



(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/EP2016/080220

(22) International Filing Date:
8 December 2016 (08.12.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
15307079.2 21 December 2015 (21.12.2015) EP

(71) Applicant: THOMSON LICENSING [FR/FR]; 1-5, rue
Jeanne d'Arc, 92130 Issy-les-Moulineaux (FR).

(72) Inventors: FRADET, Matthieu; c/o Technicolor R&D
France, 975 Avenue des Champs Blancs, CS 17616, 35576
Cesson-Sevigne (FR). MARQUANT, Gwenaëlle; c/o
Technicolor R&D France, 975 Avenue des Champs
Blancs, CS 17616, 35576 Cesson-Sevigne (FR). DE-
MARTY, Claire-Hélène; c/o Technicolor R&D France,
975 Avenue des Champs Blancs, CS 17616, 35576
Cesson-Sevigne (FR). HE, Chenzhi; c/o Institut National
des Sciences Appliquées, (INSA) de Rennes, 20 avenue
des Buttes de Coësmes, CS 70839, 35708 Rennes Cedex 7
(FR).

(74) Agents: HUCHET, Anne et al.; 1-5, rue Jeanne d'Arc,
92130 Issy-les-Moulineaux, FR (FR).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA,
MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG,
NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS,
RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY,
TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN,
ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a
patent (Rule 4.17(ii))

Published:

— with international search report (Art. 21(3))

(54) Title: METHOD, APPARATUS AND ARRANGEMENT FOR SUMMARIZING AND BROWSING VIDEO CONTENT

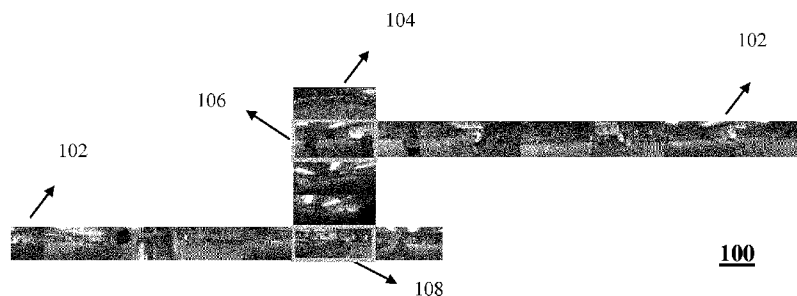


FIG. 1

(57) Abstract: A method, apparatus and arrangement for the summarizing and browsing of video sequences are provided and include at least one horizontal strip having time-sequenced video frames belonging to a single video sequence and at least one vertical strip having a plurality of video frames belonging to different video sequences, each of the plurality of video frames of the at least one vertical strip having at least one feature in common. In one instance, the at least one horizontal strip and the at least one vertical strip are arranged to intersect at a video frame of the at least one horizontal video strip having the least one feature in common with the video frames in the at least one vertical strip.



METHOD, APPARATUS AND ARRANGEMENT FOR SUMMARIZING AND BROWSING VIDEO CONTENT

5 TECHNICAL FIELD

The present principles relate generally to presenting video content and, more particularly, to summarizing and browsing video content.

BACKGROUND

10 Many works try to represent a whole video sequence into a single static or animated image. Given the desired dimensions of the output image, the idea is to incorporate as many “interesting image regions” as possible within these fixed dimensions. In such solutions, “interesting image regions” are usually key-frames. Key-frames should be mutually distinct so that near duplicates are not selected. Such
15 key-frames are usually selected either manually or by uniform subsampling in time, or automatically, for example, by using a shot and sub-shot detector. In the latter case, for each sub-shot the frame with the best quality and the maximum saliency is selected, or alternatively an objective function is minimized so that the selected key-frames (their number is supposed to be fixed) are optimal in that they are the ones
20 among all the input frames that would enable the reconstruction of the complete sequence with a minimum cost.

Once the representative frames or regions have been selected for the summary, the next problem is to arrange them such that the output representation remains compact and coherent and allows efficient browsing. Previous solutions that address
25 the browsing problem are really few.

SUMMARY OF THE INVENTION

Embodiments of the present principles are directed at least in part to addressing the deficiencies of the prior art by providing a method and apparatus and
30 arrangement for summarizing and intra and inter-video browsing video content. Various embodiments of the present principles provide a new compact representation of an input video set that enables efficient temporal browsing inside a single input video and also browsing from one video to another based on an inter-videos relationship.

