Title: A METHOD AND APPARATUS FOR GENERATING AN OUTPUT VIDEO SEQUENCE

Abstract: A video editing apparatus (101) is disclosed, comprising an extraction processor (107) operable to extract a plurality of video clips from an input video sequence, and a video characteristics processor (111) operable to determine a video clip characteristic for the video clips. An audio characteristics processor (119) determines an audio characteristic for audio segments of an audio item and a video sequence generator (121) automatically generates an output video sequence by combining the video clips with the audio items in response to the video clip characteristic and the audio characteristic. The video editing apparatus (101) thereby extracts significant video clips and synchronizes these to audio segments of the audio item. Hence, a shortened output video sequence is created with an associated soundtrack. The visual signal and soundtrack are correlated such that they have similar and/or compatible characteristics. Hence, an enhanced video sequence is created which is more interesting and entertaining to the general viewer.
Published: — with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
A method and apparatus for generating an output video sequence

The invention relates to a method and apparatus for generating an output video sequence and in particular to a system for generating an output video sequence by combining audio and video signals.

The advent of video cameras, and in the latter years, digital video cameras, has led to an increased popularity of capturing of video footage. Nowadays, video cameras are not just used for professional purposes but are widespread and frequently used by amateurs to create amateur videos. Most amateur videos are not subsequently edited due to the general lack of equipment, time and skill for performing this process. Furthermore, amateurs typically lack skill and experience for optimally filming video sequences and therefore tend to capture long and monotonous video sequences. Consequently, most home videos have a tendency to become long and uninteresting to the general viewer. This is further exacerbated by home videos typically comprising content which tends to be of more interest to the person filming than the viewer.

In recent years, the increased popularity of digital video cameras and home computers of increasing capability has resulted in the means for editing video footage being increasingly accessible to the amateur. In addition, many consumer electronic devices, such as DVD (Digital Versatile Disc) recorders, are becoming available which have some built-in video editing features. However, editing of video footage still requires a significant amount of skill to achieve good results and is furthermore very time consuming, especially for the amateur or infrequent user.

In professional applications, video sequences are typically enhanced by manually adding a soundtrack to the video footage. This further complicates the video editing process since not only must the video sequence be edited but this editing must furthermore be correlated with editing and adding of an audio soundtrack.

Accordingly, an improved system for creating an output video sequence based on an input video sequence would be advantageous and in particular a system for creating an output video sequence with an added audio component which allows for increased flexibility,
ease of use, reduced time consumption, reduced skill requirements and/or providing an improved video content would be advantageous.

Accordingly, the Invention seeks to mitigate, alleviate or eliminate one or more of the above mentioned disadvantages singly or in any combination.

According to a first aspect of the invention, there is provided a method of generating an output video sequence comprising: receiving an input video sequence; extracting a plurality of video clips from the input video sequence; determining a video clip characteristic for at least a first video clip of the plurality of video clips; receiving an audio item; determining an audio characteristic of at least a first audio segment of the audio item; and generating the output video sequence by combining the plurality of video clips with the audio item in response to the video clip characteristic and the audio characteristic.

The invention allows for an automatic combination of a video sequence and an audio item. The extraction of the plurality of video clips may for example be by dividing the entire video sequence into a plurality of video clips or by extracting only a subset of the entire video sequence thereby enabling a reduction in the duration of the video sequence. Advantageously, a plurality of video clips and audio segments may be combined in response to respective video clip characteristics and audio characteristics. The combination of the video clip and audio item may be combined without any user input based on the determined video clip characteristic and the audio characteristic. Hence, the video clip and audio segment may be combined such that they have properties that are compatible.

The invention allows for a method of generating a video sequence by automatic video editing. An improved video sequence may be generated which may be more entertaining to a viewer. Specifically, the output video sequence may comprise a video sequence which has been enhanced by addition of an audio soundtrack. For example, the audio item may be selected by a user and the video sequence may be processed to fit with the selected audio item. The invention allows for video editing which is easy to use, requires little skill, has low time consumption and/or has increased flexibility. The invention may be particularly advantageous for use with home amateur video sequence and may allow for an improved home video to be generated in an easy and fast way.

