(12) UK Patent Application (19) GB (11) 2 404 298 (13) A

(43) Date of A Publication 26.01.2005

(21) Application No: 0317306.9

(22) Date of Filing: 24.07.2003

(71) Applicant(s):
Hewlett-Packard Development Company L.P., 20555 S.H.249, Houston, Texas 77070, United States of America

(72) Inventor(s):
Stephen Philip Cheatle

(74) Agent and/or Address for Service:
Hewlett-Packard Limited
Intellectual Property Section, Building 3, Filton Road, Stoke Gifford, BRISTOL, BS34 8QZ, United Kingdom

(51) INT CL^7:
H04N 5/268 , G11B 27/031

(52) UK CL (Edition X):
H4F FGG

(56) Documents Cited:
EP 0782139 A1 **JP 100108071 A**
JP 2002218367 A

(58) Field of Search:
UK CL (Edition V) **H4F**
INT CL^7 **G11B, H04N**
Other: **Online: WPI, EPODOC, JAPIO**

(54) Abstract Title: **Use of saliency in media editing**

(57) An image signal V from a still and/or video camera recording is edited to provide a programme. A multi-valued saliency signal S accompanies the image signal or is generated therefrom, and a user specifies the value of at least one characteristic of the programme as a whole. An editing signal indicative of selected portions of the image signal having higher saliency values is generated at least partly in response to the signal S so that the value of the programme characteristic(s) equals or approximates the specified value. As shown for a video signal V, a variable saliency threshold T is set so that portions CA to CF are selected having a total length equal to a specified programme length. The vertical lines in plot S indicate adjustment of portion lengths so that no selected portion lies outside a specified length range.

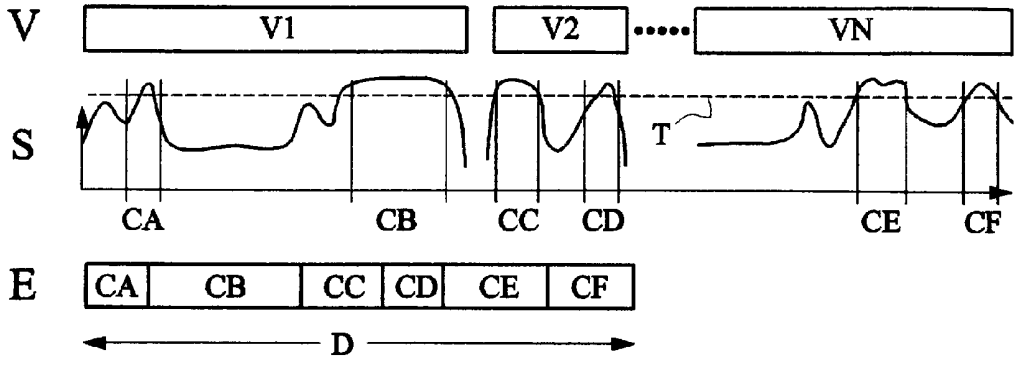

FIG. 3

GB 2 404 298 A

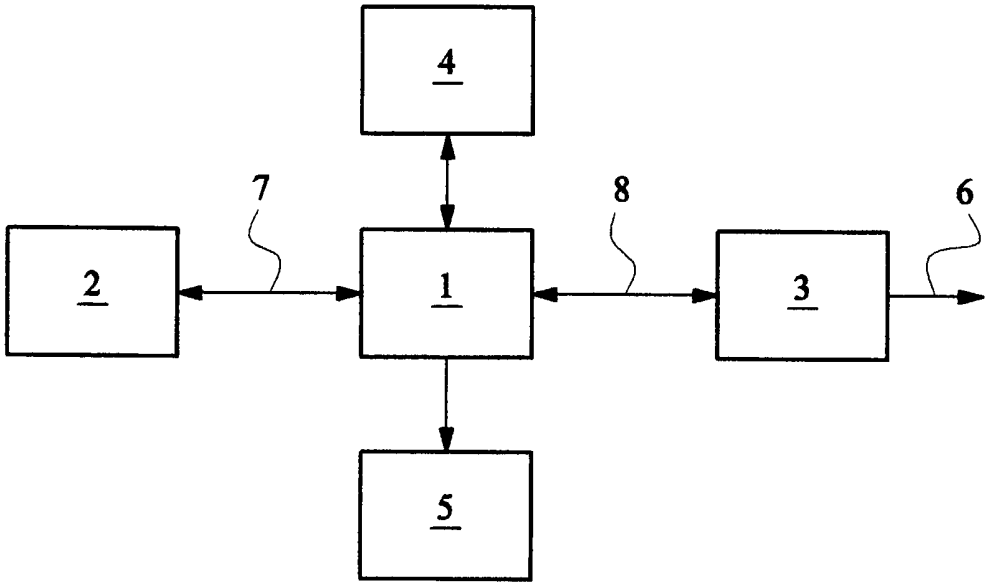
4
7
8
6
2
1
3
5

FIG. 1

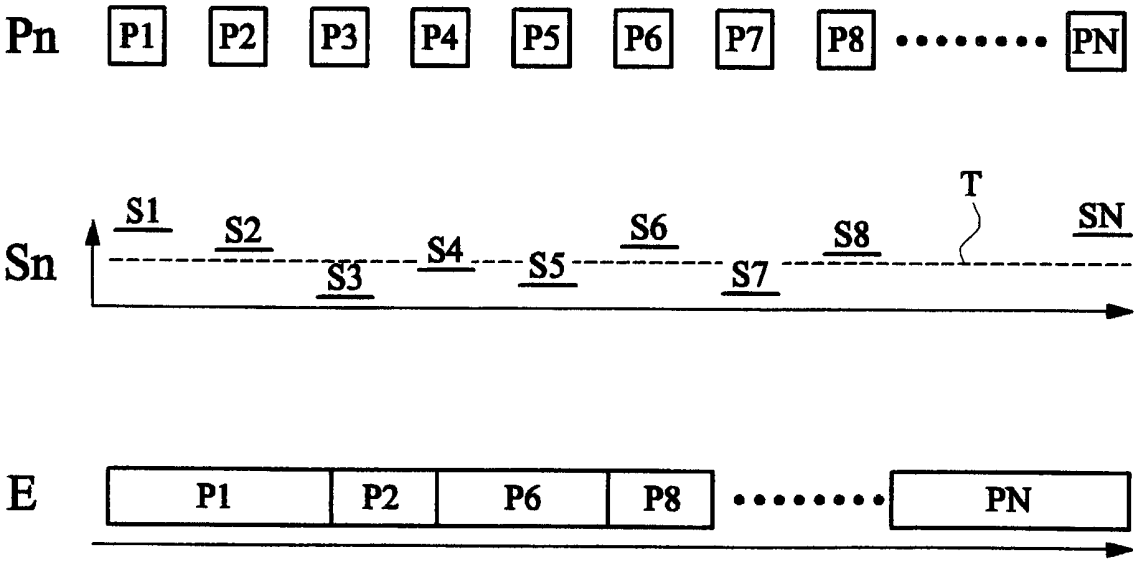
Pn
P1
P2
P3
P4
P5
P6
P7
P8
PN
Sn
S1
S2
S3
S4
S5
S6
S7
S8
T
SN
E
P1
P2
P6
P8
PN