In one embodiment of the present principles, an arrangement for the presentation of a plurality of video sequences for viewing includes at least one horizontal strip having time-sequenced video frames belonging to a single video sequence and at least one vertical strip having a plurality of video frames belonging to
5 different video sequences, each of the plurality of video frames of the at least one vertical strip having at least one feature in common. In such an embodiment the at least one horizontal strip and the at least one vertical strip are configured to intersect at a video frame of the at least one horizontal strip having the at least one feature in common with the video frames in the at least one vertical strip.

10 In an alternate embodiment of the present principles, a method for arranging video sequences for summarizing and browsing includes arranging video frames of a single video sequence in at least one strip having a first direction, the at least one strip arranged in the first direction having time-sequenced video frames, arranging video frames of different video sequences in at least one strip having a second direction, the
15 frames of the at least one strip arranged in the second direction having at least one feature in common and configuring the video frames of the at least one strip arranged in the second direction to intersect the at least one strip arranged in the first direction at a video frame of the at least one strip arranged in the first direction having the least one feature in common with the video frames of the at least one strip arranged in the
20 second direction.

In an alternate embodiment of the present principles an apparatus for arranging video sequences for summarizing and browsing includes a memory for storing at least control programs, instructions, software, video content, video sequences and data and a processor for executing the control programs and
25 instructions. In such an embodiment, when executing the control programs the processor configures the apparatus to arrange video frames of a single video sequence in at least one strip having a first direction, the at least one strip arranged in the first direction having time-sequenced video frames, arrange video frames of different video sequences in at least one strip having a second direction, the frames of the at
30 least one strip arranged in the second direction having at least one feature in common and configure the video frames of the at least one strip arranged in the second direction to intersect the at least one strip arranged in the first direction at a video frame of the at least one strip arranged in the first direction having the least one

feature in common with the video frames of the at least one strip arranged in the second direction.

In alternate embodiment of the present principles a machine-readable medium having one or more executable instructions stored thereon, which when executed by a digital processing system causes the digital processing system to perform a method for arranging video sequences for summarizing and browsing, the method includes arranging video frames of a single video sequence in at least one strip having a first direction, the at least one strip arranged in the first direction having time-sequenced video frames, arranging video frames of different video sequences in at least one strip having a second direction, the frames of the at least one strip arranged in the second direction having at least one feature in common and configuring the video frames of the at least one strip arranged in the second direction to intersect the at least one strip arranged in the first direction at a video frame of the at least one strip arranged in the first direction having the least one feature in common with the video frames of the at least one strip arranged in the second direction.

BRIEF DESCRIPTION OF THE DRAWINGS

The teachings of the present invention can be readily understood by considering the following detailed description in conjunction with the accompanying drawings, in which:

FIG. 1 depicts an exemplary representation of a collection of input video sequences arranged in crossed horizontal and vertical strips in accordance with an embodiment of the present principles;

FIG. 2 depicts an example of a temporal relationship between the two input videos of FIG. 1 in accordance with an embodiment of the present principles;

FIG. 3a depicts an example of the display of two horizontal strips and one vertical strip for summarizing and browsing video content in accordance with an embodiment of the present principles;

FIG. 3b depicts an example of the display of two vertical strips and one horizontal strip for summarizing and browsing video content in accordance with an alternate embodiment of the present principles;

FIG. 4a depicts an embodiment of the present principles in which a configuration for summarizing and browsing content includes a single horizontal strip and three vertical strips;

FIG. 4b depicts an embodiment of the present principles in which a configuration for summarizing and browsing content includes a single vertical strip and three horizontal strips;

FIG. 5 depicts an embodiment of the present principles in which a configuration for summarizing and browsing content includes a single vertical strip and two horizontal strips;

FIG. 6 depicts an example of a user interface in accordance with an embodiment of the present principles;

FIG. 7 depicts a flow diagram of a method for summarizing and browsing content in accordance with an embodiment of the present principles; and

FIG. 8 depicts a high level block diagram of an apparatus for implementing the features of the present principles in accordance with an embodiment of the present principles.

To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures. The drawings are not to scale, and one or more features may be expanded or reduced for clarity.