According to a feature of the invention, the step of generating comprises synchronising the plurality of video clips to the audio item in response to the video clip characteristic and the audio characteristic. For example, the plurality of video clips may be
synchronised to the audio item by synchronising the start of the first video clip to the start of the audio segment. The audio item and video sequence may automatically be synchronised and a synchronised soundtrack may thus automatically be added to the video sequence. The synchronisation is not necessarily a synchronisation of all aspects of the video sequence and the audio item but may for example be a partial synchronisation wherein only one parameter is synchronised for a limited duration.

According to a different feature of the invention, the synchronisation is by synchronising at least one video clip to the first audio segment when the video clip characteristic is associated with the audio characteristic. The association between the video clip characteristic and the audio characteristic preferably indicates a compatibility or suitability of the audio segment as a soundtrack for the first video clip. Hence, a synchronisation of a video clip with a suitable audio segment of the audio item is enabled.

According to a different feature of the invention, the method further comprises the step of determining an association between the video clip characteristic and the audio characteristic in response to a predetermined relationship between video characteristics and audio characteristics. Preferably, the predetermined relationship comprises stored associations between video clip characteristics and audio characteristics. This allows for a very simple and easy to implement yet efficient method of associating the video clip characteristic and the audio characteristic and thereby the first video clip and the first audio segment. The predetermined relationship may for example be determined by accessing a data store comprising a look-up table wherein possible values of the video clip characteristic are linked to possible values of the audio characteristic.

According to a different feature of the invention, the predetermined relationship comprises an association rule between video clip characteristics and audio characteristics. This allows for a high degree of flexibility and allows a high number of possible characteristics yet is simple to implement and enables an efficient and suitable association to be determined.

According to a different feature of the invention, the step of determining the association between the video clip characteristic and the audio characteristic is further in response to a user input. Preferably the determination of the association is not just automatic but may include a user input. This allows for increased flexibility and improved adaptability to a user’s specific preferences for the current video clip characteristic and audio characteristic. The user input may directly relate to the association, such as for example specifying an association rule or may indirectly relate to the association. For example, the
association may be determined by a manual linking of the audio characteristic and the video clip characteristic or of the video clip and the audio segment.

According to a different feature of the invention, the method further comprises the step of updating the predetermined relationship in response to the determined association between the video clip characteristic and the audio characteristic. Preferably, the predetermined relationship is updated when the step of determining an association is in response to the user input. This allows for the predetermined relationship to be updated to more accurately reflect a user’s preference. Hence, a learning system may be enabled wherein the association is automatically adapted to a user’s preferences.

According to a different feature of the invention, the video clip characteristic comprises at least one characteristic chosen from the group of: a picture colour characteristic; a scene change frequency characteristic; a picture brightness characteristic; and an object motion characteristic. These parameters are particularly suitable for determining if a specific audio segment is suitable for being combined with the first video clip.

According to a different feature of the invention, the step of extracting the plurality of video clips comprises extracting the plurality of video clips in response to a video content characteristic. The video content is frequently important in determining a suitable audio clip to be associated with a video clip. An improved suitability of the combination of the plurality of video clips with the audio item may thus be achieved. The video content characteristic may be determined by content analysis of the video sequence.

According to a different feature of the invention, the step of extracting the plurality of video clips comprises extracting the plurality of video clips in response to the audio characteristic. This allows for an increased correlation between the video clips and the audio item.

According to a different feature of the invention, the audio characteristic comprises at least one parameter chosen from the group of: a beat frequency parameter; a genre parameter; a voice detection parameter; and a music characteristic parameter. These parameters are particularly suitable for determining if a specific audio segment is suitable for being combined with the first video clip.

According to a different feature of the invention, the method further comprises the step of determining the audio characteristic in response to metadata associated with the audio item. This allows for a simple and fast approach to determining the audio characteristic. The metadata may for example be stored with the audio item or be embedded in the audio item.
According to a different feature of the invention, the method further comprises the step of selecting the audio item from a plurality of stored audio items in response to a user input. This provides an advantageous and simple way of obtaining audio items.