FIG. 2

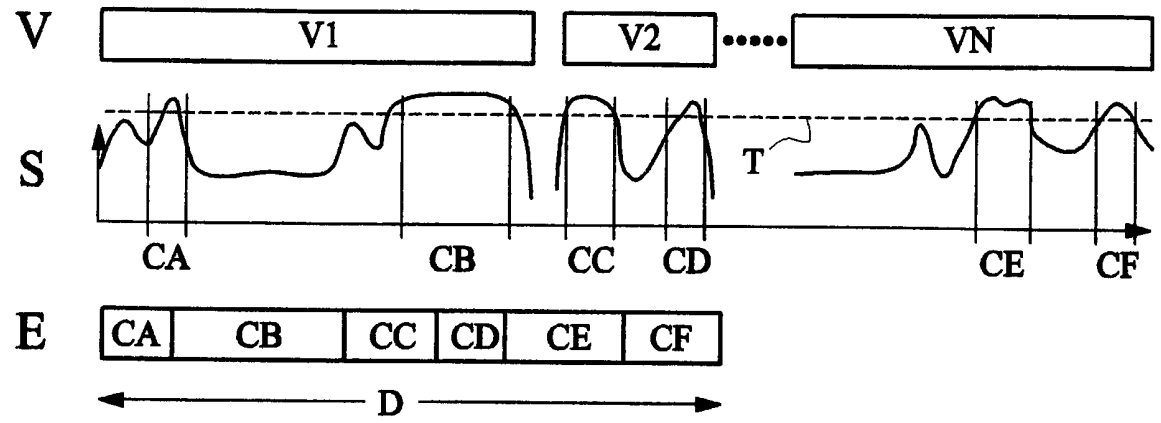

FIG. 3

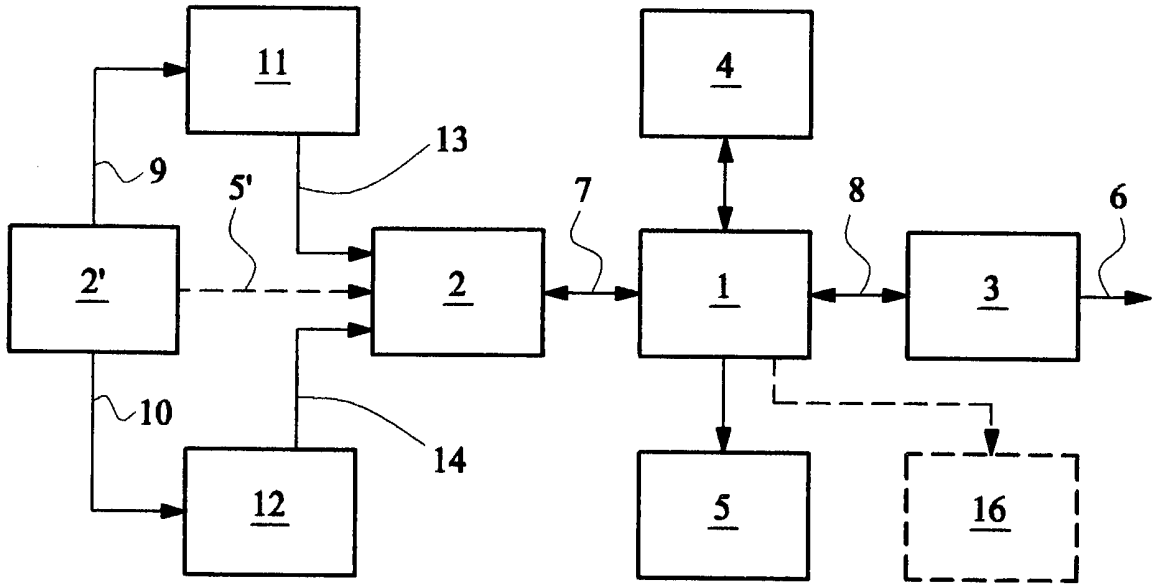

FIG. 4

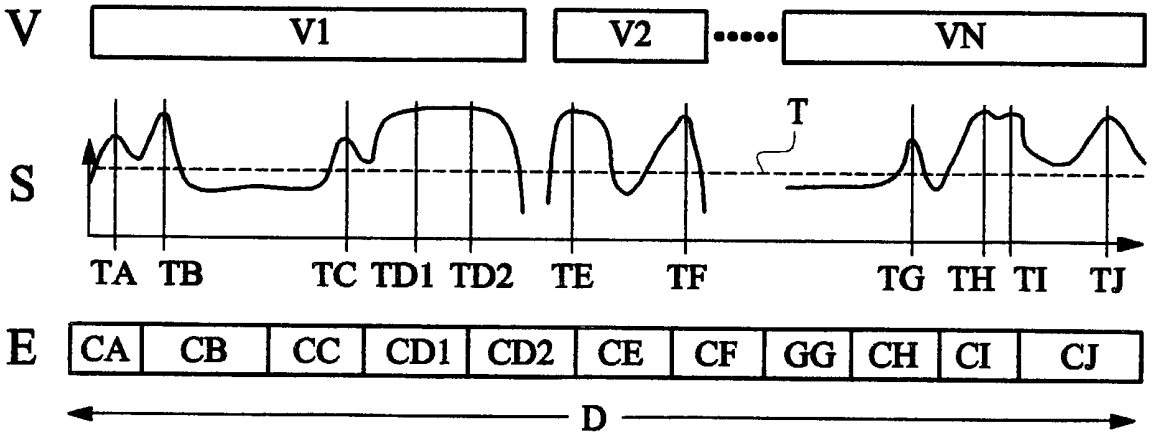

FIG. 5

Use of Saliency in Media Editing

Field of Invention

The present invention relates to the use of saliency in media editing, in particular aspects to the use of one or more saliency criteria as a basis for editing cumulative image data. Aspects of the present invention are particularly relevant to the creation of video programmes.

Description of Prior Art

As electronic still and video cameras are progressively developed, they are becoming smaller and easier to use, with improving imaging capabilities. At the same time, the media on which the camera signals are stored are also becoming smaller and cheaper, and with increasing battery lifetimes it is now very easy to capture a very large amount of audio and/or visual data over a relatively short period, possibly a single session or trip, before any downloading of data is necessary. A potential outcome is the use of wearable camcorders which are continuously recording.

Not all of the recorded material will have the same degree of interest to the user, particularly when recording continuously, and commonly it will require editing to retain selections of the material, and possibly to re-order the selected material or edit it in other ways, such as by control of video reproduction speed, selection of key frames therefrom, or the duration of a still shot. However, the amount of time that a user will want to spend on editing the recorded material is not expected to increase in proportion to the storage capability of the camera, and is more likely to remain essentially constant.