DETAILED DESCRIPTION

Embodiments of the present principles advantageously provide a method, an apparatus and an arrangement for summarizing and browsing video content. Although the present principles will be described primarily within the context of horizontal and vertical strips, the specific embodiments of the present principles should not be treated as limiting the scope of the invention. It will be appreciated by those skilled in the art and informed by the teachings of the present principles that the concepts of the present principles can be advantageously applied to video frames comprising strips oriented in substantially any direction.

The embodiments of the present principles leverage both width and height of a dedicated space of a presentation space/display screen. The interactive representation enables efficient intra and inter-video browsing while avoiding visual overload on a

presentation space/display screen contrary to multi-track representation of, for example, existing video editors.

More specifically, various embodiments of the present principles provide a compact representation of a collection of input video sequences based on crossed horizontal and vertical strips. In one embodiment, horizontal strips correspond to key-frames of filmstrips or video sequences and vertical strips are composed of key-frames from at least one different video having a common feature with a key-frame of the horizontal sequence(s). The vertical key-frames are connected together based on a common feature. For example, in one embodiment the key-frames can be connected because they all contain the same detected and recognized face, or they all correspond to the same scene captured roughly simultaneously but from different viewpoints.

Thus, in such embodiments a single horizontal strip summarizes a single input video while enabling efficient intra-video browsing, as provided by a continuous timeline on the x-axis and a single vertical strip enables efficient inter-video browsing by presenting, to a user, key-frames coming from different videos while depicting certain similarities.

Embodiments of the present principles leverage both width and height of a dedicated space of a presentation space such as a display screen. The interactive representation of the present principles enables efficient intra and inter-video browsing while avoiding visual overload.

FIG. 1 depicts an exemplary representation of an arrangement 100 of input video sequences arranged in crossed horizontal and vertical strips in accordance with an embodiment of the present principles. In the embodiment of FIG. 1, the horizontal strips 102 correspond to video sequences comprised of key-frames of a single video sequence. The vertical strip 104 of FIG. 1 is composed of key-frames from various different input videos. In accordance with various embodiments of the present principles, the key-frames for the horizontal and vertical strips are connected together/related based on some common features. For example, the horizontal and vertical strips can all contain the same detected and recognized face, or alternatively the key frames for the horizontal and vertical strips can all correspond to the same scene captured roughly at the same time but from different viewpoints, etc. In one embodiment of the present principles, common features are determined by processing programs that create exchange files containing the key frames and determine key-

frame connection information. Such files are then used to determine the common features to inter-relate the horizontal and vertical strips of the present principles.

In one embodiment of the present principles, inter-video key-frame connection is based on facial detection/recognition. The idea is to connect K_{j1}^{i1} with K_{j2}^{i2} if in both
 5 videos a common face was detected and both faces match.

In an alternate embodiment of the present principles, inter-video key-frame connection is based on image similarity. The idea is to connect K_{j1}^{i1} with K_{j2}^{i2} (defined below) using content-based image retrieval algorithms without assumption on objects that compose the scene. It should be noted that any kind of metadata could be used to
 10 establish the inter-video key-frame connections and so to deal with any video collection (e.g. same actor, same action (smoking, swimming, couple kissing...), same place, etc.).

Thus, in embodiments of the present principles, such as in the embodiment of FIG. 1, a single horizontal strip 102 enables efficient intra-video browsing of a single
 15 input video along a continuous timeline on, for example, the x-axis while a single vertical strip 104 enables efficient inter-video browsing by presenting to a user key-frames existing in different videos but linked by at least one common feature (as described above). That is, in various embodiments of the present principles an arrangement of the present principles includes at least one horizontal strip having
 20 time-sequenced video frames belonging to a single video sequence and at least one vertical strip having a plurality of video frames belonging to different video sequences, each of the plurality of video frames of the at least one vertical strip having at least one feature in common. In such embodiments, the at least one vertical strip is arranged to intersect at a video frame of the at least one horizontal video frame
 25 having the least one feature in common with the video frames in the at least one vertical strip (described in further detail below). For example, in FIG. 1, the vertical strip intersects one of the horizontal strips 102 at a first frame 106 and intersects a second of the horizontal strips 102 at a next to last frame 108.

In one embodiment in which N input videos constitute an input video set
 30 $\mathcal{V} = \{v^i\}_{i=1\dots N}$ and where key-frames are selected with the same uniform temporal subsampling for every input video, where strips are simple key-frames strips and where connections between key-frames from different input videos are based on temporal synchronization, for each input video $v^i = \{I_t^i\}$, given a temporal

subsampling step $s > 0$, M^i key-frames $\{K_j^i\}_{j=0 \dots M^i-1}$ are selected as $K_j^i = I_{j*s}^i$ where M^i directly depends on the frame number of v^i . That is, one key frame is taken for every 's' frames, so for any video, the j-th key-frame is the (j*s)-th video frame.