According to a different feature of the invention, the method further comprises the step of processing the video clip in accordance with a processing criterion determined in response to the audio characteristic. This allows for a further enhancement of the output video sequence. For example, special video effects may be applied to the first video clip. The video effects may be selected to suit the audio segment and may particularly be correlated or synchronised with characteristics of the audio segment.

According to a different feature of the invention, the step of extracting the plurality of video clips is operable to extract the video clips in response to a duration of the first audio segment. This allows for a simple and highly efficient synchronisation of the plurality of video clips and audio segments of the audio item. For example, the audio item may comprise a plurality of audio segments with known audio characteristics and of a given duration. A video clip of the video sequence may be extracted for each of these segments and selected such as to suit the audio characteristics of the individual audio segments.

According to a second aspect of the invention, there is provided an apparatus for generating an output video sequence comprising: means for receiving an input video sequence; means for extracting a plurality of video clips from the input video sequence; means for determining a video clip characteristic for at least one video clip of the plurality of video clips; means for receiving an audio item; means for determining an audio characteristic of at least a first audio segment of the audio item; and means for generating the output video sequence by combining the plurality of video clips with the audio item in response to the video clip characteristic and the audio characteristic.

These and other aspects, features and advantages of the invention will be apparent from and elucidated with reference to the embodiment(s) described hereinafter.

An embodiment of the invention will be described, by way of example only, with reference to the drawings, in which

FIG. 1 illustrates a block diagram of an apparatus for generating an output video sequence in accordance with an embodiment of the invention; and

FIG. 2 shows a flowchart of a method of generating an output video sequence in accordance with an embodiment of the invention.
The following description focuses on an embodiment of the invention applicable to a digital video editing apparatus. However, it will be appreciated that the invention is not limited to this application but may be applied to many other applications. In the following, the reference to a video sequence includes a reference to a video sequence having an associated audio signal. Thus for example, a video sequence from a digital video camera may include both recorded video and audio.

FIG. 1 illustrates a block diagram of an apparatus for generating an output video sequence in accordance with an embodiment of the invention. Specifically, the apparatus may be a video editing apparatus 101 such as a personal computer with video editing software or a consumer device with video editing functionality.

The video editing apparatus 101 comprises an interface 103 for receiving a video signal comprising a video sequence. In the preferred embodiment, the interface 103 receives the input video sequence from an external video source 105. In other embodiments, the video sequence is received from other sources, such as an internal video storage or internal video generation means. Specifically, the video editing apparatus may itself be a digital video camera comprising internal means for both generating and storing a video sequence.

The interface is coupled to an extraction processor 107. The extraction processor 107 is operable to extract a plurality of video clips from the input video sequence. In the preferred embodiment, the encode processor extracts a number of video clips from the video sequence corresponding to the most significant segments of the video sequence. Any suitable criterion for determining which segments are to be considered significant may be used without detracting from the invention. Thus, the extraction processor 107 generates a number of video clips which are to be combined to form a new video sequence that is preferably, but not necessarily, of shorter duration than the video sequence. Each video clip may in some embodiments comprise a plurality of sub-video clips. These sub-video clips may be extracted from different parts of the video sequence.

The extraction processor 107 is coupled to a video storage 109 wherein the extracted video clips are temporarily stored. The video storage 109 is coupled to a video characteristics processor 111. The video characteristics processor 111 is operable to determine a video clip characteristic for at least one video clip of the plurality of video clips. Any suitable video clip characteristic may be used without detracting from the invention. As a specific example, the video characteristics processor 111 may derive an indication of the
amount and rate of motion in the video clip from an analysis of the correlation between pictures or frames. For example, the video characteristics processor 111 may comprise functionality for object tracking as is well known in the art. In the specific example a video clip characteristic specifically relating to the degree of motion within each video clip is determined for all of the extracted video clips.

The video editing apparatus 101 furthermore comprises an internal audio storage 113 wherein a plurality of audio items is stored. The audio items may for example be various music items or sound effects. Preferably, a large variety of audio items with diverse characteristics are stored. For example, different audio items may consist corresponding to different music items (e.g. different songs) and preferably the audio storage comprises a large quantity of music items in different genres and with different characteristics. For example, some music may be slow and atmospheric whereas other music may be fast and powerful. Furthermore, each of the audio items will typically have varying characteristics in different segments of the audio item. For example a music item may have a fast and loud segment as well as slow and quieter segment, or a music item such as a song may comprise segments corresponding to instrumental parts of the song, segments corresponding to the chorus and segments corresponding to the verses. An audio item may furthermore be an amalgamation of other items or audio clips. For example, an audio item may comprise different music items (or songs), and each song may for example be considered an audio segment of the audio item.

