

19



Octrooi Centrum
Nederland

11

2012575

12 C OCTROOI

21

Aanvraagnummer: **2012575**

51

Int.Cl.:

H04N 21/44 (2011.01)

G06K 9/00 (2006.01)

H04N 21/845 (2011.01)

H04N 21/8549

(2011.01)

22

Aanvraag ingediend: **07.04.2014**

30

Voorrang:

05.04.2013 EP 13162585

73

Octrooihouder(s):

**Nederlandse Organisatie voor
Toegepast-Natuurwetenschappelijk
Onderzoek TNO te Delft.**

43

Aanvraag gepubliceerd:

-

72

Uitvinder(s):

**Anton Havekes te Delft.
Johannes Gerardus Maria Schavemaker
te Delft.**

47

Octrooi verleend:

07.10.2014

45

Octrooischrift uitgegeven:

15.10.2014

74

Gemachtigde:

ir. C.M. Jansen c.s. te Den Haag.

54

Video processing system and method.

57

The video processing system provides for automatic selection of video sequences showing goals in a video stream from cameras directed at a playing field. Classification is applied to image content from different foreground positions in an image. The set of classes comprising a first and second class corresponding players of different teams. Medians or averages of the image positions of the local image content assigned to the first and second class are determined. A condition is tested whether the results are on mutually opposite sides of a reference position and a distance between collected positions or distances between the reference position and the first and second collected positions exceed a threshold value. When the test shows that the condition is met, a preceding video sequence is selected, for example for display.

NL C 2012575

Dit octrooi is verleend ongeacht het bijgevoegde resultaat van het onderzoek naar de stand van de techniek en schriftelijke opinie. Het octrooischrift komt overeen met de oorspronkelijk ingediende stukken.

P100585NL00

Title: Video processing system and method

Field of the invention

5

The invention relates to a video processing system, a video play system and a video retrieval system comprising such a video system, and a video processing method and video play and retrieval methods.

10 Background

Automated extraction of significant clips from soccer video is known from an article by Assfalg et al, titled "Semantic annotation of soccer videos: automatic highlights identification" and published in Computer Vision and
15 Image Understanding Vol. 92, pages 285–305.

Assfalg et al consider automatic detection of highlights based on the estimation of visual cues indications of state transitions in a finite state machine model of soccer game progress. One proposed visual cue is the localization of players. Assfalg et al note that the distribution of the players' positions in specific regions of the playing field to detect placed kicks such
20 as kick-offs. Since different types of placed kicks exhibit different players' deployments, the distribution of the players' positions in specific regions provides useful information in order to recognize the type of placed kick.

Assfalg et al suggest to detect a cue for a kick-off by forming clusters of
25 players' blobs based on their colors, and checking whether the clusters' bounding boxes are almost completely separated by the midfield line.

However the use of blob clustering, which has to be performed for many images if no cues are to be missed, adds to the computational burden and time needed to detect events. This could prevent a real time response to
30 soccer video.

Summary

Among others, it is an object to provide an improved method to select video segments.

5 An automated method of selecting a trigger image from series of images according to claim 1 is provided. The series of images may contain all successive images from a camera or a temporally sub-sampled selection from these images. In this method machine classification of local image content from different foreground positions in the image to different classes
10 is used. The foreground positions may have been selected by image segmentation, or by pattern recognition, stereo image processing etc. In an embodiment, the local image content may be aggregate image data from respective foreground image regions that have been determined by image segmentation, each foreground image region corresponding to a different
15 one of the positions. The aggregate image data for a foreground image region may be an average of pixel values of pixels in the image region for example, or a histogram of such pixel values, the pixel value of some pixel in the foreground region etc. Herein a pixel value may be a color vector, or a single value. In another embodiment the local image content may be an
20 individual pixel value.

The local image content is used to assign the different positions to selected ones of the classes. A first and second collected position are determined for the first and second class, using a median or average of the image positions of the local image content assigned to the first and second class respectively.

25 The collected positions are used to detect whether a condition is met that the first and second collected positions are on mutually opposite sides of a reference position and a distance between collected positions or distances between the reference position and the first and second collected positions exceed a threshold value. When a two dimensional collected position is used,
30 each distance may optionally be determined from a component, along a

predetermined direction, of a vector to the collected position from the reference position or from the other collected position. When the condition is met for an image, this image is used as a trigger image. In a further embodiment detection whether said condition is met for at least a
5 predetermined fraction of the images for image time points in a time interval may be used as the condition. In this embodiment the detection may be associated with an image for a time point in the time interval, e.g. the image for the first or last time point in the time interval. As an expansion of the condition it may be tested whether an amount of motion in
10 the image has exceeded a predetermined value prior to the time point of the detected image, or that the amount of motion has decreased by more than a threshold prior to the detected image.

In an embodiment amounts of motion in first and second parts of the images for time points preceding the time point of the trigger image are determined.
15 Herein the first and second parts show different parts of a playing field, e.g. regions containing the goals. A detection is performed whether the amount of motion in the first part exceeds the amount of motion in the second part or vice versa. This detection can be used to decide which side has scored and to maintain a score count.

20 In an embodiment a kick off after a change of sides may be detected, by testing whether the relative aggregate positions of the first and second classes have been reversed, that is, whether if the aggregate position of the first class is larger than that of the second class in the current trigger image, the reverse was true in a preceding trigger image. Such a detection
25 may be used to assist maintenance of the score count by preventing that the score count is changed for such trigger images.

The trigger image may be used for example to trigger play of a video clip that contains images from a time interval prior to the trigger image, or to trigger transmission of a message, which may optionally contain image data
30 from such a time interval. In an embodiment, the video clip may be played

from a video storage system in response to detection of the trigger image and rendered on a display screen. The video clip may be played as an interruption of another image stream, such as a received broadcast stream or a stream of prestored images. In a further embodiment further pre-stored
5 image data may be played in combination with the video clip, a selection between different pres-stored image data being made dependent on image content at a time of the trigger image, or in a time interval adjacent to, or including that time point. In another embodiment the trigger image may be used to trigger retention of the video clip in a video storage system.

10 In an embodiment image streams are captured from a plurality of cameras directed at a same playing area. In this embodiment images from the image stream from a first camera of said plurality are used to select the trigger image. The first camera may have a field of view that includes the middle line of the playing field midway between opposite goal areas, and parts of
15 the field on both sides of that line. Subsequently a video clip is extracted that contains a part of the image stream from a second camera captured in a time interval preceding a time point at which the trigger image was captured. The second camera may have a field of view that includes one of the goal areas of the playing field.

20 In an embodiment the video clip being extracted from a combined image stream generated by switching between different ones of the cameras as source of image stream based on a content of the image streams prior to the time point of the trigger image.

Detection of the trigger image may be performed using components that are
25 also used to select the source cameras of the combined image stream.

A computer program product for performing the method is provided. The computer program product, which may be a computer readable medium such as an optical or magnetic disk or a semi-conductor memory, comprises a program of instructions for a programmable computer system that, when

executed by the programmable computer system, will make the programmable computer system execute the method.