In the past, despite the time and effort involved, such a problem has been accommodated by manual editing of the captured material to produce photo albums or edited home videos. During the editing process it is necessary to bear in mind the purpose for which the edited material is being produced, and different sets of edited material may be required for different purposes. Thus would render the editing

process even more difficult and time consuming, and in practice, multiple edits from the same source material is done rarely, if ever.

Alternatively the problem has been avoided by judicious recording, as would have been the case when recording capacity was relatively limited, as in early electronic still cameras, or relatively expensive, as in photographic film cameras. Nevertheless, as will be appreciated, a choice in real time of what to record is often difficult, and particularly interesting or desirable "magic moments" are easily missed, which is why the idea of continuous recording for later editing is such a good idea in principle.

Therefore there is a need for an aid to the editing process to shorten the time and to reduce the effort required. Prior art aids may be described as: -

(a) Manual editing tools for providing one or more edits of the same source material. These tools include paper based photo albums, their electronic equivalents such as PictureIt (Microsoft), electronic slideshow tools such as ACDSee, and video editing tools such as Adobe Premier.

(b) Automated video summarisation or abstraction systems, on which much work has been done. In this context, "summarisation" generally refers to the generation of a set of key stills which represent the video and "abstraction" generally refers to the generation of a shorter video from parts of the source video. An example is the system provided by FXPAL, as described by A Girgensohn et al, "A Semi-Automatic Approach to Home Video Editing", UIST '00 Proceedings, ACM Press, pp 81-89, 2000. This uses a fully automatic heuristic measure of "unsuitability" to break up long video shots into shorter clips. There is also the possibility of breaking clips on the basis of the audio commentary by automatic identification of sentence boundaries. While the user can specify the overall duration of the edited video, the user must also specify manually which clips are to be used and the order in which they are to be viewed. The specified duration apparently controls the threshold of "unsuitability" used to determine in/out points for each clip.

Another exemplary system is that of Intel as described by R Lienhart in "Dynamic Video Summarisation of Home Video", Proc. of IS&T/SPIE, vol 3972, pp 378-389, January 2000, which groups shots in time based on the time stamp from a digital video camera. Using a technique in which the number of clips required by a fixed sampling rate is estimated, with in/out points being based on the audio content, long shots are sampled or subdivided to generate shorter clips. Again the user can specify the length of the edited video. Based on a hypothesis that all clips are equally important, the system is arranged to select clips in a "controlled random" manner. Depending on the ratio of the specified duration to the duration of the raw material, the system chooses a few "events" at random, and then picks a sequence of clips for each "event" at random.

(c) The use of professionally constructed interactive video material to control content and detail. US 6,278, 446 (Liou) describes a "System for Interactive Organisation and Browsing of Video" which assumes an unknown, professionally edited, video source. This is broken into shots which are then clustered into scenes or some other grouping, and in this instance an interactive method is used to correct an automated shot detection system and to organise the shots into a hierarchic arrangement which can be interactively viewed. The shot boundary detection system assumes that detecting explicit edit points in the source video is sufficient, which might or might not be true for material which has already been edited professionally, but is most unlikely to be the case for raw home video which typically will consist of very long shots which need to be broken up or reduced in some way. The clustering is designed to cater for situations which do not normally occur in home video, such as alternating shots between two camera views of the same event.

US 6,038,367 (Abecassis) "Playing a Video Responsive to a Comparison of Two Sets of Content Preferences" discloses an example of a system which selects the displayed content on the basis of user preference. It is arranged for processing professionally produced material where the producer has already

identified a profile consisting of one or more attributes for each segment of video material, and the viewer specifies a preference profile which is then matched against the profile of each video segment to determine whether or not that segment should be included in the version provided to the viewer. A typical use would be to allow a viewer to control the degree of sex and/or violence which they are shown from the source material.

Summary of Invention

The present invention relates to an editing system suitable for dealing with a raw camera image signal, in conjunction with an accompanying multi-value saliency signal. By "multi-value" is meant that the saliency signal has a least two non-zero values, whether a discrete valued signal or a continuously variable signal. "Saliency signal" here indicates a signal generated according to a criterion, or from a composite of two or more criteria, representative of one or more features determined to be of potential interest to one or more persons involved in any of production of the raw image signal, editing of the image signal or derived image signals to provide an edited result, and consumption of the edited result.

The multi-value saliency signal may be generated in a number of different ways, and it may express a single dimension of saliency, such as of visual interest to the user or viewer, or of interest in the recorded scene as expressed by the camera user on the occasion of the recording, or of audio interest. It may be generated automatically or manually by the user or viewer. With an electronic camera it may be recorded together with the camera image signal on the occasion that the camera is used, or it may be generated at another time, normally later, for example by a part of the apparatus according to the invention acting on a replayed video image signal or on an accompanying audio signal, or manually by a viewer of the recorded signal. Generation and recordal of a saliency signal at the time of camera signal recordal clearly can avoid the work involved in a later analysis of the recorded signal, whether this is automated or manual.

Alternatively, the multi-value saliency signal may be a multi-dimensional quantity derived by appropriately combining two or more single dimensional saliency values, e.g. by logical or algebraic combination, for example using the sum of the values, the greatest value, or their product. Before such combination, which may be applied to one-dimensional saliency signals derived at the same time and/or to one-dimensional signals derived at different times (e.g. one generated during recording and another generated during playback), appropriate scaling or other manipulation of at least one saliency value may be necessary.

One example of a suitable saliency signal is disclosed in our copending UK Patent Application published as GB-A-02380348. This describes a saliency signal is generated by analysis of the movement of a feature within the viewed scene, either at the time of recordal, when it may be recorded together with the camera image signal, or it may be generated on a subsequent occasion from the recorded signal. Our copending UK Patent Application No. 0225304.5 describes a saliency signal which may be provided directly by the user on the occasion of recordal using some form of manual or mechanical input to generate the signal according to the degree of interest or desire for the picture felt by the user during the occasion.

Such a saliency signal may generally be provided simultaneously with the camera image signal, but in certain instances the user may predetermine the saliency signal, for example where a particularly desirable image needs to be recorded.

The present invention is, in aspects, concerned with provision of an edited programme from raw source material, and embodiments of apparatus according to the invention include a user operable control for specifying a characteristic of the programme.

Exemplary programme characteristics that might be specified by the user are the duration of the programme, the pace of the programme, or the minimum saliency of the programme. Commonly, one of these characteristics will be set by the user as a primary parameter, but other secondary parameters could also be under user control as will become clear later. The control may provide preset choices or a continuous control.

For the purposes of illustration, most of the description below will relate to control of the duration of the programme. Nevertheless similar principles will be easily derived in respect of other programme characteristics, The reader will appreciate that the primary characteristics are not normally entirely independent of one another. For instance the user may determine a saliency threshold, thereby providing a variable programme length, and in that case there could be supplementary control of the average shot (selected portion) time, or of minimum/maximum shot times, or with a smart choice of video in/out points. Minimum and/or maximum shot times (or alternatively a mean shot time and allowable variations therefrom) will affect the pace of the programme and could alternatively be the parameters primarily controlled by the user.