5 In one embodiment, for constructing the N horizontal strips, it is assumed that all input videos have the same frame width and frame height. If not, conversion to a common format is performed. For each i , given the key-frame set $\{K_j^i\}$, the horizontal strip image H^i is defined as a simple horizontal image stack, in one embodiment, in accordance with equation one (1) which follows:

10

$$H^i(x1,y1) = K_j^i(x2,y2) \text{ with } y1 = y2 \text{ and } x1 = x2 + j * \text{frame_width}, \quad (1)$$

$$\forall x2 \in [0, \text{frame_width}[, y2 \in [0, \text{frame_height}[, j \in [0, M^i - 1]$$

15

where $(x1,y1)$ and $(x2,y2)$ correspond to pixel locations in the domain of the horizontal strip image H^i (whose width may change for every i since M^i may change) and the image domain $[0, \text{frame_width}[\times [0, \text{frame_height}[$ respectively.

20

FIG. 2 depicts an example of a temporal relationship between two input videos in accordance with an embodiment of the present principles. In the embodiment of FIG. 2, each input video $\{v^i\}$ has its own time duration, as well as its own absolute start time and end time. For a given pair of videos (v^{i1}, v^{i2}) with respective absolute start times t_0^{i1} and t_0^{i2} and a temporal overlap, the time offset $\delta^{i1,i2}$ between the two input videos is related to the absolute start times according to equation two (2) which follows:

25

$$t_0^{i1} + \delta^{i1,i2} = t_0^{i2}. \quad (2)$$

30

To provide proper temporal synchronization between the two input videos, the time offset has to be determined. In various embodiments of the present principles, such synchronization information can be determined using metadata associated with the video files for the input videos if capture devices for the different videos were previously synchronized. Alternatively, such information can also be determined using audio or image feature matching between the two input videos, as discussed by Bagri et al., "A Scalable Framework for Joint Clustering and Synchronizing Multi-

Camera Videos", European Signal Processing Conference (EUSIPCO), 2013; and by Elhayek et al., "Feature-Based Multi-video Synchronization with Subframe Accuracy", DAGM 2012 (Deutsche Arbeitsgemeinschaft für Mustererkennung DAGM e.V. - German Association for Pattern Recognition).

5 In one embodiment of the present principles, a time threshold u is defined such that $0 < u < s/2$ and two key-frames K_{j1}^{i1} and K_{j2}^{i2} from different input videos v^{i1} and v^{i2} are considered. The two frames, K_{j1}^{i1} and K_{j2}^{i2} , are considered as "connected" if $|j1 * s + \delta^{i1,i2} - j2 * s| < u$. That is, the $j1$ -th key-frame of v^{i1} and the $j2$ -th key-frame of v^{i2} are captured at times separated by less than u . Note that
10 with $u < s/2$, one key-frame of a video v^{i1} cannot be connected with more than one key-frame of a video v^{i2} .

In accordance with an embodiment of the present principles, for the construction of vertical strips for each key-frame, K_j^i , the following set of key-frames $C_j^i = K_j^i \cup \{C_p^{i,j}\}_{p=1 \dots P_j^i}$ containing K_j^i itself and the P_j^i key-frames $C_p^{i,j}$ connected with
15 K_j^i is considered. To simplify the notations, $C_j^i = \{C_p^{i,j}\}_{p=0 \dots P_j^i}$ is considered. As such, the vertical strip image V_j^i is defined as a simple vertical image stack, in one embodiment, in accordance with equation three (3) which follows:

$$V_j^i(x1,y1) = C_p^{i,j}(x2,y2) \text{ with } x1 = x2 \text{ and } y1 = y2 + p * \text{frame_height}, \quad (3)$$

$$20 \quad \forall x2 \in [0, \text{frame_width}[, y2 \in [0, \text{frame_height}[, p \in [0, P_j^i]$$

where $(x1,y1)$ and $(x2,y2)$ correspond to pixel locations in the domain of the vertical strip image V_j^i (whose height may change for every (i,j) since P_j^i may change) and the image domain $[0, \text{frame_width}[\times [0, \text{frame_height}[$ respectively.

