, are more interested in watching the most exciting events again rather than the entire game. Avid fans that missed the game are indeed very interested in all the important events of the game. The time during which the ball is kicked into the goal, a red card occurs, a yellow card occurs, a corner kick occurs, a penalty kick occurs, well played one-on-one ball movement, a shot on goal, a "save" by the goalkeeper, a cross (ball crossed relatively close in front of the goal), a nutmeg (player passes the ball between the legs of his opponent and runs around his opponent to get the ball), bicycle kick, a header, a player is tackled to the ground, etc., are normally the exciting part of the game. The remaining time during the soccer match is typically not exciting to watch on video, such as for example, nearly endless commercials, the time during which the players move the ball around the center of the field, the time during which the players pass the ball back and forth, etc. While it may indeed be entertaining to sit in a stadium for hours watching a 90 minutes soccer match, many people who watch a video of a soccer match find it difficult to watch all of the game, even if they are loyal fans. A video summarization of the soccer video, which provides a summary of the match (e.g., game) having a duration shorter than the original soccer video, may be appealing to many people. The

video summarization should provide nearly the same level of the excitement (e.g. interest) that the original match provided

[0030] Upon initial consideration, soccer would not be a suitable candidate to attempt automated video summarization. Initially, there are nearly an endless number of potential plays that may occur which would need to be accounted for in some manner. Also, there are many different types of plays, such as a corner kicks, throw-ins from the sideline, attempted goals, headers, etc., that likewise would need to be accounted for in some manner. In addition, each of these plays involves significant player motion which is difficult to anticipate, difficult to track, and is not consistent between plays. Moreover, the ball would normally be difficult, if not impossible, to track during a play because much of the time it is obscured from view. In addition, much of the play regarding soccer involves the motion of a significant number of players spread across the field, which is difficult to interpret in an automated fashion. Based upon such considerations soccer has been previously considered impractical, if not impossible, to attempt to summarize.

[0031] It is conceivably possible to develop highly sophisticated models of a typical soccer video to identify potentially relevant portions of the video, and many attempts have been made as previously described in the background. However, such highly sophisticated models are difficult to create and are not normally robust. Further, the likelihood that a majority of the highly relevant portions of the soccer video will be included in such a video summarization is low because of the selectivity of the model. Thus the resulting video summarization of the soccer video may simply be unsatisfactory to the average viewer.

[0032] After consideration of the difficulty of developing highly sophisticated models of a soccer video to analyze the content of the soccer video, as the sole basis upon which to create a soccer summarization, the present inventors determined that this technique is ultimately flawed as the models will likely never be sufficiently robust to detect all the desirable content. Moreover, the number of different types of model sequences of potentially desirable content is difficult to quantify. In contrast to attempting to detect particular model sequences, the present inventors determined that the desirable segments of the soccer match, which are by their nature generally "non-stop" matches, are preferably selected based upon characteristics of the broadcast video. The characteristics of the broadcast video which the present inventors identified as providing an indication of an exciting portion of the soccer match is the "replay". Without the "action-stop" nature of some sporting activities, such as football and baseball, it is difficult to otherwise determine such events. Any method may be used for detecting replays, such as for example, N. Babaguchi, Y. Kawai, Y. Yasugi and T. Kitahashi, "Linking Live and Replay Scenes in Broadcasted Sports Video," *Proceedings of ACM Multimedia 2000 Workshop on Multimedia Information Retrieval (MIR2000)*, pp.205-208, November 2000; Y. Kawai, N. Babaguchi and T. Kitahashi, "Detection of Replay Scenes in Broadcasted Sports Video by Focusing on Digital Video Effects," *Trans. IEICE(D-II)*, Vol.J84-D-II, No.2, pp.432-435, February 2001 (in Japanese); H. Pan, P. van Beek, and M. I. Sezan, "Detection of Slow Motion Replay Segments in Sports Video for Generating Sports Video Highlights," *ICASSP 2001*; and H. Pan, B. Li, and M. I. Sezan, "Automatic

Detection of Replay Segments in Broadcast Sports Programs by Detection of Logos in Scene Transitions," *ICASSP 2002*, incorporated by reference herein.