In a first aspect, the present invention hence provides apparatus for providing an editing signal for editing an image signal from a camera recording to provide a programme, the image signal being accompanied by a multi-valued saliency signal, the apparatus comprising a user operable control for specifying the value of at least one characteristic of the programme as a whole and a signal processor responsive at least in part to the output of the user operable control and to the multi-valued saliency signal for generating a said editing signal indicative of selected portions of the image signal having higher saliency values, the selected portions being determined so that the value(s) of the programme characteristic(s) at least approximates to the specified value(s).

While the editing signal could be stored per se, for later use with the image signal, the apparatus preferably includes selection means responsive to the editing signal for selectively transmitting the selected portions of said image signal either for immediate display or transmission, or to a store forming part of the apparatus. This arrangement enables a viewer to adjust the edited programme at the time of viewing to suit their needs. For example, it is quite possible to arrange the apparatus to respond effectively to a requirement that the viewer needs to see the highlights of a particular source material in a specified time, e.g. two minutes, or that the viewer wants to be entertained with interesting (or salient) material from the same source material for a

specified time, e.g. 15 minutes, or even that the viewer wants to interact with the editing process (as described later) so that certain items of the source material are shown to a greater or lesser extent than would otherwise be the case based on the saliency level alone. The latter feature could be implemented by a viewer operated slider or rotary control (e.g. presented as a "go faster" or "slow down" control) which leads to dynamic adjustment of the overall duration and/or saliency threshold as the source material is being edited and viewed. It will be understood that in this way the same source material can be viewed in different ways to suit the needs of individual viewers by the use of high level viewer controls, as opposed to explicit manual editing as would be necessary with prior art arrangements, and that this is facilitated by the use of the multi-level saliency signal.

Indeed, it is possible to introduce a random or apparently random (or variable) element into the editing process, so that the same source material will be edited differently on successive viewings, and this feature could be particularly useful when the viewer merely wants to be entertained. Randomness may be obtained by introducing a random factor at any stage of the editing process, for example by adding a relatively low level time varying random factor (possibly plus 1, zero, or minus one) to the saliency signal before it is further processed, or by appropriately adjusting the output of a high level viewer control in a similar manner, or by introducing a decision step in the editing process which is randomly controlled. Variableness may be introduced in a similar manner, but with more control over the adjustment of the editing process.

For example, it may be that the number of saliency values is restricted to an extent that it does not entirely determine which images or portions are selected. Thus when selecting from a number of stored still images (e.g. 300 images) it may not be possible to view all of the images having the greatest saliency level (e.g. 50 images) because a minimum time for viewing individual images has been set (e.g. allowing only 25 images to be shown in the allocated time). In such a case a decision needs to be made as to which images are to be shown. Possible choices are to show the same 25 images on each viewing, or to make a random selection at each viewing (which will be more

interesting for the same viewer), or to deliberately vary the selection at each viewing so that over time all the selected images are viewed the same number of times which could mean making a record when each image is viewed. A variant of the latter process is to ensure that all 300 images are eventually viewed a number of times proportional to their saliency level, in which case the lower saliency level images are preferably fairly evenly distributed throughout the edited programme.

Thus in a further aspect the present invention also provides apparatus for editing an image signal from a camera recording to provide an edited signal providing a programme, the image signal being accompanied by a multi-valued saliency signal, the apparatus comprising a user operable control for specifying a characteristic of the overall edited programme and a signal selection circuit means responsive at least in part to the output of the user operable control and to the saliency signal for selecting portions of the image signal associated with higher values of the saliency signal, the selected portions being determined so that the value(s) of the programme characteristic(s) at least approximates to the specified value(s).

The aforesaid said multi-value saliency signal may be derived by methods such as those exemplified above.

Advantageously, where the recording consists of video material, the signal processor is arranged to maximise the saliency (e.g. the total or time integral of the multi-valued signal of all the selected portions) of the edited programme. Where the recording includes still images, the signal processor may also be arranged to maximise the saliency provided there are additional constraints on the lengths of time that any individual still image may be shown, to prevent a single highest saliency signal being shown continuously.

In one embodiment of apparatus according to the invention the signal processor is arranged to set the durations of the different selected portions to reflect the multi-valued saliency level thereof. This is particularly applicable to still images, but can also be applied to video clips.

Alternatively, when dealing with a video signal, the processor may be arranged to select those portions having a multi-valued saliency level above a threshold value T, and to set the value T so that the selected portions have a combined length substantially equal to the specified programme length.

In a somewhat related arrangement, the processor may be arranged to select those portions having a multi-valued saliency level above a threshold value T, to set the value T so that the selected portions have a combined length substantially equal to the specified programme length, and subsequently to adjust the length of any portions having a length greater than said maximum duration to said maximum duration and/or to adjust the length of any portions having a length less than said minimum duration to said minimum duration. This tends to assume that a video sequence may be cut at any arbitrary point, but it is well known that there are some points (for example where there is minimum motion in the image) at which cutting provides a better result. Accordingly in a yet further arrangement, the processor may be arranged to select those portions having a multi-valued saliency level above a threshold value T, and to set the value T so that the selected portions have a combined length substantially equal to the specified programme length, and subsequently to modify the length of at least one portion so as to adjust its in and/or out point to a more favourable time. In such a case, in one preferred embodiment the processor is arranged so that the timing error in programme length introduced by said adjustment is carried forward to the determination of the length of a subsequent selected portion or portions.

It may happen that one or more of the selected video clips is so long as to dominate the edited programme at that stage, or as to produce a boring programme. While this may be dealt with on the basis of settable or predetermined time constraints as outlined above, the processor may be arranged to respond to such a situation by subdividing it into a plurality of selected portions, for example on the basis of variations in the level of the multi-level saliency signal, and/or in response to features in an audio signal accompanying the recorded image signal.

In some embodiments, the apparatus is arranged so that the overall saliency (for example the saliency value integrated over the length of the edited programme) of the

edited signal is maximised. In other embodiments, in particular those for dealing with video image signals, the apparatus is arranged so that the overall saliency tends to a high value which however is slightly less than the maximum because it takes in other restricting but desirable factors such as the implementation of appropriate video in and/or out points based on other criteria (that is to say, the saliency value is maximised consistent with other constraints).

The employment of a saliency signal having a plurality of values, as opposed to a binary signal (which could be considered as equating effectively to turning a camera on and off or operating a camera release button), enables a flexible framework in which sensible and effective decisions to be made concerning the edited video, as will become clear particularly when considering the embodiments of the invention.

Preferably, for greater flexibility or freedom in editing, the apparatus according to the invention additionally includes a saliency adjuster for interactively adjusting the level of a said saliency signal while the edited signal is being provided. For example by employing the user operable control the user may be enabled to pause the editing process at a user selected stage, and to adjust the prevailing level of the multi-value saliency signal, either to include in the edited programme previously excluded portions of the recorded source signal, or to exclude previously included portions.