25 Equation (3) describes the construction of a vertical strip from bottom to top with a selected key-frame on the lowest row. Of course constructing similarly a vertical strip from top to bottom with selected key-frame on the highest row is straight forward, as well as constructing a vertical strip with selected key-frame on an intermediate row. This means that that there are several options for the vertical
30 arrangement of the key-frames, $C_j^i = \{C_p^{i,j}\}_{p=0 \dots P_j^i}$, that compose a vertical strip, among which the two following solutions can be found (1) by increasing video index

(key-frame of v^1 if any is displayed just above or just under selected key-frame) and (2) by decreasing similarity (the highest similarity to the selected key-frame, the closest to it). It should be noted that the option that would correspond to display key-frames of v^i on the i^{th} row in any case would lead to some holes in vertical strips that do not contain any key-frames of v^i .

In one embodiment of the present principles, the vertical arrangement is deduced automatically and in real-time from what is actually displayed on the screen to maintain on a same row the key-frames from a same video. That is, in accordance with an embodiment of the present principles, if an already displayed vertical strip contains a key-frame of video v^i on a certain row and that the user asks for the display of another vertical strip also containing a key-frame of video v^i , the arrangement is done so that both key-frames of video v^i appear on the same row. If it is not possible to do so without introducing holes in the required vertical strip, optimization is performed to satisfy the rule requiring maintaining vertical locations for the largest possible number of key-frames of an embodiment of the present principles. For example, in one embodiment of the present principles all configurations can be tested by considering a vertical segment of length of the number of key frames in the second vertical strip, sliding this segment along the y-axis and retaining the position that maximizes the number of rows of this segment containing one key frame of the first vertical strip corresponding to a video for which another key-frame has to be displayed in the second vertical strip

In accordance with embodiments of the present principles, when displaying simultaneously a vertical strip image and a horizontal strip image having a common key-frame, they are crossed at their key-frame in common. For a given pair (i, j) and for p such that $C_p^{i,j} = K_j^i$,

$$H^i(x + j * \text{frame_width}, y) = V_j^i(x, y + p * \text{frame_height}),$$

$$\forall x \in [0, \text{frame_width}[, y \in [0, \text{frame_height}[$$

That is, at an intersection, pixels of horizontal and vertical strip images have color values in common since they intersect at a common key-frame.

In one embodiment of the present principles, to avoid visual overload during display and collisions or inconsistencies between strips, simultaneous display of two

horizontal (or more) and two vertical strips (or more) is prevented. For example, assume that at the initialization, only one first horizontal strip is shown, then that a user requests for the display of a first vertical strip, then requests again for the display of a second horizontal strip. If the user requests again for the display of a second vertical strip during the display of the second horizontal strip, the display of the first horizontal strip will be removed/hidden before the display of the second horizontal strip. For example, FIG. 3a depicts an example of the display of two horizontal strips 102 and one vertical strip 104 for summarizing and browsing video content in accordance with an embodiment of the present principles. Similarly, FIG. 3b depicts an example of the display of two vertical strips (104) and one horizontal strip (102) for summarizing and browsing video content in accordance with an alternate embodiment of the present principles. As described above, in one embodiment of the present principles, to avoid visual overload during display and collisions or inconsistencies between strips, simultaneous display of two (or more) horizontal and two (or more) vertical strips is prevented. Although in the embodiments of FIG. 3a and FIG. 3b the simultaneous display of two (or more) horizontal and two (or more) vertical strips is prevented, in alternate embodiments of the present principles in which display/presentation space is not an issue and as such visual overload during display and collisions or inconsistencies between strips will not occur, the display of two (or more) horizontal and two (or more) vertical strips is enabled.

In various embodiments of the present principles, if a single strip exists in either the horizontal or vertical direction, the display of multiple strips in the other direction is possible. For example, FIG. 4a depicts an embodiment of the present principles in which a configuration for summarizing and browsing content includes a single horizontal strip 102 and three vertical strips 104. Similarly, FIG. 4b depicts an embodiment of the present principles in which a configuration for summarizing and browsing content includes a single vertical strip 104 and three horizontal strips 102.