[0033] Referring to FIG. 1, one technique for the summarization of soccer video is to identify replay segments, such as replay 1, replay 2, and replay 3. Then a modified video stream may be created based upon the detection of the replay 1, replay 2, and replay 3. The modified stream may include the replays without additional material there between, or may include additional video material between the replay segments. In any case, the resulting video stream is of a shorter duration than the original video stream, or otherwise the replay segments are temporally spaced closer together in the modified video stream. Moreover, the modified video stream may be a set of pointers to the original video stream indicating those portions of the original video stream that comprise the summarization, while avoiding the need to actually modify the video stream itself.

[0034] After further consideration of the typical broadcast of soccer matches it was determined that a replay of a potentially exciting action occurs after the presentation of the "live" action. Since in the broadcast of a soccer match the "live" action and the replay of the action are presented with different camera angle and/or field of view, the presentation of both to the viewer would provide added excitement to the event, as shown in FIG. 2. The result is more than the presentation of the same segment twice or two uncorrelated events. In this manner it was determined that the presentation of both the replay event and the live event, both having a different perspective of the same event, would provide additional desirable summarization content for the viewer. The live event is normally temporally previous to the replay of the same event. Preferably, the live event and the replay of the event include at least 25%, 50%, or 75% or more of the same event, albeit potentially using different camera angles and different fields of view (e.g., zoomed in or zoomed out).

[0035] While the presentation of the "live event" and the "replay event" results in content that is far more than merely the presentation of the individual events alone because of the resulting synergy for the user that the combination creates, the present inventors determined that normally a limited amount of time elapses between the "live" and "replay" events, where this region of the video frequently includes "close-up" scenes of players, referees, etc. Such information is also likely to be of interest to the user because it frequently includes facial expressions of excitement or disgust. Referring to FIG. 3, in this manner it was determined that the presentation of the replay event, the live event, and the intervening material, would provide additional desirable summarization content for the viewer. Preferably, the intervening video material includes at least 25%, 50%, or 75% or more of the temporal video sequence between the "live" and "replay"

[0036] While the presentation of the "live event" and the "replay event" results in content that is far more than merely the presentation of the individual events alone because of the resulting synergy for the user that the combination creates, the present inventors also determined that normally a limited amount of the "setup" time to the live action is also likely to be of interest to the user because it frequently includes contextual information to the "live event" and "replay

event" to make the resulting summarization more interesting. The "setup" material in addition is normally semantically relevant to the "live event" and/or the "replay event". Referring to FIG. 4, in this manner it was determined that the presentation of the setup, the live event, and the replay event, would provide additional desirable summarization content for the viewer. Preferably, the setup video material includes at least 25%, 50%, or 75% or more of the duration of either the corresponding "live" and "replay" material.

[0037] Referring to FIG. 5 by way of illustration, FIGS. 5A-5C illustrate a goal from a "live" broadcast, while FIGS. 5D-5F illustrate the corresponding "replay" broadcast. FIGS. 5A-5C show different camera angles and different zoom-factors than used in the "replay" of FIGS. 5D-5F. While the semantic content of the two versions are the same, the live version usually employs a wide-angle camera shooting from a distance while the replay version usually employs a narrow-angle camera shooting in close range. In other words, the frame in the replay appears structurally different from those in the live version. As previously discussed, it would be at best difficult to use content analysis technology to determine that the contents are the same.

[0038] Referring to FIG. 6, a system suitable for the creation of summarizations of soccer content is shown. The video is initially processed by the replay detection module to identify portions of the video that are replays. Any suitable replay detection technique may be used, as desired. The replay detection module identifies the replay segment and the associated live segment.