5 According to another aspect a video processing system is provided with a classifier for classifying local image content, a position collector, configured to determine a median or average of the image positions of the local image content assigned to a first and second class respectively, and a detector to detect a trigger image if a condition is met that the first and second collected positions are on mutually opposite sides of a reference position and a
10 distance between collected positions or distances between the reference position and the first and second collected positions exceed a threshold value.

The video capture system may comprise a selection module configured to select respective ones of the cameras as sources of images for successive
15 image time points, based on content of the image streams, the video storage system being configured to play the video clip using the images from the cameras selected for the image time points in the video clip, or to make the video clip available for display on demand e.g. using twitter, or transmit the video clip. Making the video clip available for display on demand may
20 comprise transmitting signals to devices that are capable of rendering the video clip when a trigger image has been detected. The camera selection will be referred to as focus selection, i.e. selection of a camera that will be used for a special purpose, distinguished from other cameras.

25 Brief description of the drawing

These and other objects and advantageous aspects will become apparent from a description of exemplary embodiments, with reference to the following figures.

30 Figure 1 shows a video system

Figure 2 shows a flow chart of trigger image selection

Figure 3 shows a video system

Detailed description of exemplary embodiments

5

Figure 1 shows a video system, comprising a plurality of cameras 10a-c, a video storage system 12 and a video processing system 14. By way of example a display system with a video display 16 is shown. Video storage system 12 has inputs coupled to cameras 10a-c. Video processing system 14 is coupled to video storage system 12, for receive images from first camera 10a. Although an embodiment is shown wherein video processing system 14 receive images from first camera 10a via video storage system 12, it should be appreciated that instead a direct connection first camera 10a to video processing system 14 may be used, and/or video storage system 12 may receive images from video processing system 14 instead of directly from cameras 10a-c. In the illustrated embodiment, video storage system 12 has an output coupled to video display 16. Video storage system 12 has an optional input 120 for receiving image data from an external source.

In an example of operational deployment, cameras 10a-c are directed at different sub-areas of a soccer playing field. The field of view of first camera 10a contains the middle line of the playing field, including parts of the playing field on mutually opposite sides of that middle line in the field of view. A second and third camera 10b,c may have fields of view that include the penalty area and the goals on opposite sides of the playing field respectively.

A schematic organization of video processing system 14 is shown, comprising an image segmentation module 140, a segment parameter extractor 142, a classifier 144, a position collector 146 and a detector 148 coupled in series. As shown by way of illustration, image segmentation module 140 and segment parameter extractor 142 receive image data

separately from video storage system, but video processing system 14 may furthermore comprise an image memory (not shown), which image segmentation module 140 and segment parameter extractor 142 may use to access data from an image. Alternatively image segmentation module 140 and segment parameter extractor 142 may have their own image memories for this purpose. Detector 148 is coupled to video storage system 12.

Video processing system 14 may comprise a programmable computer, or a combination of computers with a program or programs to make it perform operations of the various components of shown in the figure, as described in the following. A computer program product such as a magnetic or optical disk, or a semi-conductor memory may be provided which stores these programs. As used herein, for the embodiment with the programmable computer or computers, statements that video processing system 14, or a component of video processing system 14 is configured to perform operations, or that video processing system 14 performs operations, mean that the program contains instructions to make the programmable computer do so. Alternatively, video processing system 14 may be configured to perform the operation under control of dedicated circuits.

In operation streams of video data from cameras 10a-c are stored in video storage system 12 and optionally played to video display 16, i.e. retrieved from video storage system 12 and output to video display 16 in a form to make video display 16 render the images. In addition to video data from cameras 10a-c, video storage system 12 may play other data to video display 16, such as locally stored commercials and/or received data such as a TV program stream obtained using external input 120.

Figure 3 shows an alternative embodiment, wherein cameras 10a-c are coupled directly to video processing system 14 and video processing system 14 supplies image data to video storage system. In this embodiment image segmentation module 140 and segment parameter extractor 142 process images from all cameras 10a-c and the output of classifier 144 is coupled to

a focus selection module 36, which used classification results to select for each image time point in a combined video stream from which of cameras 10a-c an image will be used. In this embodiment video processing system 14 comprises image memories for different ones of cameras 10a-c (only one
5 image memory 30 labeled) and multiplexers 32, 34. Although separate image memories 30, it should be appreciated that instead one memory with storage areas for images from different cameras may be used, and that the function of multiplexers 32, 34 may be performed by selection between these storage areas. A first multiplexer 32 serves to supply images from
10 different cameras 10a-c to image segmentation module 140 and segment parameter extractor 142. A second multiplexer 34 serves to provide image storage system 12 with images from the cameras 10a-c selected by focus selection module 36, for storing data representing the selected images.

Although image segmentation module 140 and segment parameter extractor
15 142 may process all images from all cameras 10a-c, this is not essential: alternatively they may process temporally sub-sampled selections of the images. This reduces processing load and it may also reduce the amount of image memory that needs to be used.

Video processing system 14 at least partly controls selection of images that
20 will be displayed by video display 16. Video processing system 14 processes images from first camera 10a, in order to make a dynamically changing selection of one of the cameras 10a-c from which the images will be fed to video display 16. In order to do so, video processing system 14 processes images from first camera 10a, in order to detect a predetermined type of
25 event. Upon detection video processing system 14 triggers video storage system 12 to play selected video data to video display 16.

Figure 2 shows a flow chart of image processing by video processing system 14. In a first step 21, video processing system 14 receives data representing an image obtained by first camera 10a. The steps of the flow-chart may be
30 executed for all images captures by first camera 10a. However, in an

embodiment the steps may be executed for a sub-samples set of these images, e.g. for one image per second. In a second step 22, image segmentation module 140 segments the image. Segmenting, also called segmentation, for which segmentation algorithms are known per se, 5 comprises selecting regions of locations in the image on the basis of image content.

In a simple form of segmentation, a region may be selected by identifying pixels in the image that have a property value within a property value range defined for the region, such as a pixel value in a pixel value range defined 10 for the region, and determining a region of pixel locations of such pixels, by selecting pixel locations that are each a neighbor of at least one other pixel location in the region. Segmentation may be used to identify foreground regions and a background region surrounding the foreground regions. The background region may correspond to a part of the image shows a playing 15 field for example, and a number of foreground regions may correspond to parts of the image that show players.

Other forms of segmentation that may be used include recognition based segmentation. An exemplary form of recognition comprises computing histograms of image value gradients for each of a plurality of image blocks 20 in the image, computing values of a distance measure between the computed histogram and respective reference histograms, and identifying the block as a foreground region or not dependent on the computed differences, for example using a support vector machine, which is known per se. Herein the reference histograms may be derived from training examples.

25 Another example of recognition is shape matching, which may involve detecting edges in the image, or determining gradients, and comparing the detected edges or gradient of each region of the image with reference contours and identifying the region as a foreground region dependent on computed match scores with the reference contours.