Preferably for greater control over the resulting edited programme the user operable control is arranged (e.g. in conjunction with the processor) to also permit the setting of at least one additional time constraint to affect the duration of said selected portions in the programme (thus affecting the pace of the edited programme), for example the setting of maximum and/or minimum durations for the selected portions (as more particularly described later), or the average duration.

Operation of this control may involve the shortening of an otherwise unduly long high saliency portion, for example either to provide one shorter extract or by cutting it into shorter spaced lengths which are then joined together. It may additionally or alternatively involve the rejection of short portions which would otherwise be selected for their higher saliency values. However another option is to lengthen such

portions to include temporally adjacent lower saliency parts until the minimum duration is reached, either for all such portions or at least for the ones among such portions having the relatively higher saliency values.

Apparatus according to the invention can be arranged to provide an edited programme based on any combination of still and or moving (video) images, and can therefore be arranged to receive signals from a stills camera or a video camera, or a hybrid of the two. In this context it is important to bear in mind that some cameras are capable of being operated in a number of different modes, including the provision of one or more of video sequences, single still images, and "burst" sequences consisting of a plurality of closely temporally spaced still images typically at between 1 and 5 frames per second and of higher resolution than frames of the video signal.

It is a consideration that under the conditions of home video, or the conditions under which the apparatus of the invention might be used, the quality of video sequences can be significantly impaired for example by excessive camera movement, so that in the case a jerky video clip for example it might be more acceptable in the edited programme to replace part or all of the clip with a good single still frame derived from the clip, or a single frame from a "burst" sequence if available (preferably a frame with the highest salience), or even part or the whole of a burst sequence if available.

Where the apparatus provides a programme consisting of a sequence of selected portions representing still images (from a still camera or from a video camera as just discussed), it may be arranged to set the durations of the different selected portions (commonly corresponding to single frames) to reflect the saliency thereof, normally so that the more salient stills are available for longer viewing in the edited programme. This is all subject to the constraint of specified programme length, so that the more time that is allocated to very salient frames, the less time is available for other frames, some of which may need to be discarded as a result, for example.

Where the apparatus provides a programme consisting of a sequence of selected portions which are video clips, it may be arranged to set the duration of each selected

portion to reflect an associated saliency value, for example its integrated or peak saliency value, so that more of the more interesting clips are viewed (as mentioned above, there may be other constraints such as in/out points to take into account). The associated saliency value could additionally or alternatively be employed in other ways – for example, a very salient clip could be repeated, or a clip with the highest salience could be repeated in slow motion, or a selected clip with the lowest saliency value could be played faster than normal. Again all these variations need to be put into the framework of the specified programme length.

When the apparatus provides a programme consisting of a sequence of selected portions which includes both video clips and still images, then more salient stills may be set to have longer durations, and more salient clips may be viewed for longer periods, as in the two preceding paragraphs. However, since the methods of allocating time are different for the two cases, it may be necessary to provide a rule for determining the absolute timings. For example, the stills may be treated for this purpose as video clips having their associated saliency values, so that their absolute durations within the edited programme are determined in the same way as the video clip periods. In an alternative option, the total number $\underline{s}$ of stills and the total number $\underline{v}$ of clips is initially assessed, and the edited programme length $\underline{l}$ is divided so that the stills occupy a total time of $l.s/(s+v)$ and the clips occupy the remainder of the programme time, with each of these time allocations then being divided according to relative saliency levels for example as in the two preceding paragraphs. Clearly there would be other ways of dividing the time between video clips and stills, depending on the desired end result.

In a still further aspect, the invention further provides a method of creating an editing signal for editing an image signal from a camera recording to provide an edited programme, the image signal being accompanied by a multi-valued saliency signal, the method comprising the steps of specifying a characteristic of the programme as a whole and generating a said editing signal indicative of selected portions of the image signal having higher saliency values in response at least in part to the specified characteristic and to the multi-valued saliency signal, with the selected portions being

determined so that the value of the programme characteristic at least approximates to the specified value. The editing signal can be recorded together with the image signal, i.e. the source material, for later use (in which case it is arranged to be temporally associated with the image signal, such as by incorporating the same time stamps, or by defining the times at which the image signal needs to be cut), or it is used as it is produced. When producing an edited programme the editing signal is supplied to a selection means, such as a switch receiving the image signal, for selectively transmitting the selected portions.

In a yet further aspect, the invention further provides a method method of editing an image signal, the image signal comprising a sequence of images with one or more associated multi-valued signals representative of saliency of content associated with the corresponding image, the method comprising determining a length of programme, identifying portions of the sequence of images with high values of at least one of the multi-valued saliency signals, and including said portions in an edited programme of the determined length.

Brief Description of Drawings

Additional features and advantages of the invention will become clear upon a consideration of the appended claims, to which the reader is referred, and also upon a reading of the following more detailed description of exemplary embodiments of the invention made with reference to the accompanying drawings, in which:

Figure 1 is a schematic block circuit diagram of a first embodiment of apparatus according to the invention;

Figure 2 is a diagram explaining the operation of the circuit of Figure 1 when used for recorded still photographs with accompanying recorded saliency signals;

Figure 3 is a diagram explaining the operation of the circuit of Figure 1 when used for recorded video signals with accompanying recorded saliency signals;

Figure 4 is a schematic block circuit diagram of a second embodiment of apparatus according to the invention, for use with recorded video and sound signals, but not necessarily recorded saliency signals; and

Figure 5 is a diagram explaining the operation of the circuit of Figure 4.

Where appropriate the same reference sign is used for closely corresponding items in the different figures.

Description of Specific Embodiments

The apparatus shown in Figure 1 is for use with a recorded camera source signal which is accompanied by a corresponding saliency signal having at least two non-zero values. It includes a central processor circuit 1 operatively coupled to first and second memories 2 and 3, and receiving outputs provided by a user control 4. Processor 1, which is preferably software driven, but may be implemented in hardware, also serves to transmit image and/or other signals to a display 5 as required by user control 1. The memory 2 is either arranged to receive signals downloaded from a camera memory, in which case memories 2 and 3 may be part of the same memory, or it may be a camera memory, e.g. a removable card or even part of the camera itself. Its output 7 to the processor 1 provides both the recorded image and recorded saliency signals. The memory 3 serves to receive and store the edited programme signals 8 from the processor 1, and provides a programme output 6 for later use or downloading as required under control of the processor 1 and user control 4.

Either the user control is suitably arranged so that the user can specify the type of image signal (e.g. still or video) in the memory 2, or the processor is arranged to identify the type of material automatically, e.g. from the signal format. The user control is also arranged so that the user can (a) specify the length of the edited programme; (b) instruct the apparatus to provide the source or edited signal to the display 5 for viewing; and (c) interactively alter the saliency signal recorded in the

memory 2. Other functions may be available, as will become clear when operation of the apparatus is described below.