[0039] To identify close-up segments preceding replays that are subsequent to the corresponding live segment, visual features of the video may be used. The close-up segments are varied in their content, zoom factor, color characteristics, all of which are without any particular invariant characteristics. Therefore it would appear difficult to determine the content as being a close-up. However, the present inventors determined that the close-ups are different from other video, so a determination that it is sufficiently dissimilar is sufficient. Accordingly, the first visual feature relies on detecting the absence of the anticipated color of a soccer field, which is typically generally green. One technique to detect the absence of a particular color is to determine the ratio of the number of pixels belonging to the generally green color of a soccer field to the total number of pixels in a frame. In close-up segments, players and referees are shot from a close distance, as illustrated in FIG. 7B. Therefore, the frame is not dominated by the colors of the soccer field. In contrast, the dominant color of a soccer field which comprises a large portion of a frame is generally green, as illustrated in FIG. 7A.

[0040] In some cases the dominant color of a soccer field varies from time to time, and game to game. However, soccer fields still tend to be generally green in most cases. One technique to determine the ratio of the number of green pixels to the number of pixels in a frame may be calculated in the normalized red-green-blue color space, with normalized red and green color components being considered. The normalized values may be calculated as:

$$R' = \frac{R}{R+G+B} \quad G' = \frac{G}{R+G+B}$$

[0041] The normalized color space is illustrated in **FIG. 8**. One characterization for whether a pixel is generally green may be if its normalized values $R' < 0.5$, $G' > 0.5$, and $R'+G' \leq 1.0$. The generally green pixels may be counted and compared again the total number of pixels in the entire frame to determine if the scene is sufficiently green. Other characterizations and/or color spaces may likewise be used. Referring to **FIG. 9**, an example of the ratio of the number of green pixels to the total number of pixels in a frame over a period of time is illustrated. For typical soccer the ratio is close to 1, while in close-up shots the ratio is close to 0; while in replay segments the ratio tends to vary.

[0042] The second visual feature relies on detecting sufficient scene changes during the video. Usually there is a sudden scene cut between the live action and the close-up segment. Periodically, a close-up segment between the live action and its replay consists of several different shots of players, coaches, and audiences, and therefore includes several scene cuts, as illustrated in **FIG. 10**. The close-up segment shown in **FIG. 10** includes three segments and four scene cuts. The close-up segment may be characterized by the low ratio of the number of green pixels to the number of the total pixels in a frame. The earliest scene cut in this time period with a low ratio may be selected as the starting point of the close-up segment and the end point of the live action, labeled as scene cut **0** of **FIG. 10**.

[0043] Any suitable scene cut detection technique may be used. In the preferred system the scene cuts are determined as the difference between color histograms of every two adjacent frames. The color histogram is calculated in the normalized red-green-blue color space and each of the normalized red and green color components are evenly divided into 16 bins. As a result, there are a total number of $16 \times 16 = 256$ bins. The difference between two color histograms is measured using a mean-square error criterion. A sufficient change, such as greater than a predetermined threshold, in the color histogram difference identifies a scene cut.

[0044] After determining the end points of the live action segments, the system then determines their starting point. Due to the nature of soccer games most of the events do not have objective starting points of exciting events for a user, except for perhaps a corner kick and a free kick. Moreover, in most exciting actions the different events in a soccer game proceed in an unbroken sequence. In many cases, a portion of a game that is continuous may last for several minutes and uses camera pan and zoom, but lacks sharp scene cuts or gradual transitions. Therefore, it is not effective to use scene cut or transition detection techniques to determine the "starting point" of event segments. Accordingly, the system may subjectively select the starting points of events associated with the detected replay segments as being a given duration, such as 15 or 30 seconds prior to the starting point of the close-up segments or end point of the live action segments. It is to be understood that the duration of the "live" segment may be selected in any manner, such as in a manner based in some way on the duration of the "replay" segment.

[0045] A system may have multiple durations of the corresponding live segments. For example, the system may subjectively select the starting point of events associated with the detected replay segments as being either 15 or 30 seconds prior to the starting points of the close-up segments (or end point of live action segments). The criteria for selecting between 15 or 30 seconds may involve defaulting to 15 seconds unless a criteria is satisfied, otherwise 30 seconds is selected.