Other forms of segmentation that may be used are based on measurements to determine a deviation from background depth. Stereo image processing may be used for example. This may be realized for example by measuring disparity between the image and one or more further images obtained
5 simultaneously by one or more further cameras (not shown) directed at a same part of the field as first camera 10a, from a different camera position. In this form of segmentation a map of pixel locations may be used, that define, for each of a plurality of pixel positions in the image from first camera 10a, pixel positions in each of the one or more further images, where
10 a same position on the playing field is visible. In this form of segmentation pixel positions in foreground regions may be determined by detecting whether the pixel value at those positions from first camera 10a differ by more than a threshold amount from the pixel values of the mapped pixel position in the one or more images from the further cameras. A direct
15 comparison of pixel values may be used, but alternatively a pixel value mapping may be defined, to compute a pixel value to be compared for a camera from a pixel value of another camera.

In a third step 23, segment parameter extractor 142 extracts parameter values of foreground regions that have been selected in second step. In one
20 example, an average pixel value of the pixel values of pixels in the region may be computed.

In a fourth step 24, classifier 144 classifies the foreground regions based on the extracted parameter values. Classification assigns at least part of the foreground regions to a first or second class based on the extracted
25 parameter values. In one example, the foreground regions may be assigned to the classes by testing whether the size of the difference between an extracted parameter value of the foreground region and reference parameter values for the first and second class does not exceed a threshold. Optionally a vector containing a plurality of parameter values may be used. The
30 classification need not be exclusive: foreground regions with a parameter

value or vector that differs by more than a threshold from the reference parameter values for both classes may be left unassigned to either class. These foreground regions may correspond to the referee or a goal keeper for example. Alternatively, more than two classes may be used for
5 classification, only the first and second class being used for subsequent detection as described in the following.

In an embodiment, predefined classes may be used. In another embodiment, the classes may be defined based on the image itself, for example by clustering. Clustering, which is known per se, may make use of a clustering
10 quality measure for a combination of first and second reference parameter values or vectors, defined as a sum of distance measures of the extracted parameter value or vector of the foreground region to the closest of the first and second reference parameter value or vector, summed over the detected foreground regions. Clustering may be used to select a first and second (sets
15 of) reference parameter values that minimizes such a sum. These first and second (sets of) reference parameter values may be used for assignment to the classes. In another embodiment, clustering to define the classes may be performed independent of detection in fourth step 24, e.g. periodically.

In an embodiment classifier 144 may suppress unstable classifications. For
20 example, classifier 144 may track image positions of detected regions through successive images, and suppress a classification of a region in the current image if its image position is not located in a range of positions predicted from tracking of regions with the same classification in preceding images. In a simpler embodiment, classifier 144 may suppress a
25 classification of a region if less than a predetermined number of successive images including the current image contains regions of the same classification within a predetermined range of positions relative to the position of the region in the current image.

In a fifth step 25, position collector 146 determines a first and second joint
30 position for the foreground regions that have been assigned to the first and

second class respectively and, in the embodiment with suppression, not suppressed. The collected position may be a median or an average of the image position of the foreground regions in the same class. In this way, the effect of false region classifications at outlying image positions is reduced. A
5 one dimensional collected position may suffice, for example a position component along a horizontal direction in the image from first camera 10a, or a position of a projection onto a line that runs between the goals. But a two dimensional collected position may be determined.

The image position of a foreground region may be determined as an average
10 or median of the pixel locations in the region for example. But instead the position of an arbitrary pixel in the foreground region may be used as the position of the region, or averages of maximum and minimum coordinate values in the region. This does not make a significant difference when the regions are small.

15 In a sixth step 26, detector 148 determines whether the first and second collected positions satisfy a predetermined condition that reflects spatial separation of the classes by the middle line of the playing field. This condition depends on the first and collected positions relative to a reference position that corresponds to the position of the middle line of the playing
20 field.

In an embodiment the predetermined condition is a first condition that the first and second collected positions should be on different sides of the reference position, and that both the difference between the first collected position and the reference position, and the difference between the second
25 collected position and the reference position exceed a predetermined threshold. In another embodiment, the predetermined condition is a second condition that the first and second collected positions should be on different sides of the reference position and that the difference between the first and second collected position exceeds a predetermined threshold.

In an embodiment, the predetermined condition is a third condition that the first condition (or alternatively the second condition) is met at least a predetermined fraction of the images for image time points in a time interval of predetermined duration prior to a time point of the trigger image.

5 This makes detection more robust. Equivalently, the time interval used in the third condition may follow the time point of the trigger image, or contains images for time points before and after that of the trigger image, The predetermined time interval may have a duration of between 1 second and 20 seconds for example, and the predetermined fraction may lie
10 between 60% and 100% for example. Detector 148 may store a result of a test whether the first condition (or the second condition) is met into a memory (not shown), for use to evaluate the third condition. Once the third condition is met, detector 148 may disable the stored result of the test whether the first condition (or the second condition) was met for the
15 predetermined time interval.

In an embodiment, the predetermined condition is a fourth condition that is a combination of the first condition (or alternatively the second condition, or the third condition) and a fifth condition. In this embodiment, the fifth
20 condition is that a detected amount of movement of the detected foreground regions. The amount of movement for an image time point may be detected from motion vectors representing movement of image data from a previous image, or by taking lengths of velocity vectors obtained tracking the foreground regions for example, or from an absolute value of a component of such a velocity vector or an image to image motion vectors, and summing
25 the lengths or absolute values obtained for the different foreground regions. In a further embodiment, the detected amount of movement may be a sum of amounts of movement for image time points during a further time interval of a predetermined length ending at the time point at which the first or second condition was detected to be met, or at the beginning of the
30 time interval used in the third condition, if the third condition is used.

In an embodiment, the predetermined condition is a sixth condition that is a combination of the fifth condition and a seventh condition. In this embodiment, the seventh condition is that a temporal decrease in the amount of movement of the detected foreground regions between a first time point for which the fifth condition was met and a second time point a predetermined time interval (e.g. one second) prior to the first time exceeds a predetermined threshold.

If the predetermined condition used in sixth step 26 is not satisfied, the process repeats from first step 21 for a subsequent image. If the predetermined condition is satisfied, detector 148 executes a seventh step 27, using the fact that the predetermined condition was detected to make the image a trigger image. In an embodiment, detector 148 sends a command to video storage system 12 to play video data that represents images from second and/or third camera 10b,c captured during a video clip time interval that precedes the time point of reception of the trigger image. In an embodiment, video storage system 12 may be configured to use a time interval of predetermined duration at a predetermined temporal position relative to the detected time point of sixth step 26. Alternatively, video processing system 14 may be configured to select the video clip time interval. Subsequently, the process repeats from first step 21 for a subsequent image.

In an embodiment, detector 148 may be configured to suppress or qualify detection of a trigger image having indicia associated with the start of a game or game half sixth step 26. Detector 148 may also be configured to detect which side has scored and add a detection result as a qualification. As used herein a “qualification” is information stored or transmitted in association with a trigger signal identifying a trigger image.

Detection of the start of a game or game half may be performed based on the average or median positions of the foreground regions of the first and second class respectively. Trigger images may be assigned to a first or second type

dependent on whether the average or median positions of the foreground regions of the first class is larger than the average or median positions of the foreground regions of the second class. A change between the first and second assigned type may be used to detect the start of a game half. In an embodiment, detector 148 is configured to determine the type for each detected trigger image and to suppress or qualify use of the image as a trigger image in response to a change of the assigned type.