The user control can take any known form, for example individual controls for each function, or a programmed type of control function where the user is led through the options on the display 5 and makes choices of parameters which are then confirmed on the display. The chosen parameters may remain at a convenient position on the display when the image signals are viewed, for example along the foot of the image.

As indicated in Figure 2, when used for still photographs the source signal provides a series of still images Pn (shown as P1 to PN), each accompanied by a saliency signal Sn (shown as S1 to SN). As shown, the signal Sn has 7 non-zero values, so is easily represented by a three digit binary number, and it is used both for selecting which of the images Pn are to be included, and for deciding how long each individual image will be viewed. The latter aspect is important for avoiding a monotonous image display rate and for enabling a longer look at the more interesting images.

The user operates the control 4 to specify to processor 1 the overall duration D of the programme arising from the edited signal, and the maximum (d_{max}) and minimum (d_{min}) shot durations for the images to be selected. Recognition of the signal type, e.g. still image, video or "burst", is commonly effected automatically, but otherwise the control 4 is also operated to specify that the image signals are for still photographs. Upon completion of input from the user control, detected automatically or by a further user input, the processor proceeds to compute the number of images (N_d) which can be displayed within the duration by dividing duration D by the mean shot length:

$$N_d = 2D/(d_{max} + d_{min})$$

The level of an inclusion saliency threshold T can then be adjusted until N_d images lie above it, and these images will have a saliency lying between a maximum value S_{max} and a minimum value S_{min}. Then for any image Pn in the selected set of N_d images an initial estimate of shot duration d^n_{init} can be made from the saliency value S^n for that image. An exemplary manner of doing this is by way of linear interpolation:

$$d^n_{init} = d_{min} + (d_{max} - d_{min})*(S^n - S_{min})/(S_{max} - S_{min})$$

At this stage the duration of the programme will be the sum V of the values d^n_{init} for all N_d of the selected images, and since the distribution of saliency values has not been taken into account, V may differ appreciably from D by an amount $\Delta = V - D$. Possible ways of adjusting this, if necessary, are (a) to distribute the amount Δ between some or all of the images at random; (b) to change the duration of each of the images by Δ / N_d; or (c) to distribute the amount Δ systematically over the selected images according to the saliency values.

The resulting edited programme E is indicated in the bottom line of Figure 2, where it can be seen that images P1, P2, P6, P8 and PN have been selected and that their durations increase with increasing height of Sn above the threshold value T. This programme E is appropriately stored in the memory 3, for example as the selected still images associated with their respective durations in the edited programme.

Figure 3 shows a plot similar to that of Figure 2 but for a source video sequence V composed of a number of source clips V1 to VN accompanied by a saliency signal S with continuously variable level. The user operates control 4 to specify that the signal in memory 2 is video. With a very simple approach, the processor 4 would be arranged to adjust the saliency threshold T until the total length of the selected portions CA to CF of the video equals D; it would then proceed to extract the selected portions from the memory 1, to join them using known techniques, and to store the edited programme E in the memory 3.

However, as shown, by operation of the user control 4 the maximum length of any clip is limited to d_{max} and/or the minimum length of any selected clip is made equal to d_{min}. These two aspects are respectively shown in the edited programme depicted in the bottom line of Figure 3 for clip B (where some parts lying above T are not used), and for clips A and D (where some parts lying below T are used).

While above video approaches do provide an edited programme with high saliency, they fail to take account of optimal in/out points for each clip, and they do not attempt to generate a variety of clip durations in the range between d_{max} and d_{min}.

In a more sophisticated approach, the processor 1 is arranged to operate so that the inclusion threshold T is used merely to select where to take clips from the source material, and does not set the in/out points on the basis of the intersections of T and the saliency signal, or directly on maximum and minimum durations of extracted video portions.

Figure 4 shows the block circuit diagram of an embodiment of the invention employing this approach. It is useful for processing recorded signal from a video camera but it differs from the apparatus of Figure 1 in that it can be used with recorded video signals lacking a recorded saliency signal S. As shown a memory 2' of or from the camera, or of the apparatus, contains recorded video image signals 9 and recorded audio signals 10, which in initial use of the apparatus are passed under the control of processor 1 respectively through an image saliency circuit 11 and an audio saliency circuit 12. Image saliency circuit 11 analyses the video image to derive therefrom a measure of visual saliency, for example by the method similar to that described in GB-A-02380348, to provide a visual saliency output 13. Audio saliency circuit 12 analyses the audio signal 10 to derive therefrom by known techniques a measure of audio saliency to provide an audio saliency output 14. Signals 13 and 14 are then recorded in a memory 2 in association with the source signals 9 and 10. Optionally, if a further saliency signal S' has been recorded in memory 2', for example by the method described in our copending UK Patent Application No. 0225304.5, this is also stored in the memory 2 in association with the other signals.

For completeness, the obtaining of a saliency signal according to the approaches described in GB-A-02380348 and UK Patent Application No. 0225304.5 respectively will now be briefly described (for fuller details of implementation, the skilled person is directed to refer to the specifications directly). In GB-A-02380348, features in the field of view are identified as significant and then tracked from frame to frame. A measure of saliency is calculated from the proximity of a feature to the centre of the frame, preferably with a weighting being given to the feature by indications of user interest such as movement (for example of a user's head for a head-mounted camera)

to recentre the feature in the field of view or the number of frames for which the feature is close to the centre of the field of view. The result can be a number of signals for each feature identified as being salient (or potentially salient) – these will be analogue signals in the first instance, but may be converted to digital signals (either binary, by comparison with a threshold) and signals from separate features may also be combined to provide an overall saliency. In UK Patent Application No. 0225304.5, user generated saliency is discussed, with a user input being provided to give a multi-valued saliency signal. This user input may be positively determined by the user, or may be sensed from, for example, a physiological property of the user (such as heartrate or sweating).

As determined by the user operating the control 4, the processor 1 uses as the multi-value saliency signal S, Figure 5, either the signal S' or the signal 13 or a desired combination of the two. Optionally, the audio signal 14 may also be involved in the formation of the signal S. However, in the operation of this embodiment particularly described with respect to Figure 5 it serves only in defining certain time points as explained later, and the signal S equates to the signal 13.

As with the embodiment of Figure 3, the mean shot length is used by the processor to determine the number N_d of shots to be selected, and the threshold T is adjusted until that number of shots lie above the saliency threshold. For each selected shot, a time point Tn corresponding to its maximum saliency value is then determined, as indicated by the vertical lines TA to TJ in the saliency plot S of Figure 5.

Optionally, long shots thus defined are broken by processor 1 into shorter sections so as to provide more than one time point per shot. For example a time point may be defined at each well defined local maximum of the saliency level, as shown for points TA and TB derived from a single portion lying continuously above T; or, if the saliency function is generally flat over a significant period, as in clip D, time points TD1 and TD2 may be determined by some other feature such as audio or visual content, and in the particularly described operation the processor is arranged (or instructed from the user control) to base such decisions on the audio saliency signal

14 so that time points such as TD1 and TD2 are located for example where audio saliency reaches separate local maxima.