In some embodiments and applications, an audio segment may correspond to the entire audio item.

The audio storage 113 is coupled to an audio storage interface 115. The audio storage interface 115 is operable to receive an audio item from the audio store 113. In the preferred embodiment, the audio storage interface 115 comprises means for receiving a user input from a user and means for retrieving an audio item from the audio storage in response thereto. Hence, a user may select an audio item in the audio storage 113 and the audio storage interface 115 may retrieve the selected audio item therefrom. In other embodiments, the audio item may be received from other means such as an external audio item source or an internal audio item generator.

The audio storage interface 115 is coupled to an audio buffer 117 wherein the selected audio item is temporarily stored. The audio buffer 117 is coupled to an audio characteristics processor 119. The audio characteristics processor 119 is operable to determine an audio characteristic of at least a first audio segment of the audio item. In some embodiments, the audio characteristics processor 119 comprises functionality for dividing the
audio item into a number of different segments and for determining an audio characteristic of each of the segments. In other embodiments, the audio item is divided into segments when fed to the audio characteristics processor 119. For example, the audio item may be stored in segments in the audio storage 113 or may be stored with data indicative of different segments. Specifically, an audio item may consist in a single audio segment. As a specific example, the audio item may be a song and the audio characteristics processor 119 may be operable to determine the pace of the music in different segments. Specifically, it may determine the beat rate throughout the song and divide the audio item into segments according to the changes in the rate of the beat. Thus, some segments may be generated that are relatively slow and some segments may be generated which are relatively fast.

The video editing apparatus 101 furthermore comprises a video sequence generator 121. The video sequence generator 121 is coupled to the video characteristics processor 111 and the audio characteristics processor 119 and receives the audio characteristics and video clip characteristics from these.

Furthermore, the video sequence generator 121 is coupled to the audio buffer 117 and the video storage 109 and is operable to retrieve the audio item and the video clips from these. The video sequence generator 121 is operable to generate an output video sequence by combining the plurality of video clips with the audio item in response to the video clip characteristic and the audio characteristic. In the preferred embodiment, the video sequence generator 121 combines each video clip with an audio segment of the audio item having an audio characteristic compatible with the video clip characteristic of the video clip. Thus, audio segments and video clips are preferably paired such that they have similar and/or compatible characteristics. In the specific example, video clips and audio segments may be combined such that video clips having a video clip characteristic indicating a high degree and rate of motion are combined with audio segments having a high beat rate; and video clips having a video clip characteristic indicating a relatively lower degree and rate of motion are combined with the audio segments having an audio characteristic indicating a relatively lower beat rate.

The combined audio segments and video sequences are thus combined into an output video sequence. Preferably, this is done by a simple consecutive arrangement of the combined video clips and audio segments. In the preferred embodiment, the audio item is unchanged and thus the order and duration of the audio segments are unchanged whereas the video clips are re-ordered in order to suit the order of the audio segments in the audio item. Specifically, in the preferred embodiment an audio item is thus selected by a user and the
video sequence is processed or manipulated to suit or match the characteristics of the selected audio item. However, in other embodiments, the audio segments may be re-arranged to fit an order of the video clips or both the video clips and the audio segments may be re-arranged. The output video sequence is consequently output to a video source 123 which specifically may be an internal or external video storage device.

In the preferred embodiment, the video editing apparatus may thus receive a video sequence from an external source and combine it with a selected audio item. The video editing apparatus may extract significant clips and synchronise these with audio segments of the audio item. Hence, a preferably shortened output video sequence may thus be created with an associated soundtrack. The visual signal and soundtrack are furthermore preferably correlated such that they have similar and/or compatible characteristics. Hence, an enhanced video sequence is created which is more interesting and entertaining to the general viewer.

The approach may be particularly useful for automatically or semi- automatically creating music video type video sequences from a home video sequence.