Detection which side has scored can be performed based on the side of the playing field where most motion was detected in a predetermined time interval prior to the trigger image. In an embodiment, detector 148 is configured to determine the first and second amounts of motion in respective parts of the images, corresponding to respective parts of the playing field or respective goal areas, and to select an indication of the scoring side associated with each trigger image dependent on which of the first and second amounts of motion was largest in a predetermined time interval prior to the trigger image. This may be used to update a score value and/or to select information such images to be displayed in conjunction with images preceding the trigger images.

In another embodiment, detector 148 may be configured to detect which side has scored by testing a sign of the direction of movement of the average or median positions of the foreground regions of the first and second class in the predetermined time interval prior to the trigger image.

In another embodiment, detector 148 may be configured to detect which side has scored by testing whether a sum of the average or median positions of the foreground regions of the first and second class exceeds a threshold corresponding to the position of the centre line of the playing field in the predetermined time interval prior to the trigger image.

Video processing system 14 may be configured to select one or more of the 10a-c as a source of the images for the video clip. In an embodiment video processing system 14 may be configured to make a continuous focus

selection, assigning focus to images from different ones of the cameras 10a-c in different time intervals (focus is used here in terms of selecting between images rather than in the optical sense). This means that for any time point one of the cameras will be assigned to have the focus. In this embodiment
5 video processing system 14 may include, in the command to video storage system 12, an indication of the camera or cameras 10a-c that have the focus during the video clip time interval.

As a result of the command, video storage system 12 causes video display 16 to render video data for the selected video clip time interval as a video clip.
10 In other embodiments, video storage system 12 may be configured to make a video clip available for display on demand e.g. using twitter, or to transmit the video clip, the video clip containing image captured in a time interval preceding the time of capture of the detected trigger image. Making the video clip available for display on demand may comprise transmitting
15 signals to devices that are capable of rendering the video clip when the trigger image has been detected, so that the users of the devices are enabled to select to play the video clip.

In an embodiment, video storage system 12 is configured to play a background stream of video images, and to interrupt play of this backup
20 stream to play the video clip in response to the command. The background stream may comprised pre-stored images from video storage system 12, such as commercial messages, or a stream derived from an external source, such as a TV channel, images obtained from an Internet website, or images with text received from an external source.

25 In the embodiment shown in figure 3, video storage system 12 may be configured to play a stream of images from the cameras 10a-c continuously, and to interrupt play of this stream to play the video clip, with earlier images from one of those cameras 10b,c, in response to the command. Video storage system 12 may store a combined stream that, for each time point,
30 contains only image data representing images from the camera 10a-c that

had the focus at that time point. In this case, it is not needed to include, in the command to play the video clip, the indication of the camera or cameras 10a-c that had the focus.

Alternatively video processing system 14 may be configured to send further
5 commands to video storage system 12 to indicate focus selection during continuous play. In this case, video storage system 12 may store images from the cameras according to the commands, or store image data from all cameras 10a-c as well as data from the commands, for use to select images during play.

10 In the embodiment of figure 3, video processing system 14 may detect foreground regions from the images from all cameras 10a-c. In this case video processing system 14 may assign focus to one cameras 10a-c dependent on detected foreground regions from the different cameras. In this case, detection of the trigger image may make use of results that are
15 produced as part of a continuous focus selection process. Optionally, the foreground regions may be classified, and optionally suppressed using verification by tracking as described, as part of the focus selection process. In this embodiment only foreground regions assigned to the first and second class will be used in the focus selection process. In one example, focus may
20 be switched when video processing system 14 detects that motion has caused a count of foreground regions in an image to rise by a predetermined number.

In the embodiment wherein video processing system 14 determines focus selection during continuous play, the focus selection process provides for
25 segmentation and optionally also classification and suppression, these intermediate steps of the focus selection process may be used as the corresponding steps in the process described using figure 2. Thus, only the overhead of determining the first and second collected positions and testing the predetermined condition needs to be added for selecting the video clip.

However, it should be noted that other methods of focus selection may be used. Alternative methods to assign the focus are known per se.

In an embodiment, video processing system 14 may select the video clip time interval based on focus selection. For example video processing system 14
5 may select the end of this interval as a time of a last previous focus change to the first camera 10a from one of the second and third cameras 10b,c. A predetermined temporal length up to that end may be used in the video clip time interval for example, or a time interval from the last preceding focus change.

10 In an embodiment, video storage system 12 is made to play one or more further image sequences in response to detection of the trigger image, before and or after the video clip with images from cameras 10a-c. The one or more further video clips may comprise commercial messages for example. Video storage system 12 may be configured to respond to the command to play the
15 video clip by retrieving the one or more further image sequences and playing them to image display 16. In another embodiment video processing system 14 includes an indication of the one or more further image sequences in the command to video storage system 12.

In a further embodiment, the one or more further image sequences are
20 selected dependent on image processing results. This may be used to implement a selection by a sponsor to show his commercial messages only if a team of his selection has scored, or to signal which team has scored before playing the video clip.

In one embodiment, video storage system 12 stores one or more further
25 image sequences in association with the first and second class used in the classification of the regions, e.g. in a table of such associations or stored with the one or more further image sequences. Detector 148 may be configured to select one of the first and second class with the trigger image and to indicate the selected class to video storage system 12, video storage
30 system 12 being configured to play the one or more further image sequences

associated with the indicated class, if any, in response to the command to play the video clip. Alternatively, video processing system 14 may have a memory storing data representing such an association and detector 148 may be configured to select one or more of the further image sequences, if any, based on the selected class and its association with further image sequences, and to send a command to video storage system 12 to play the selected one or more of the further image sequences, if any.

Detector 148 may be configured to select the one of the first and second class for example by testing whether the collected position of the class is on the same side of the reference position as a side associated with one of the last of the second and third camera 10b,c that last had the focus before the detection of the trigger image.

In an embodiment, the definition of the classes is determined automatically, for example by clustering as described. Automated class definition makes it superfluous to use human intervention to define the first and second class for individual games. Automated class definition may be performed on an ongoing basis, not limited to trigger image detection. In an embodiment, respective class definitions determined by clustering respective different trigger images are compared. The comparison may be used to identify classes from trigger different images (by looking for the class with closest reference parameter value or vector) and/or to detect game changes (when at least one of the reference parameter value or vector differs by more than a threshold from both of the previous reference parameter values or vectors). Although a display system has been shown by way of example, it should be appreciated that the video processing system as described could be used in other applications. For example, in a video storage application, the video processing system may be configured to control video storage system 12 to retain selectively video data for the selected video clip time interval, other video data being discarded. Discarding may take place for example when the age of the other video data is older than a predetermined buffer age. The

stored data may be used as part of data that is accessible via a predetermined website for example.

In a video retrieval application, the video processing system may be configured to store identifications of selected video clip time intervals in addition to stored video data. In this case video storage system 12 may be configured to use these identifications to retrieve and play only those selected video clip time intervals when operating in a selected play mode. Other play modes may include a normal play mode wherein video data outside the selected video clip time intervals is retrieved as well, e.g. from cameras that were assigned to have the focus.