Once the shots have been selected, each has its in and out points set by the processor to define the corresponding video clip. There is more flexibility than if saliency alone is used to set the in/out points, and the in and/or out points may even lie below the threshold T if appropriate, similarly to clips CA and CD of Figure 3. In general there are a number of constraints to be satisfied:

1. The overall programme duration D. There may also be desired pace limits d_{max} and d_{min}, which set the maximum and minimum clip lengths.

2. The distribution of shot lengths – if possible long and short shot lengths should be interspersed for maximum visual variety.

3. The overall saliency of the programme, which desirably tends to a maximum value.

4. Natural constraints for each clip, for example, the start and end of the source material, and "unsuitability" metrics such as are used in Girgensohn mentioned previously.

Many techniques for selecting in and out points are possible. In this embodiment, the following process is used. The region about each time point is analysed by the processor 1 to determine the earliest possible position for the in point and the latest possible position for the out point, for example on the basis of "unsuitability" of earlier/later material, end of the raw shot in the source material, etc. Subsequently, potential in/out points are identified by the processor 1 in the region between the earliest in point and the latest out point. As is recognised in the art, good in/out points are those which contain no or little motion in the scene, or breaks in speech, for example. To ensure that the maximum saliency point lies between any in point and any out point, only in points which precede the time point, and out points which follow the time point are identified by the processor.

Next the area under the saliency curve between the earliest in point and the latest out is determined by the processor for each shot, followed by linear interpolation to calculate a target duration for a clip from that shot which lies between d_{max} and d_{min}.

Each shot is then processed in turn, selecting the pair of in and out points from the alternative identified potential points so as to arrive at a clip duration which is closest to the target duration. The error between the calculated and actual clip durations is used to modify the target shot length for the next clip from the following shot, or it is dispersed among a plurality of the following shots. Finally the processor acts to extract the appropriate clips from the memory 2, to join them in the desired sequence, and to direct the resulting signal to memory 3 for storage.

In this way, the duration and size of the saliency signal for each shot determines the length of the corresponding clip, and the pace of the video, in terms of the number and duration of clips, is controllable at a high level to meet the user's preferences. Also the detailed clip boundary conditions are decided with full regard to the content of the video to avoid unnatural cuts.

This more sophisticated type of approach tends to cause the threshold level T to be lowered more than in the simple approaches described above, and consequently the resulting programme should contain more variety, and potentially prevent the problem of all of the time D being allocated to a relatively small number of long clips with high salience.

In a variation of the above video process, in and out points for the selected clips having a saliency value above a second, higher threshold value T' may be set so as to select the closest longer clip length to the target length (even if there is a closer shorter length possible) and other selected clips are allocated a closest shorter length to the target length, so that slightly more of the higher saliency material appears in the edited programme.

In either embodiment, the user control may be set so that the edited programme may be reviewed on the display 5, and so that it may be paused or otherwise processed at any point as required. Optionally, the source material may also be viewed in a similar

manner, but preferably in such a case either an addition display 16 is provided, or the display 5 is adapted to show a split screen to provide for both images.

Where interactive control is desired, the processor may be arranged so that the saliency values recorded in memory 2 or 15 may be adjusted by the user on the user control 4 when reviewing either source material and/or edited programme, so as to modify what goes into the edited programme. For example, the user may be able to stop the edited programme at any time, scan the source material about the corresponding time, and alter the saliency of parts thereof so as to include or exclude them from the edited programme. It will then normally be necessary to re-run the entire editing process, or at least to revise the editing of the remaining part of the programme, to ensure that the programme length is as specified by the user.

Optionally, the alterations in the recorded saliency signal(s) are recorded while the original values are retained, and also optionally provision is made to record sets of adjusted saliency values for a plurality of users, each set being associated with a user identifier. In this way different users may obtain different edited programmes from the same source material.

Additionally, or alternatively, the user may be enabled to stop the review of the edited programme at any point, and adjust the set parameters, such as overall time or maximum/minimum durations, and to instruct the processor to re-edit the remaining part of the programme accordingly.

A further refinement takes account of the fact that while the multi-value saliency signal may be fairly consistently generated for each individual recording session, with variations reflecting the associated saliency parameters, there may be considerable variation in absolute values between different sessions, and this could result in contributions from one session unreasonably dominating the edited programme.

The apparatus of the present invention may therefore include a facility to pre-scan the saliency signal over the entire length of the material to be edited, so as to attempt to normalise the saliency signal for different recording sessions prior to editing. As an example, the saliency signal may be adjusted so that the means and range over the

whole of the material are generally consistent, e.g. by determining a local average of the signal, and its mean, over a moving time window, followed by shifting and scaling to ensure that the local average and range remain constant for the duration of the material to be edited.

A yet further refinement takes account of the fact that the relative lengths of parts of the source material is unlikely to be what is needed to be reflected in the edited programme. Thus there may be a long portion with relatively high saliency, but all concerning the same event, whereas a second equally interesting and salient event is represented by only a short portion of the source material. This would be expected to result in a similar imbalance in the edited result, and some form of correction is desirable.

It is normally possible to identify different events for example from an analysis of time stamps recorded with a camera signal so as to obtain an indication of the relative contribution each event makes to the source material. A corrective factor for each event may then be determined which is used for adjusting the contribution each event makes to the edited programme. As a simple illustration, if a first event contributes three times as much to the source material as a second event, and they both have the same overall degree of saliency (i.e. the portions selected for the edited programme on a simple saliency basis would also be in the three to one ratio), the saliency signal for the first event could be scaled down until the relative contributions of the first and second events in the edited programme become more equal (i.e. an automatic saliency adjustment prior to production of a final editing signal), or even more simply, a time limit could be imposed on the contribution from the first event.

The apparatus of the invention therefore can be arranged so that it is possible to assign an overall saliency figure to each identified event or sequence in the source material, the relative overall saliency figures for the events being used to adjust the contributions from the individual events in the edited programme.

Other modifications of the invention include the facility to alter the speed of playback of at least some of the selected video clips, either per se, or additional to the normal

playback of such clips. It may be useful for example when a longer programme duration is specified, to repeat high saliency selected clips containing significant motion at a lower speed; or if a shorter programme is specified, selected clips with little or no motion could be played back at a higher speed.

Another optional facility would be to utilise the visual or other saliency peaks to provide keyframes for indexing the full video content, but providing a fast viewable feature in their own right.

Although the invention has been described in terms of editing an image signal provided by an electronic still video or hybrid camera, it should be noted that the invention extends to circumstances where the image signal is derived in other ways, for example by the scanning of still or cine images to provide an electrical or other image signal.

CLAIMS

1. Apparatus for providing an editing signal for editing an image signal from a camera recording to provide a programme, the image signal being accompanied by a multi-valued saliency signal, the apparatus comprising a user operable control for specifying the value of at least one characteristic of the programme as a whole and a signal processor responsive at least in part to the output of the user operable control and to the multi-valued saliency signal for generating a said editing signal indicative of selected portions of the image signal having higher saliency values, the selected portions being determined so that the value of the programme characteristic(s) at least approximate(s) to the specified value(s).