[0046] The first criteria for selecting 30 seconds may be the energy of the audio signal at frequencies below 1000 hertz and/or above 6400 hertz. Energies of the audio track below 1000 hertz and above 6400 hertz are usually low in unexciting moments and high in exciting moments. The energy may be calculated as shown in **FIG. 11**. It is to be understood that the frequencies may be changed, as desired. If the average audio energy over the 30-second duration is larger than the average energy over the 15-second duration, then a 30-second live event segment is preferred, otherwise a 15-second live event segment is preferred.

[0047] The system may also use the audio pitch contour to reflect the degree of excitement of commentators. For example, a higher pitch may be associated with a higher degree of excitement.

[0048] Another criteria is the duration of the replay segments. The lengths of the replay segments tend to be correlated with the excitement of the actions in the replays. For example, a replay of a goal which is considered to be quite exciting is much longer than a replay of a foul which is considered only mildly exciting. When an action is more exciting, it is more desirable to provide more "live" event to see how the exciting event occurred. One way to make this determination is if the replay exceeds a threshold, then a 30-second long "live" event is selected; otherwise, a 15 second "live" event is selected.

[0049] A summarization presented to the user may be hierarchical in nature, if desired. The desired summary may be selected by the user. Referring to **FIG. 12**, the summary may be the concatenation of the replay segments. The next summary may be the concatenation of the live segments. The next summary may be the concatenation of the live and replay segments. The next summary may be the concatenation of the live, close-up, and replay segments. Also, the techniques taught herein may be applicable to other sporting broadcasts, such as for example, ice hockey.

[0050] The summary description defines those portions of the video sequence that contain the relevant segments for the video summarization. The summary description may be compliant with the MPEG-7 Summary Description Scheme or TV-Anytime Segmentation Description Scheme. A compliant media browser may apply the summary description to the input video to provide summarized viewing of the input video without modifying it. Alternatively, the summary description may be used to edit the input video and create a separate video sequence. The summarized video sequence may comprise the selected segments which excludes at least a portion of the original video other than the plurality of segments. Preferably, the summarized video sequence excludes all portions of the original video other than the plurality of segments.

[0051] Referring to **FIG. 13**, the first layer of the summary is constructed using the detection technique. The second and

third layers (and other) are extracted as being of increasingly greater excitement, based at least in part, on the audio levels of the respective audio of the video segments. Also, it would be noted that the preferred audio technique only uses the temporal domain, which results in a computationally efficient technique. In addition, the level of the audio may be used as a basis for the modification of the duration of a particular play segment. For example, if a particular play segment has a high audio level then the boundaries of the play segment may be extended. This permits a greater emphasis to be placed on those segments more likely to be exciting. For example, if a particular play segment has a low audio level then the boundaries of the play segment may be contracted. This permits a reduced emphasis to be placed on those segments less likely to be exciting. It is to be understood that the layered summarization may be based upon other factors, as desired.

[0052] Referring to FIG. 14, the video summarization may be included as part of an MPEG-7 based browser/filter, where summarization is included within the standard. With different levels of summarization built on top of the aforementioned video summarization technique, the system can provide the user with varying levels of summaries according to their demands. Once the summary information is described as an MPEG-7 compliant XML document, one can utilize all the offerings of MPEG-7, such as personalization, where different levels of summaries can be offered to the user on the basis of user's preferences described in an MPEG-7 compliant way. Descriptions of user preferences in MPEG-7 include preference elements pertaining to different summary modes and detail levels.

[0053] In the case that the summarization is performed at a server or service provider, the user downloads and receives the summary description encoded in MPEG-7 format. Alternatively, in an interactive video on demand (VOD) application, the media and its summary description reside at the provider's VOD server and the user (e.g., remote) consumes the summary via a user-side browser interface. In this case, the summary may be enriched further by additional information that may be added by the service provider. Further, summarization may also be performed by the client.