In a messaging application, the video processing system may be configured to generate a message in response to detection of the trigger image and cause a transmitter to transmit that message. A stored list of one or more target addresses may be used to provide destination addresses of the message. In a further embodiment, the video processing system may cause the selected video clip and optional one or more further image sequences to be included in the message as well.

Although the described processes may be applied to all images received from first camera 10a, it should be appreciated that it may suffice to use only part of the received images, for example a temporally sub-sampled sub-set of images. Furthermore, once a first trigger image has been detected, detection of trigger images within a predetermined time distance from the first trigger image may be suppressed, or at least transmission of additional command to play the video clip based on trigger images within this time distance may be suppressed.

Although an embodiment has been described wherein the images are segmented, resulting foreground regions are classified and positions are determined for the segments, it should be appreciated that alternatively individual pixels may be classified, and the collected class positions for the first and second class may be determined as a median or average of the pixel

positions that have been assigned to the first and second class respectively. This reduces computational demands at the cost of some robustness. Individual pixels may be assigned to a class based on a difference between the pixel value of the pixel and a reference pixel value for the class, or based
5 on a difference between reference data and a data extracted from pixel values in a surrounding of the pixel location the pixel. Alternatively, foreground pixels may be selected on a pixel by pixel basis, and the selected foreground pixels may subsequently be assigned to the classes. Deviation from background depth may be used to select foreground pixels for example,
10 by means of stereo image processing.

The omission segmentation may make tracking more difficult, and thereby also suppression of unstable class assignments. However, for static image configurations, such as at kick off, tracking is not needed. Hence, suppression of an unstable class assignment to a pixel location may be
15 realized by detecting whether the pixel location, or pixel locations in a window surrounding the pixel location, has or have the same class assignment in at least a predetermined threshold number of images in a time window that contains the current image.

Although exemplary embodiments have been described in terms
20 of video images of a soccer match it will be appreciated that the a number of the described technical measures apply as well to the detection of separation events of other groups of objects as triggers for subsequent action. In the examples, the objects are soccer players, or more specifically areas in video images produced due to the players. But the technical measures may also be
25 applied to video registrations of other sports that involve separation of players, like hockey, and even to separation events in other sports that do not have goal lines. Outside sports separation events of any kind of objects, e.g. during sorting of different types pellets, migration of bacteria etc, may be detected. However, the technical measures are especially effective for the
30 properties of images that result from soccer games.

Conclusies

1. Een geautomatiseerde werkwijze voor het selecteren van een activerend beeld uit een reeks van beelden, welke werkwijze het opnemen van beelden omvat en voor elk beeld in de reeks
 - classificeren van lokale beeldinhoud uit verschillende voorgrondposities in het beeld elk in één van een verzameling klassen, gebaseerd op een eigenschap van de lokale beeldinhoud, waarbij de verzameling klassen een eerste en tweede klasse omvat;
 - het bepalen van een eerste en tweede verzamelde positie voor de eerste en tweede klassen, gebruik makend van een mediaan of gemiddelde van de beeldposities van de lokale beeldinhoud die is toegewezen aan respectievelijk de eerste en tweede klas;
 - detecteren of aan een volwaarde voldaan is dat de eerste en tweede verzamelde posities aan onderling tegenoverliggende kanten van een referentiepositie liggen en afstand tussen de verzamelde posities of afstanden tussen de referentiepositie en de eerste en tweede verzamelde positie boven een drempelwaarde liggen;
 - selecteren van het beeld als het activerende beeld afhankelijk van of aan de genoemde voorwaarden voldaan is.
2. Een werkwijze volgens conclusie 1, waarin het selecteren van het beeld als het activerende beeld het detecteren omvat of aan genoemde voorwaarde voldaan is voor ten minste een vooraf behaalde fractie van de beelden voor beeldtijds punten in een tijdsinterval dat het activerende beeld omvat.
3. Een werkwijze volgens een der voorafgaande conclusies, omvattende het bepalen van een hoeveelheid beweging in relatie met het beeld, detecteren of aan een verdere voorwaarde voldaan is dat de hoeveelheid beweging onder een eerste drempel ligt en/of een afname in de hoeveelheid beweging vergeleken met een voorafgaand beeld een tweede drempel

overschrijdt en het beeld selecteren als activerend beeld afhankelijk van of aan genoemde verdere voorwaarde ook voldaan is.

4. Een werkwijze volgens een der voorafgaande conclusies, omvattende bepalen of een eerste volgorde van de eerste en tweede verzamelde posities langs een vooraf bepaalde richting in het activerende beeld omgekeerd is ten opzichte van een tweede volgorde langs de genoemde vooraf bepaalde richting van de eerste en tweede verzamelde posities uit een voorafgaand activerend beeld, en onderdrukken of kwalificeren van het activerende beeld afhankelijk van een resultaat van genoemde bepaling.
5. Een werkwijze volgens een der voorafgaande conclusies, omvattende het bepalen van een hoeveelheid beweging in eerste en tweede delen van de beelden voor tijdstippen die vooraf gaan aan het tijdstip van het activerende beeld, waarbij de eerste en tweede delen verschillende delen van een speelveld in beeld brengen, detecteren of de hoeveelheid beweging in het eerste deel groter is dan de hoeveelheid beweging in het tweede deel of vice versa en het kwalificeren van het activerende beeld afhankelijk van een resultaat van genoemde detectie.
6. Een werkwijze volgens een der voorafgaande conclusies, omvattende het opnemen van beeldstromen uit een veeltal camera's die gericht zijn op eenzelfde speelveld, gebruik van de beelden van de beeldstroom van een eerste camera van genoemde veeltal als genoemde reeks beelden, en in reactie op het genoemde selecteren van het beeld als het activerende beeld, extraheren van een videoclip die een gedeelte van de beeldstroom van een tweede camera omvat, opgenomen in een tijdsinterval dat vooraf gaat aan een tijdstip waarop het activerende beeld werd opgenomen.
7. Een werkwijze volgens conclusie 6, omvattende het genereren van een gecombineerde beeldstroom uit de beeldstromen van een veeltal camera's, door te schakelen tussen camera's van het veeltal op basis van een inhoud van de beeldstromen, waarbij de videoclip daarna geëxtraheerd wordt uit de gecombineerde beeldstroom.