As a special example, a video editing apparatus in accordance with an embodiment of the invention may comprise an algorithm that combines home videos and songs into a home music video. The user may select a home video sequence and an audio item in the form of a song. The video may be analyzed and the more important scenes extracted. Scene changes and other relevant characteristics, such as colour and brightness of the images, are determined. Additionally or alternatively the extraction of the scenes may be performed in response to the characteristics. The audio item may be analyzed to extract characteristics, such as beats, melody changes, singing parts and solos, and the audio item may be segmented accordingly. The music selected by the user may furthermore be used as a parameter for selecting the type of music video to be generated and possibly the effects to be applied to the images. For example, if the selected music audio item genre is rap, fast changes and short video clips are applied whereas if a slow song is used longer video clips may be used. The extracted video clips may be combined with the audio segments such that the content characteristics of the video clips correlate with the audio characteristics of the segments. Additionally, special effects may be applied to the video images in response to the determined video clip characteristics or audio characteristics. The information related to e.g. the music genre and speed may be extracted directly from the audio item or may be communicated by the user, or retrieved from a service by using the Internet.

FIG. 2 shows a flowchart of a method 200 for generating an output video sequence in accordance with an embodiment of the invention. The method is applicable to
the video editing apparatus 101 of FIG. 1 and will be described with reference to this. It will be appreciated that, whereas the method is illustrated as a sequential execution of separate steps, any suitable order or time of execution of these steps may be used without subtracting from the invention. Specifically, some or all steps may as appropriate be executed sequentially as illustrated or may be executed in parallel. Furthermore, for clarity and brevity, the steps of the method are illustrated as separate steps but it will be apparent that the illustrated steps may be combined or that steps may be divided into further sub-steps as appropriate for the specific embodiment.

The method initiates in step 201 wherein an input video sequence, such as a home video, is received from the external video source, such as a digital video camera. Step 201 is followed by step 203 wherein the plurality of video clips is extracted from the input video sequence by the extraction processor 107. Any suitable algorithm or criteria for extracting video clips may be used. However, in the preferred embodiment, the video clips are extracted in response to a video picture content characteristic. Hence, a content analysis may be performed on the video sequence and, the most significant scenes and sequences may be extracted in response to this analysis. For example, if the video sequence is a video sequence of a football game, audio contents of the video soundtrack may be extracted to detect times of increased spectator volume. This is typically associated with highlights of the game, such as goals or penalties. Thus, the video clips may automatically be extracted to correspond to the highlights of the football game.

In some embodiments, the extraction of video clips comprises a shortening of the existing scenes recorded. For example, it is typical for home videos that very long clips are recorded. Often, several minutes are recorded wherein little occurs. In some embodiments, the extraction may for example include extraction of only the first 10-20 seconds of each recording clip thereby creating a much shortened video sequence with significantly increased amount of activity.

In some embodiments, the extraction of video clips is in response to at least one audio characteristic of the audio item or one or more of the audio segments. For example, if an audio item is selected which is in a genre characterised by high intensity, fast rhythm and loud volumes, the extraction processor 107 predominantly extracts video clips characterised by high degrees and rates of movement and rapid scene changes. However, if the audio item is characterised by being slow, atmospheric and gentle, video clips will be extracted which predominantly comprise little and/or slow movement and slow scene
changes. Thereby, the video sequence generated from the video clips will closely correlate with the characteristics of the selected audio item.

In some embodiments, the extraction processor 107 is operable to extract the video clips in response to a duration of segments of the audio item. For example, an audio item may comprise a first audio segment of 15 seconds and a second audio segment of 20 seconds duration. In this case, the extraction processor 107 may determine a video segment which has a close correlation with the characteristics of the first audio segment and generate a first video clip by extracting the first 15 seconds of this video segment. It may then proceed to determine a video segment which has a close correlation with the characteristics of the second audio segment and generate a second video clip by extracting the first 20 seconds of this video segment. This may be repeated for as many audio segments as comprised in the audio item.