[0054] Referring to FIG. 15, the output of the module that automatically detects important segments may be a set of indices of segments containing events (e.g., plays) and important parts of the input video program. A description document, such as an MPEG-7 or TV-Anytime compliant description is generated in The Description Generation module. Summary segments are made available to the Post-Processing module by The Extraction of Summary Segments module which processes the input video program according to the description. A post-processing module processes the summary Segments and/or the description to generate the final summary video and final description. The post-processing module puts the post-processed segments together to form the final summary video. The post-processing module may transcode the resulting video to a format different than that of the input video to meet the requirements of the storage/transmission channel. The final description may also be encoded, e.g., binarized if it is generated originally in textual format such as XML. Post-processing may include adding to the original audio track a commentary, insertion of advertisement segments, or metadata. In contrast to play detection, post-processing may be completely, or in part,

manual processing. It may include, for example, automatic ranking and subset selection of events on the basis of automatic detection of features in the audio track associated with video segments. This processing may be performed at the server and then the resulting video transferred to the client, normally over a network. Alternatively, the resulting video is included in a VOD library and made available to users on a VOD server.

[0055] The terms and expressions which have been employed in the foregoing specification are used therein as terms of description and not of limitation, and there is no intention, in the use of such terms and expressions, of excluding equivalents of the features shown and described or portions thereof, it being recognized that the scope of the invention is defined and limited only by the claims which follow. All references cited herein are hereby incorporated by reference.

1. A method of processing a video including soccer comprising:

(a) identifying a plurality of segments of said video based upon it being a replay of another portion of said video, where each of said segments includes a plurality of frames of said video; and

(b) creating a summarization of said video by including said plurality of segments, where said summarization includes fewer frames than said video

2. The method of claim 1 wherein one of said segments is a goal shot.

3. The method of claim 1 wherein one of said segments is a corner kick.

4. The method of claim 1 wherein said summarization is a set of pointers to said video.

5. A method of processing a video including soccer comprising:

(a) identifying a plurality of first segments of said video based upon it being a replay of another portion of said video, where each of first said segments includes a plurality of frames of said video;

(b) identifying a plurality of second segments of said video, wherein each of said second segments corresponds with a corresponding one of said first segments containing the same event; and

(c) creating a summarization of said video by including said plurality of first segments and second segments, where said summarization includes fewer frames than said video.

6. The method of claim 5 wherein said summarization includes a set of segments in the following order, (1) a first one of said second segments corresponding with said first one of said first segments, (2) a first one of said first segments, (3) a second one of said second segments corresponding with said second one of said first segments; (4) a second one of said first segments, (5) a third one of said second segments corresponding with said third one of said first segments, (6) a third one of said first segments.

7. The method of claim 6 wherein additional portions of said video exist between at least one of said first and second segments of said set

8. The method of claim 6 wherein said a first one of said first segments and said first one of said second segments are at least 25% of the same event

9. The method of claim 6 wherein said a first one of said first segments and said first one of said second segments are at least 50% of the same event.

10. The method of claim 6 wherein said a first one of said first segments and said first one of said second segments are at least 75% of the same event.

11. The method of claim 5 wherein one of said another portion is the same as one of said second segment.

12. A method of processing a video including soccer comprising:

- (a) identifying a first segment of said video based upon it being a replay of another portion of said video, where said first segment includes a plurality of frames of said video;
- (b) identifying a second segment of said video, wherein said second segment corresponds with said first segment containing the same event; and
- (c) identifying a third segment of said video temporally between said first segment and said second segment;
- (d) creating a summarization of said video by including said first segment, said second segment, and said third segment, where said summarization includes fewer frames than said video.

13. The method of claim 12 wherein said summarization includes a plurality of said first segments, a plurality of corresponding said second segments, a plurality of corresponding said third segments, and a set of segments in the following order, (1) a first one of said second segments, (2) a first one of said third segments, (3) a first one of first segments corresponding with said first one of said second segments, (4) a second one of said second segments, (5) a second one of said third segments, (6) a second one of said first segments corresponding with said second one of said second segments, (7) a third one of said second segments, (8) a third one of said third segments, (9) a third one of said first segments corresponding with said third one of said second segments.

14. The method of claim 13 wherein additional portions of said video exist between a corresponding two at least one of said first and second segments of said set.

15. The method of claim 13 wherein said a first one of said first segments and said first one of said second segments are at least 25% of the same event.