2. Apparatus according to claim 1 wherein the apparatus includes an input for a saliency signal which has been generated and recorded on the occasion when the image signal was recorded.

3. Apparatus according to claim 1 or claim 2 and arranged for generation of a saliency signal by manual input from a viewer during a preview of the recording, and for recording thereof, prior to generation of said editing signal.

4. Apparatus according to any preceding claim wherein the apparatus includes at least one saliency signal generator for generating a corresponding saliency signal from the recorded image signal.

5. Apparatus according to any preceding claim wherein the signal selection circuit means derives the said multi-valued saliency signal by combining a plurality of separate saliency signals.

6. Apparatus according to any preceding claim and including a saliency adjuster for interactively adjusting the level of a said saliency signal while the edited signal is being provided.

7. Apparatus according to any preceding claim wherein the signal processor is arranged to set the durations of respective said selected portions to reflect the multi-valued saliency level thereof.

8. Apparatus according to any preceding claim wherein a said characteristic is the minimum saliency level of said selected portions.

9. Apparatus according to any preceding claim wherein a said characteristic is the duration of said programme, with the durations of the selected portions being determined so that the programme length at least approximates to the specified duration.

10. Apparatus according to any preceding claim wherein a said characteristic is the pace of the programme.

11. Apparatus according to claim 10 wherein the pace of the programme is determined either (a) by the maximum and/or minimum duration for each said selected portion; or (b) by selecting the mean duration and the allowable variation from the mean duration.

12. Apparatus according to any one of claims 8 to 11 for use with an image signal from a video camera recording wherein the signal processor is arranged to maximise the saliency of the edited programme.

13. Apparatus according to any one of claims 1 to 7 wherein a said characteristic is the duration of said programme, with the durations of the selected portions being determined so that the programme length at least approximates to the specified duration, the processor being arranged to select those portions having a multi-valued saliency level above a threshold value T, and to set the value T so that the selected portions have a combined length substantially equal to the specified duration.

14. Apparatus according to any one of claims 1 to 6 for use with an image signal from a video camera recording, wherein a said characteristic is the duration of said programme, with the durations of the selected portions being determined so that the programme length at least approximates to the specified duration, the processor being arranged to select those portions having a multi-valued saliency level above a threshold value T, any portions having a length greater than a specified maximum duration being reduced to selected portions with said maximum duration and/or any

portions having a length less than a specified minimum duration being lengthened to selected portions with said minimum duration, and to set the value T so that the selected portions have a combined length substantially equal to the specified duration.

15. Apparatus according to any preceding claim wherein the processor is arranged to modify at least one selected portion so as to adjust its in and/or out point to a more favourable time.

16. Apparatus according to claim 14 wherein the processor is arranged so that the timing error in programme length introduced by said adjustment is carried forward to the determination of the length of a subsequent selected portion or portions.

17. Apparatus according to any preceding claim, wherein the processor is arranged effect a subdivision of at least one said portion which would otherwise be selected into a plurality of spaced selected portions.

18. Apparatus according to claim 16 wherein the processor is arranged to effect said subdivision at least in part in response to variations in the level of said multi-valued saliency signal.

19. Apparatus according to claim 16 or claim 17 wherein the processor is arranged effect said subdivision at least in part in response to features in an audio signal accompanying the recorded image signal.

20. Apparatus according to any one of claims 1 to 8 wherein said characteristic is selected from the duration of the programme, the pace of the programme and the minimum saliency of the programme.

21. Apparatus according to any preceding claim and including selection means responsive to said editing signal for selectively transmitting said selected portions of said image signal.

22. Apparatus according to claim 21 and including a store for storing the output of said selection means.

23. Apparatus according to any preceding claim and having means to introduce a random factor or a controlled variable factor into the process of generating the editing signal.

24. Apparatus according to claim 23 wherein said factor is introduced into said multi-valued saliency signal before use and/or into a said characteristic.

25. A method of creating an editing signal for editing an image signal from a camera recording to provide an edited programme, the image signal being accompanied by a multi-valued saliency signal, the method comprising the steps of specifying the value of at least one characteristic of the programme as a whole and generating a said editing signal indicative of selected portions of the image signal having higher saliency values in response at least in part to the specified characteristic value(s) and to the multi-valued saliency signal, with the selected portions being determined so that the value(s) of the programme characteristic(s) at least approximate(s) to the specified value(s).

26. The method according to claim 25 wherein the multi-valued saliency signal is derived at least in part directly from said recording.

27. The method according to claim 25 or claim 26 wherein the multi-valued saliency signal is derived at least in part from said image signal from said recording.

28. The method according to any one of claims 25 to 27 wherein the multi-valued saliency signal is derived at least in part from the recording of a saliency signal derived from a manual input of a viewer previewing the camera recording.

29. The method according to any one of claims 26 to 28 wherein the specified value is selected from the duration of the programme, the pace of the programme and the minimum saliency of the programme.

30. The method according to any one of claims 25 to 29 and including the further step of specifying the maximum and/or minimum length of any selected portion.

31. The method according to any one of claims 25 to 30 and including the further step of manually or automatically adjusting said multi-value saliency signal for parts of said image signal prior to production of a final said editing signal.

32. The method according to any one of claims 25 to 31 and including the further step of introducing a variable factor into the process of generating a said editing signal.

33. The method according to claim 32 wherein the variable factor is a random factor or a controlled variable factor.

34. The method according to claim 32 or claim 33 wherein the factor is introduced into the multi-valued saliency signal before it is used and/or into a said characteristic.

35. A method of producing an edited programme comprising performing the method as defined in any one of claims 25 to 34 and the step of selectively transmitting said image signal under control of said editing signal.

36. The method of claim 35 and including the step of recording said selectively transmitted signal.

37. A method of editing an image signal, the image signal comprising a sequence of images with one or more associated multi-valued signals representative of saliency of content associated with the corresponding image, the method comprising determining a length of programme, identifying portions of the sequence of images with high values of at least one of the multi-valued saliency signals, and including said portions in an edited programme of the determined length.

38. Apparatus substantially as hereinbefore described with reference to the accompanying drawings.

39. A method substantially as hereinbefore described with reference to the accompanying drawings.

Application No: GB 0317306.9 **Examiner:** Richard Baines
Claims searched: 1 - 39 **Date of search:** 1 June 2004

Patents Act 1977 : Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance	
X	1 - 4, 6 - 10, 12 - 14, 20, 21, 25 - 29, 31, 35 & 36	EP 0,782,139 A1	(SUN) - abstract, figures, column 2 lines 50 - 52 and column 5 lines 26 - 54
X	1 & 25 at least	JP 2002 218,367	(SONY) - abstract and figure
A	-		JP 10,108,071

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC[W]:

H4F

Worldwide search of patent documents classified in the following areas of the IPC[7]:

H04N, G11B

The following online and other databases have been used in the preparation of this search report:

Keywords in EPODOC, WPI, JAPIO