Step 203 is followed by step 205. In step 205, the video characteristics processor 111 extracts a video clip characteristic for at least one and preferably all of the extracted video clips. In the preferred embodiment, the video clip characteristic comprises one or more of the following parameters:

A picture colour characteristic: A general colour characteristic or a colour of a part of the picture can be used to associate the video clip with an audio segment. For example, a picture comprising predominantly bright primary colours may indicate that the video sequence comprises a cartoon or other video content suited for children, and it may therefore preferably be combined with audio segments also being suited for children.

A picture brightness characteristic: A video clip having a high level of brightness may be combined with audio segments having a cheerful and bright nature whereas a low level of brightness may indicate a suitability for combination with slow and menacing music.

A scene change frequency characteristic: For example, fast scene changes indicates a high level of action and therefore a suitability for combination with a fast paced audio item.

An object motion characteristic: Similarly, a high degree and rate of motion indicates a suitability for combination with fast paced music and a low degree and rate of motion indicates a suitability for combination with slow paced music.

Step 205 is followed by step 207 wherein an audio item is received by the storage interface 115 from the audio storage 113 and stored in the audio buffer 117. In the preferred embodiment, a user input is received and an audio item is selected and retrieved in
response. Specifically, the video editing apparatus 101 may comprise a display listing all the audio items stored in the audio storage 113 and the user input may indicate a selection of a displayed audio item. The user may for example select an audio item prior to receiving the video sequence, which may accordingly be processed to suit the selected audio item.

Step 207 is followed by step 209 wherein the audio characteristics processor 119 determines an audio characteristic of at least one audio segment of the selected audio item. The audio characteristics processor 119 may for example determine an audio characteristic comprising one or more of the following parameters:

A beat frequency parameter: For example, the frequency of a drum beat may be determined from a low pass filtering of the audio segment followed by detection of peaks in the resulting signal. The beat frequency is indicative of a speed of the music and may thus be used to combined the video clips with the audio segments.

A genre parameter: For example, the genre of the music may be indicative of which video clips the music may advantageously be combined with. For example, audio items related to children may advantageously be combined with video clips predominantly comprising bright colours.

A voice detection parameter: For example, voice detection may be used to determine whether an audio item is an instrumental music item and a suitable video clip may advantageously be determined in accordance therewith.

A music characteristic parameter: The audio characteristic may, for example, comprise an indication of whether the audio item comprises predominantly rhythmic (e.g. rock music) or non rhythmic music (e.g. classical music) and this may be used to determine a suitable video clip.

In the preferred embodiment, the audio items are pre-categorised and stored with information related to characteristics of the audio item. Specifically, meta-data may be embedded in the audio item data or stored with the audio items and this meta-data preferably comprises information related to for example the above described parameters. Thus the audio characteristics processor 119 may simply determine the audio characteristic by extracting the meta-data associated with the selected audio item.

In the preferred embodiment, steps 203 and 205 are performed in parallel to steps 207 and 209.

Step 209 is followed by step 211 wherein the output video sequence is generated by the video sequence generator 121 by combining the plurality of video clips with the audio item in response to the video clip characteristic and the audio characteristic. In the
preferred embodiment, the video sequence generator 121 pairs the video clips and audio segments such that the audio characteristic of an audio item and the corresponding video clip has the best match in accordance with a combination criterion.

In the preferred embodiment, the combination is in response to a predetermined relationship between video characteristics and audio characteristics. For example, an association between possible values of an audio item and possible values of a video clip may be predetermined.

The association may be based on an association rule. An example of an association rule may be to determine an association value as the difference between a scene change frequency of a video clip and fifty times the beat rate frequency. The pairing may then be performed such that this association value is minimised. Thus the video clip will be selected which has a scene change frequency most closely matching fifty times the beat frequency. This will result in an output video sequence wherein a scene typically lasts for around fifty beats. Thus if fast music is selected, fast scene changes occur in the output video sequence, and if slow music is selected slow scene changes occur.

Alternatively or additionally, the predetermined relationship may comprise stored associations between video clip characteristics and audio characteristics. For example, a preferred relationship between picture brightness and a music genre may be stored. Hence, the music genre may be used to access the storage and a preferred associated brightness level may be retrieved. The video clip may then be selected to most closely attain this value.