16. The method of claim 13 wherein said a first one of said first segments and said first one of said second segments are at least 50% of the same event.

17. The method of claim 13 wherein said a first one of said first segments and said first one of said second segments are at least 75% of the same event.

18. A method of processing a video including soccer comprising:

- (a) identifying a first segment of said video based upon it being a replay of another portion of said video, where said first segment includes a plurality of frames of said video;
- (b) identifying a second segment of said video, wherein said second segments corresponds with said first segment containing the same event;
- (c) identifying a third segment of said video temporally prior to said first segment and semantically significant to at least one of said first segment and said second segment; and

(d) creating a summarization of said video by including said first segment, said third segment, and said second segment, where said summarization includes fewer frames than said video.

19. The method of claim 18 wherein said summarization includes a plurality of said first segments, a plurality of corresponding said second segments, a plurality of corresponding said third segments, and a set of segments in the following order, (1) a first one of said third segments, (2) a first one of said second segments corresponding with said first one of said first segments, (3) a first one of said first segments, (4) a second one of said third segments, (5) a second one of said second segments corresponding with said second one of said first segments; (6) a second one of said first segments, (7) a third one of said third segments, (8) a third one of said second segments corresponding with said third one of said first segments, (9) a third one of said first segments.

20. The method of claim 19 wherein additional portions of said video exist between a corresponding two of at least one of said first and second segments of said set, and between a corresponding two of at least one of said third and first segments.

21. The method of claim 19 wherein a third segment is at least 25% of the duration of at least one of the corresponding first and second segments.

22. The method of claim 19 wherein a third segment is at least 55% of the duration of at least one of the corresponding first and second segments.

23. The method of claim 19 wherein a third segment is at least 75% of the duration of at least one of the corresponding first and second segments.

24. A method of processing a video including soccer comprising:

- (a) identifying a first segment of said video based upon it being a replay of another portion of said video, where said first segment includes a plurality of frames of said video;
- (b) identifying a second segment of said video semantically including close-up images temporally prior to said first segment and; and
- (c) creating a summarization of said video by including said first segment, and said second segment, where said summarization includes fewer frames than said video.

25. The method of claim 24 further comprising identifying a third segment of said video, wherein said third segment corresponds with said first segment containing the same event.

26. The method of claim 25 wherein said summarization includes a plurality of said first segments, a plurality of corresponding said third segments, a plurality of corresponding said second segments, and a set of segments in the following order, (1) a first one of said third segments, (2) a first one of said second segments, (3) a first one of said first segments corresponding with said first one of said third segments, (4) a second one of said third segments, (5) a second one of said second segments, (6) a second one of said first segments corresponding with said second one of said third segments; (7) a third one of said third segments, (8) a third one of said second segments, (9) a third one of said first segments corresponding with said third one of said third segments.

27. The method of claim 24 wherein said second segment is identified based upon not detecting sufficient generally green within the frames of said second segment.

28. The method of claim 24 wherein said second segment is identified based upon detecting scene changes within said video.

29. The method of claim 28 wherein generally green is defined by a value of $R' < 0.5$, $G' > 0.5$, and $R' + G' \leq 1.0$ where $R' = R / (R + G + B)$ and $G' = G / (R + G + B)$.

30. A method of processing a video comprising:

- (a) identifying a first segment of said video based upon it being a replay of another portion of said video, where said first segment includes a plurality of frames of said video;
- (b) identifying a second segment of said video temporally previous to said first segment based upon sufficient dissimilarity to said first segment; and

(c) creating a summarization of said video by including said first segment, and said second segment, where said summarization includes fewer frames than said video.

31. The method of claim 30 wherein said second segment semantically includes close-up images.

32. The method of claim 30 wherein said video is soccer.

33. The method of claim 30 wherein said dissimilarity is based upon color.

34. The method of claim 33 wherein said color is generally green.

35. The method of claim 34 wherein said dissimilarity is based upon determining ratios of pixels.

36. The method of claim 30 wherein said dissimilarity is based upon a scene cut.

* * * * *