8. Een werkwijze volgens conclusie 6 of 7, omvattende weergeven van de videoclip op een videodisplay, het beschikbaar maken van de videoclip voor weergave, en/of verzenden van de videoclip, in reactie op selectie van het beeld als het activerende beeld.
- 5 9. Een werkwijze volgens conclusie 6 of 7, omvattende het selectief bewaren van de videoclip in een video opslagsysteem in reactie op selectie van het beeld als het activerende beeld.
10. Een werkwijze volgens een der voorafgaande conclusies, omvattende het segmenteren van het beeld en het afleiden van genoemde beeldinhoud uit voorgrondgebieden die door genoemde segmenteren geïdentificeerd zijn.
- 10 11. Een werkwijze volgens conclusie 6, omvattende het genereren van een gecombineerde beeldstroom uit de beeldstromen van het veelal camera's, door te schakelen tussen camera's van het veelal op basis van een inhoud van de beeldstromen, waarbij de werkwijze een camera selectieproces bevat om te schakelen tussen de camera's, waarin het camera selectieproces het
- 15 segmenteren omvat van beelden van beeldstromen van elk van de camera's, bepalen van beeldposities van de voorgrond segmenten uit de beelden van de beeldstromen van elk van de camera's en het maken van omschakelingen tussen de camera's gebaseerd op de beeldposities, waarin de segmentatie die
- 20 uitgevoerd wordt in genoemde camera selectieproces wordt gebruikt als genoemde segmentatie en bepaling van de beeldposities voor het activerende beeld uit de beeldstroom voor de eerste camera.
12. Een videoverwerkingsstelsel voorzien van
- een ingang voor het ontvangen van een reeks beelden;
- 25 - een classificator voor het classificeren van lokale beeldinhoud uit verschillende voorgrondposities in elk van de beelden, waarbij de lokale beeldinhoud van elke verschillende voorgrondpositie wordt toegekend aan een van een verzameling vooraf bepaalde klassen, op basis van een eigenschap van de lokale beeldinhoud, waarbij de verzameling vooraf
- 30 bepaalde klassen een eerste en een tweede klas omvat;

- een positie verzamelaar ingericht om een eerste en tweede verzamelde positie te verzamelen voor de eerste en tweede klassen in elk van de beelden, gebruik makend van een mediaan of een gemiddelde van de beeldposities van de lokale beeldinhoud die respectievelijk is toegekend aan
5 de eerste en tweede klassen;
 - een detector geconfigureerd om te detecteren of voldaan is aan een voorwaarde dat de eerste en tweede verzamelde posities aan onderling tegenoverliggende kanten van een referentiepositie liggen en een afstand tussen verzamelde posities of afstanden tussen de referentiepositie en de
10 eerste en tweede verzamelde posities boven een drempelwaarde liggen.
13. Een videoverwerkingsysteem volgens conclusie 11, voorzien van een beeldsegmentatie module geconfigureerd om het beeld te segmenteren in voorgrondgebieden en achtergrondgebied, waarbij de classifier geconfigureerd is om, als genoemde lokale beeldinhoud voor de classificatie,
15 gebruik te maken van geaggregeerde beelddata uit voorgrondgebieden die door genoemde segmentatie module geïdentificeerd zijn.
14. Een videoverwerkingsysteem volgens 12 of 13, waarin het video opslag systeem is ingericht om een video clip af te spelen die beelden bevat die zijn opgenomen in een tijdinterval voorafgaand aan een tijd van opname
20 van het activerende beeld, in reactie op de detectie dat aan de voorwaarde voldaan is, en/of middelen omvat om in reactie op de detectie de videoclip te verzenden naar inrichtingen die in staat zijn om de videoclip af te spelen een signaal te sturen te signaleren dat de videoclip beschikbaar is voor weergave, en/of in reactie op de detectie een signaal te sturen naar de
25 inrichtingen dat de videoclip beschikbaar is voor weergave.
15. Een video opnamesysteem, voorzien het videoverwerkingsysteem volgens conclusie 12, 13 of 13, een veeltal camera's en een video opslagsysteem, waarbij de ingang van het video processingsysteem direct of indirect gekoppeld is aan een eerste camera van genoemde veeltal camera's,
30 waarbij de detector geconfigureerd is om een signaal af te geven aan een

video opslagsysteem in reactie op detectie dat aan de voorwaarden voldaan is voor een huidig beeld, om een videoclip af te spelen, de video clip te verzenden en/of boodschappen te versturen om te signaleren dat de videoclip beschikbaar is, waarbij het video opslagsysteem geconfigureerd is om in de
5 videoclip een deel van de beeldstroom van een tweede camera op te nemen dat is opgenomen in een tijdsinterval dat vooraf gaat aan een tijdstip van opnemen van het huidige beeld.