In an embodiment in which key frames in a vertical strip are selected, there can exist frames in horizontal strips that have no feature in common. For example, FIG. 5 depicts an embodiment of the present principles in which a configuration for summarizing and browsing content includes a single vertical strip and two horizontal strips. In the embodiment of FIG. 5, after the key frames for the vertical strip are selected the two horizontal strips are created using key frames of two of the different

videos presented in the vertical strip. As depicted in FIG. 5, the horizontal strips are just above one another with no space in between. Such an arrangement can lead to a confusing presentation if the key frames of the two horizontal strips have no feature in common. As such, in accordance with an embodiment of the present principles, borders 502 (e.g., thick borders as depicted in FIG. 5) are displayed between the neighboring key frames of the horizontal strips that have no feature in common as depicted in FIG. 5.

In at least one embodiment of the present principles key-frame selection is based on saliency, activity and/or aesthetic estimation. Frames with a local maximum of saliency or aesthetic score or local minimum of activity are considered. Alternatively, key frame selection can be performed manually by a user.

However, when key-frame selection is not based on uniform temporal subsampling, inter-video key-frame connection, if still based on synchronization, must be adapted. That is, in such embodiments, one key-frame is connected to the closest key-frame of each other video if their temporal distance does not exceed a given threshold.

In at least one embodiment of the present principles, the key frames described above can comprise videos as well as still pictures. As such, vertical strip images can contain still pictures that have been connected to the considered video key-frame(s) of a horizontal strip and vice versa.

In addition, in accordance with various embodiments of the present principles, to enable a user to quickly recall frames of interest previously watched, video indices and key-frame indices or other such references to previously used key-frames are stored in a memory/queue. In such embodiments, corresponding thumbnails can be displayed in a dedicated display space. As such, when a thumbnail is selected, the associated horizontal and vertical strips are displayed.

In various embodiments of the present principles, a computer readable medium (e.g., memory, storage device, removable media, and so on) is provided with stored program instructions, which, when executed by a processor, will cause a method to be implemented, such as described above according to one or more embodiments of the present principles.

FIG. 6 depicts an example of a user interface 600 in accordance with an embodiment of the present principles. That is, an embodiment of the present

principles includes a user interface 600 including various commands for selecting video frames for browsing. The user interface 600 of FIG. 6 illustratively includes two top sections (602, 604) and a single lower section (606). In the embodiment of FIG. 6 the top left section (602) depicts bar graph representations of a number of input videos to a device, such as a video editor of the present principles. The number of bars corresponds to the number of videos, the length of a bar depicts the time duration of the corresponding video, and the horizontal alignment depicts the absolute start time of the corresponding video. The top left section (602) also provides a toggle button to enable a user to select at least one of the input videos depicted as horizontal bars and display the at least one associated horizontal strip in the lower section (606). In FIG. 6, the selected input video is depicted as the highlighted horizontal bar in the top left section (602). As also depicted in the embodiment of FIG. 6, the top left section (602) of the user interface 600 includes a vertical timeline bar which a user can use, by for example touching, clicking, sliding, to select a region of a horizontal strip, for example, the selected horizontal strip, to select a corresponding key-frame of the considered summarized video and display in the lower section (606) the associated vertical strip that shows other key-frames all coming from different videos but all connected to the selected key-frame based on certain established relationships as described above.

The top right section (604) of the user interface 600 of FIG. 6 illustratively depicts a representation of the selected key frame of the selected input video.

In the single lower section (606) of the user interface 600 of FIG. 6, a horizontal representation of the key frames of the selected horizontal video and vertical representations of the corresponding key frames of the other videos is depicted in accordance with an embodiment of the present principles.

In the lower section (606) of the user interface 600 of the embodiment of FIG. 6, a user can touch/click/slide horizontally one finger on any region of a vertical strip (not shown) to select the corresponding key-frame and display the horizontal strip that summarizes the video to which the selected key-frame belongs. In addition, a user can drag the horizontal strip left or right to show previous or future instances or frames if a strip is longer than a screen and similarly drag a vertical strip top or bottom to show other key-frames if a strip is higher than the screen. Similarly, a user can touch/click/slide vertically one finger on any region of a horizontal strip (not

shown) to select the corresponding key-frame and display the vertical strip that shows other key-frames all coming from different videos but all connected to the selected key-frame based on certain established relationships as described above.

The user interface 600 of the present principles further enables a user to
5 double touch or click on any region of a horizontal or vertical strip shown in the lower section (606) to play the corresponding video from the corresponding instant in the top right section (604), or in a separate window, or a second screen.