In the preferred embodiment, a plurality of parameters of the audio characteristic and video clip characteristic are used to determine a preferred association between video clips and audio segments. In some embodiments, the association is furthermore determined in response to a user input. The user input may directly specify an association rule or a specific association. For example, a set of rules for association may be inputted and may override or modify the existing association rules. However, the user input may also directly pair a video clip and audio segment and an association or association rule may be derived from this pairing. The user input is preferably used to modify the stored associations or the association rule whereby the user's preference may gradually be included in the automatically determined associations.

In the preferred embodiment, the video clips are thus synchronised to the audio segments. Thus, typically, when one audio segment is followed by a new audio segment, a new video clip will replace the previous video clip. Hence, the variations in the
audio item are followed by variations in the output video sequence. For example, the rate of scene changes may vary in line with the beat frequency of the music.

Step 211 is followed by step 213 wherein the video sequence generated by combining the video clips and the audio segments are output to the video source 123.

In some embodiments, the video clips may furthermore be processed in accordance with a processing criterion determined in response to the audio characteristic. Specifically, the processing may comprise adding a visual special effect to the video clip. The visual special effect may be selected in accordance with the audio characteristic of the associated audio segment. For example, if the audio segment corresponds to slow music, diffusion effects may be applied to the images of the video clip.

In the preferred embodiment, the video sequence comprises not only video but also comprise associated audio. In this case, the audio of the video sequence is preferably combined with the audio of the audio item. For example, the audio item may be mixed with the audio of the video sequence but at a reduced volume thereby providing an underlay audio soundtrack but allowing the audio of the video sequence to be heard.

Thus, the preferred embodiment of the invention allows for a simple, fast and automated method of generating an entertaining video sequence. In particular a music video type video sequence may be generated from a home video sequence.

The invention can be implemented in any suitable form including hardware, software, firmware or any combination of these. However, preferably, the invention is implemented as computer software running on one or more data processors and/or digital signal processors. The elements and components of an embodiment of the invention may be physically, functionally and logically implemented in any suitable way. Indeed the functionality may be implemented in a single unit, in a plurality of units or as part of other functional units. As such, the invention may be implemented in a single unit or may be physically and functionally distributed between different units and processors.

Although the present invention has been described in connection with the preferred embodiment, it is not intended to be limited to the specific form set forth herein. Rather, the scope of the present invention is limited only by the accompanying claims. In the claims, the term comprising does not exclude the presence of other elements or steps. Furthermore, although individually listed, a plurality of means, elements or method steps may be implemented by e.g. a single unit or processor. Additionally, although individual features may be included in different claims, these may possibly be advantageously combined, and the inclusion in different claims does not imply that a combination of features
is no feasible and/or advantageous. In addition, singular references do not exclude a plurality.
Thus references to "a", "an", "first", "second" etc do not preclude a plurality.

In summary, the invention relates to a video editing apparatus (101) comprising an extraction processor (107) operable to extract a plurality of video clips from an input video sequence, and a video characteristics processor (111) operable to determine a video clip characteristic for the video clips. An audio characteristics processor (119) determines an audio characteristic for audio segments of an audio item and a video sequence generator (121) automatically generates an output video sequence by combining the video clips with the audio items in response to the video clip characteristic and the audio characteristic. The video editing apparatus (101) thereby extracts significant video clips and synchronises these to audio segments of the audio item. Hence, a shortened output video sequence is created with an associated soundtrack. The visual signal and soundtrack are correlated such that they have similar and/or compatible characteristics. Hence, an enhanced video sequence is created which is more interesting and entertaining to the general viewer.
CLAIMS:

1. A method of generating an output video sequence comprising the steps of:
   receiving (201) an input video sequence;
   extracting (203) a plurality of video clips from the input video sequence;
   determining (205) a video clip characteristic for at least a first video clip of the
   plurality of video clips;
   receiving (207) an audio item;
   determining (209) an audio characteristic of at least a first audio segment of
   the audio item; and
   generating (211) the output video sequence by combining the plurality of
   video clips with the audio item in response to the video clip characteristic and the audio
   characteristic.

2. A method of generating an output video sequence as claimed in claim 1
   wherein the step of generating (211) comprises synchronising the plurality of video clips to
   the audio item in response to the first video clip characteristic and the audio characteristic.