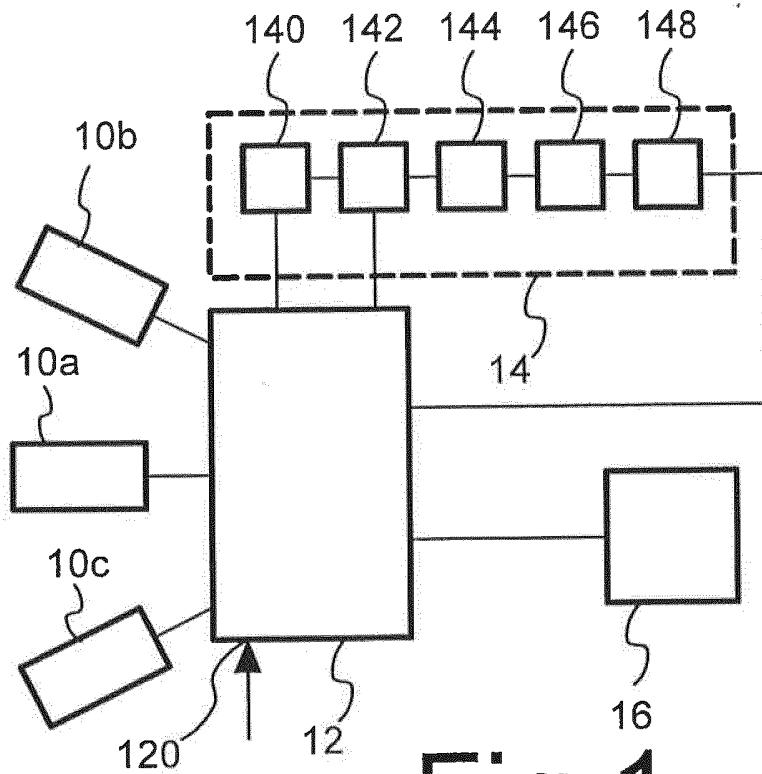


Fig. 1

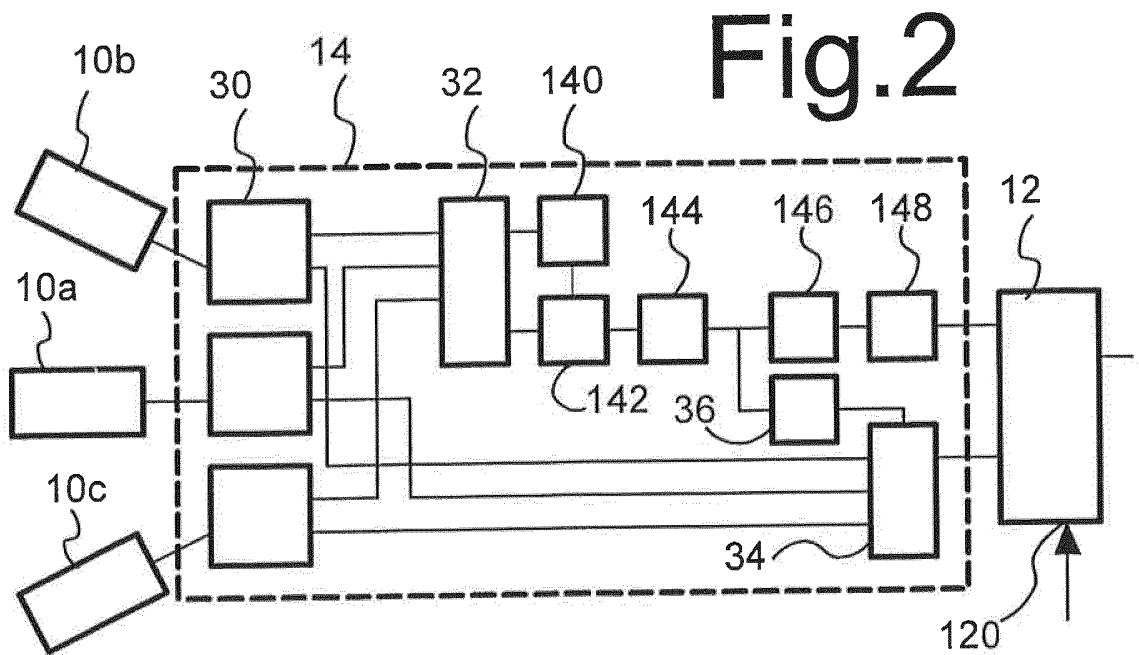


Fig. 2

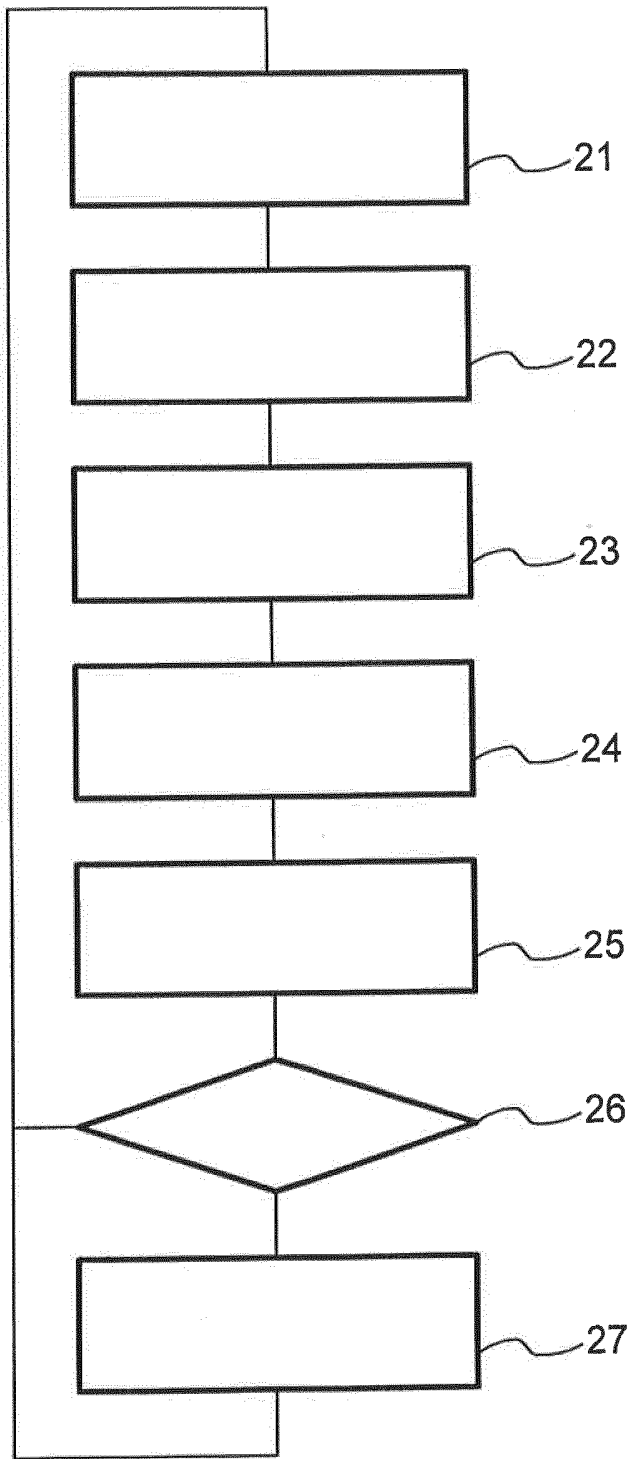
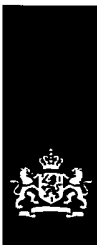


Fig.3



ONDERZOEKSRAPPORT

BETREFFENDE HET RESULTAAT VAN HET ONDERZOEK NAAR DE STAND VAN DE TECHNIEK

RELEVANTE LITERATUUR			
Categorie ¹	Literatuur met, voor zover nodig, aanduiding van tekstgedeelten of figuren.	Van belang voor conclusie(s) nr.	Classificatie (IPC)
X,D	ASSFALG J ET AL: "Semantic annotation of soccer videos: automatic highlights identification", COMPUTER VISION AND IMAGE UNDERSTANDING, ACADEMIC PRESS, US, deel 92, nr. 2-3, 1 november 2003 (2003-11-01), bladzijden 285-305, XP004472303, ISSN: 1077-3142, DOI: 10.1016/J.CVIU.2003.06.004 * samenvatting * * figuur 10 * * bladzijde 297, regels 6-13 * -----	1-15	INV. H04N21/44 G06K9/00 H04N21/845 H04N21/8549
A	BERTINI M ET AL: "Highlights modeling and detection in sports videos", PATTERN ANALYSIS AND APPLICATIONS, SPRINGER-VERLAG, LO, deel 7, nr. 4, 1 december 2004 (2004-12-01), bladzijden 411-421, XP019381468, ISSN: 1433-755X, DOI: 10.1007/S10044-004-0234-1 * bladzijden 416-417 * -----	1-15	Onderzochte gebieden van de techniek H04N G06K
Indien gewijzigde conclusies zijn ingediend, heeft dit rapport betrekking op de conclusies ingediend op:			
Plaats van onderzoek:	Datum waarop het onderzoek werd voltooid:	Bevoegd ambtenaar:	
's-Gravenhage	15 juli 2014	Naci, Suphi Umut	
¹ CATEGORIE VAN DE VERMELDE LITERATUUR			
<p>X: de conclusie wordt als niet nieuw of niet inventief beschouwd ten opzichte van deze literatuur</p> <p>Y: de conclusie wordt als niet inventief beschouwd ten opzichte van de combinatie van deze literatuur met andere geciteerde literatuur van dezelfde categorie, waarbij de combinatie voor de vakman voor de hand liggend wordt geacht</p> <p>A: niet tot de categorie X of Y behorende literatuur die de stand van de techniek beschrijft</p> <p>O: niet-schriftelijke stand van de techniek</p> <p>P: tussen de voorrangsdatum en de indieningsdatum gepubliceerde literatuur</p> <p>T: na de indieningsdatum of de voorrangsdatum gepubliceerde literatuur die niet bezwaard is voor de octrooiaanvraag, maar wordt vermeld ter verheldering van de theorie of het principe dat ten grondslag ligt aan de uitvinding</p> <p>E: eerdere octrooi(aanvraag), gepubliceerd op of na de indieningsdatum, waarin dezelfde uitvinding wordt beschreven</p> <p>D: in de octrooiaanvraag vermeld</p> <p>L: om andere redenen vermelde literatuur</p> <p>&: lid van dezelfde octrooifamilie of overeenkomstige octrooipublicatie</p>			