3. A method of generating an output video sequence as claimed in claim 2
   wherein the synchronising is by synchronising the first video clip to the first audio segment
   when the video clip characteristic is associated with the audio characteristic.

4. A method of generating an output video sequence as claimed in claim 1 further
   comprising the step of determining an association between the video clip characteristic and
   the audio characteristic in response to a predetermined relationship between video
   characteristics and audio characteristics

5. A method of generating an output video sequence as claimed in claim 4
   wherein the predetermined relationship comprises stored associations between video clip
   characteristics and audio characteristics.
6. A method of generating an output video sequence as claimed in claim 4 wherein the predetermined relationship comprises an association rule between video clip characteristics and audio characteristics.

7. A method of generating an output video sequence as claimed in claim 4 wherein the step of determining the association between the video clip characteristic and the audio characteristic is further in response to a user input.

8. A method of generating an output video sequence as claimed in claim 4 further comprising the step of updating the predetermined relationship in response to the determined association between the video clip characteristic and the audio characteristic.

9. A method of generating an output video sequence as claimed in claim 1 wherein the video clip characteristic comprises at least one characteristic chosen from the group of:
   a picture colour characteristic;
   a scene change frequency characteristic;
   a picture brightness characteristic; and
   an object motion characteristic.

10. A method of generating an output video sequence as claimed in claim 1 wherein the step of extracting (203) the plurality of video clips comprises extracting the plurality of video clips in response to a video picture content characteristic.

11. A method of generating an output video sequence as claimed in claim 1 wherein the step of extracting (203) the plurality of video clips comprises extracting the plurality of video clips in response to the audio characteristic.

12. A method of generating an output video sequence as claimed in claim 1 wherein the audio characteristic comprises at least one parameter chosen from the group of:
    a beat frequency parameter;
    a genre parameter;
    a voice detection parameter; and
    a music characteristic parameter.
13. A method of generating an output video sequence as claimed in claim 1 further comprising the step of determining the audio characteristic in response to metadata associated with the audio item.

14. A method of generating an output video sequence as claimed in claim 1 further comprising the step of selecting the audio item from a plurality of audio items in response to a user input.

15. A method of generating an output video sequence as claimed in claim 1 further comprising the step of processing the first video clip in accordance with a processing criterion determined in response to the audio characteristic.

16. A method of generating an output video sequence as claimed in claim 1 wherein the step extracting (203) the plurality of video clips is operable to extract the video clips in response to a duration of the first audio segment.

17. A computer program enabling the carrying out of a method according to claim 1.

18. A record carrier comprising a computer program as claimed in claim 16.

19. An apparatus (101) for generating an output video sequence comprising:
   means (103) for receiving an input video sequence;
   means (107) for extracting a plurality of video clips from the input video sequence;
   means (111) for determining a video clip characteristic for at least one video clip of the plurality of video clips;
   means (115) for receiving an audio item;
   means (119) for determining an audio characteristic of at least a first audio segment of the audio item; and
   means (121) for generating the output video sequence by combining the plurality of video clips with the audio item in response to the video clip characteristic and the audio characteristic.
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G11B27/034

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G11B G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>WO 02/052565 A (ALTMAN EDWARD JAMES; KELLOCK PETER ROWAN (SG); KENT RIDGE DIGITAL LAB) 4 July 2002 (2002-07-04) page 39, line 29 - page 42, line 27</td>
<td>1-19</td>
</tr>
<tr>
<td>X</td>
<td>WO 00/39997 A (DEKEL ELAN; EARTHNOISE COM INC (IL)) 6 July 2000 (2000-07-06) the whole document</td>
<td>1-19</td>
</tr>
</tbody>
</table>

Other documents are listed in the continuation of box C.

Patent family members are listed in annex.

Date of the actual completion of the international search: 21 May 2004

Date of mailing of the international search report: 01/06/2004

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax. (+31-70) 340-3016

Authorized officer:
Mourik, J.
<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family number(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>EP 1354318 A1</td>
<td>22-10-2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GB 2380599 A, B</td>
<td>09-04-2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2004027369 A1</td>
<td>12-02-2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 0039997 A2</td>
<td>06-07-2000</td>
</tr>
</tbody>
</table>