SCHRIFTELIJKE OPINIE

DOSSIER NUMMER NO138976	INDIENINGSDATUM 07.04.2014	VOORRANGSDATUM 05.04.2013	AANVRAAGNUMMER NL2012575
CLASSIFICATIE INV. H04N21/44 G06K9/00 H04N21/845 H04N21/8549			
AANVRAGER Nederlandse Organisatie voor Toegepast-Natuurwetenschappelijk Onderzoek TNO			

Deze schriftelijke opinie bevat een toelichting op de volgende onderdelen:

- Onderdeel I Basis van de schriftelijke opinie
- Onderdeel II Voorrang
- Onderdeel III Vaststelling nieuwheid, inventiviteit en industriële toepasbaarheid niet mogelijk
- Onderdeel IV De aanvraag heeft betrekking op meer dan één uitvinding
- Onderdeel V Gemotiveerde verklaring ten aanzien van nieuwheid, inventiviteit en industriële toepasbaarheid
- Onderdeel VI Andere geciteerde documenten
- Onderdeel VII Overige gebreken
- Onderdeel VIII Overige opmerkingen

	DE BEVOEGDE AMBTENAAR Naci, Suphi Umut
--	---

SCHRIFTELIJKE OPINIE

Aanvraag nr.:
NL2012575

Onderdeel I Basis van de Schriftelijke Opinie

1. Deze schriftelijke opinie is opgesteld op basis van de meest recente conclusies ingediend voor aanvang van het onderzoek.
2. Met betrekking tot **nucleotide en/of aminozuur sequenties** die genoemd worden in de aanvraag en relevant zijn voor de uitvinding zoals beschreven in de conclusies, is dit onderzoek gedaan op basis van:
 - a. type materiaal:
 - sequentie opsomming
 - tabel met betrekking tot de sequentie lijst
 - b. vorm van het materiaal:
 - op papier
 - in elektronische vorm
 - c. moment van indiening/aanlevering:
 - opgenomen in de aanvraag zoals ingediend
 - samen met de aanvraag elektronisch ingediend
 - later aangeleverd voor het onderzoek
3. In geval er meer dan één versie of kopie van een sequentie opsomming of tabel met betrekking op een sequentie is ingediend of aangeleverd, zijn de benodigde verklaringen ingediend dat de informatie in de latere of additionele kopieën identiek is aan de aanvraag zoals ingediend of niet meer informatie bevatten dan de aanvraag zoals oorspronkelijk werd ingediend.
4. Overige opmerkingen:

SCHRIFTELIJKE OPINIE

Aanvraag nr.:
NL2012575

Onderdeel V Gemotiveerde verklaring ten aanzien van nieuwheid, inventiviteit en industriële toepasbaarheid

1. Verklaring

Nieuwheid	Ja: Conclusies 1-15 Nee: Conclusies
Inventiviteit	Ja: Conclusies Nee: Conclusies 1-15
Industriële toepasbaarheid	Ja: Conclusies 1-15 Nee: Conclusies

2. Citaties en toelichting:

Zie aparte bladzijde

Onderdeel VIII Overige opmerkingen

De volgende opmerkingen met betrekking tot de duidelijkheid van de conclusies, beschrijving, en figuren, of met betrekking tot de vraag of de conclusies namerkbaar zijn, worden gemaakt:

Zie aparte bladzijde

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

Reference is made to the following documents; the numbering will be adhered to in the rest of the procedure.

- D1 ASSFALG J ET AL: "Semantic annotation of soccer videos: automatic highlights identification",
COMPUTER VISION AND IMAGE UNDERSTANDING, ACADEMIC PRESS, US,
vol. 92, no. 2-3, 1 November 2003 (2003-11-01), pages 285-305,
XP004472303,
ISSN: 1077-3142, DOI: 10.1016/J.CVIU.2003.06.004
- D2 BERTINI M ET AL: "Highlights modeling and detection in sports videos",
PATTERN ANALYSIS AND APPLICATIONS, SPRINGER-VERLAG, LO,
vol. 7, no. 4, 1 December 2004 (2004-12-01), pages 411-421,
XP019381468,
ISSN: 1433-755X, DOI: 10.1007/S10044-004-0234-1

- 1 In view of the clarity problems mentioned under Item VIII below, it is not possible to perform a thorough examination of the claimed invention and its technical properties. However, when interpreted in view of the description, the subject-matter of claim 1 does not appear to involve an inventive step for the following reasons:
- 1.1 In D1(p.297, l.6-13 and fig.10), a method for detecting kick off is disclosed, which performs the operations of detecting the players on the field, clustering them into teams based on their colors and checking whether the team players are distributed on each side of the midfield line.
- 1.1.1 Therefore the subject-matter of claim 1, when interpreted in view of the application as a whole, appear to differ from this known D1 in that it discloses a method for checking if the two clusters are sufficiently separated.
- 1.1.2 The solution proposed for this purpose, i.e. comparing the distance between the mean points of two clusters against a threshold, is the most straightforward way of doing this and, consequently, lacks inventive step.

- 2 The same arguments also apply to claim 12 which is the corresponding system claim.
- 3 Dependent claims of the application also do not appear to contain any additional features which, in combination with the features of any claim to which they refer, involve an inventive step.
- 3.1 One outstanding feature in the dependent claims is the use of motion amount, as a further feature for the estimation of the kick off scenes. This feature, however, also lacks inventive step, since the fact that the relative lack of motion in a scene is an indication of a kick-off event is well-known in the art, as disclosed in D2(p.416, second column, first paragraph under Section 3.3).

Re Item VIII

Certain observations on the international application

- 1 Claims 1 and 12 are not clear.
- 2 Claim 1 is not supported by the description, as its scope is broader than justified by the description and drawings:
 - 2.1 The application as a whole is related to detecting the kick-offs in a football match by extracting the players, classifying them into two teams and checking the distribution of the two groups of players with respect to the midfield line.
 - 2.2 In claim 1, however, the scope of the claim is defined broadly by abstracting *kick-off*, *players*, *teams* and *mid-field line* to *trigger image* (activerend beeld), *classes* (klassen) and *reference point* (referentiepositie), respectively.
 - 2.3 However, given that only the kick-off detection during a football match is given as an example, the disclosure of the application does not justify such a broad claim that covers the detection of any even in any kind of multimedia data.
 - 2.4 Hence the scope of the claim should be redefined to be limited to this only scenario disclosed in the application. Otherwise it is not clear for the skilled person in the art how to select the objects, classes and the reference point if, for example, the trigger image is the first frame of a film.
- 3 The term, *verzamelde posities liggen aan onderling tegenoverliggende kanten van een referentiepositie*, used in claim 1 is vague and unclear and leaves the reader in doubt as to the meaning of the technical features to which it refers, thereby rendering the definition of the subject-matter of said claims unclear.
 - 3.1 How two opposite sides of a point (*referentiepositie*) are defined is not clear. However, the *reference position* (*referentiepositie*) appears to be the midfield line, hence a line and NOT a point, in view of the application.
- 4 The same arguments also apply to claim 12 which is the corresponding system claim.