FIG. 7 depicts a flow diagram of a method 700 for summarizing and browsing content in accordance with an embodiment of the present principles. The method 700
10 begins at step 702 during which video frames of a single video sequence are arranged in at least one strip having a first direction (e.g., a horizontal direction), the strip comprising time-sequenced video frames. The method 700 can then proceed to step 704.

At step 704, video frames of different video sequences are arranged in at least
15 one strip having a second direction (e.g., a vertical direction), the frames of the at least one strip arranged in the second direction having at least one feature in common. The method 700 can then proceed to step 706.

At step 706, the video frames of the at least one strip arranged in the second direction are configured to intersect the at least one strip arranged in the first direction
20 at a video frame of the at least one strip arranged in the first direction having the least one feature in common with the video frames of the at least one strip arranged in the second direction. The method 700 can then be exited.

Optionally, in one embodiment of the present principles, only one strip is arranged in the first direction if there is more than one strip arranged in the second
25 direction, and only one strip is arranged in the second direction if there is more than one strip arranged in the first direction. An arrangement of the present principles can be limited as such to prevent visual confusion when displaying an arrangement of the present principles.

FIG. 8 depicts a high level block diagram of an apparatus 800 for
30 implementing the features of the present principles in accordance with an embodiment of the present principles. The apparatus 800 of FIG. 8 can comprise a device, such as a video editor, in various embodiments of the present principles. The apparatus 800 of FIG. 8 comprises a processor 810 as well as a memory 820 for storing control

programs, instructions, software, video content, video sequences, data and the like. The processor 810 cooperates with conventional support circuitry 830 such as power supplies, clock circuits, cache memory and the like as well as circuits that assist in executing the software routines stored in the memory 820. As such, it is contemplated

5 that some of the process steps discussed herein as software processes may be implemented within hardware, for example, as circuitry that cooperates with the processor 810 to perform various steps. The apparatus 800 of FIG. 8 also includes input-output circuitry 840 that forms an interface between the various respective functional elements communicating with the renderer.

10 Although the apparatus 800 of FIG. 8 is depicted as a general purpose computer that is programmed to perform various control functions in accordance with the present principles, the invention can be implemented in hardware, for example, as an application specified integrated circuit (ASIC). As such, the process steps described herein are intended to be broadly interpreted as being equivalently

15 performed by software, hardware, or a combination thereof.

While the foregoing is directed to various embodiments of the present principles, other embodiments of the invention may be devised without departing from the basic scope thereof. For example, one or more features described in the examples above can be modified, omitted and/or used in different combinations.

20 Thus, the appropriate scope of the invention is to be determined according to the claims that follow.

CLAIMS

1. A method (700) for intra and inter-video browsing, the method comprising:
 - Obtaining data representative of a collection of video sequences, said data
5 connecting key frames of said video sequences according to at least one feature;
 - Displaying a first arrangement of a subset of said key frames as crossed horizontal and vertical strips of key frames according to connections in said data;
 - 10 – Computing a second arrangement of said a subset of said key frames as crossed horizontal and vertical strips of key frames according to said first arrangement, the selection of a key frame of the first arrangement and connections in said data; and
 - Displaying the second arrangement.
- 15 2. The method of claim 1, wherein a strip of the arrangements gathers time-sequenced key frames of one video sequence of said collection of video sequences.
3. The method of claim 1, wherein a vertical strip of the arrangements gathers key frames of at least one video sequence of said collection of video sequences
20 connected according to at least one visual or audio feature.
4. The method of claim 1 or 2 wherein the selection of a key frame in a horizontal strip of a first arrangement having only one horizontal strip develops a vertical strip in the second arrangement.
5. The method of claim 1 or 2 wherein the selection of a key frame in a vertical
25 strip of a first arrangement having only one vertical strip develops a horizontal strip in the second arrangement.
6. The method of claim 1 or 2 wherein the selection of a key frame in a first
30 horizontal strip of a first arrangement having two or more horizontal strips removes horizontal strips but the first horizontal strip from the first arrangement and adds a vertical strip to the first arrangement to compute the second arrangement.

7. The method of claim 1 or 2 wherein the selection of a key frame in a first vertical strip of a first arrangement having two or more vertical strips removes vertical strips but the first vertical strip from the first arrangement and adds a horizontal strip to the first arrangement to compute the second arrangement.
- 5 8. The method of claim 1, wherein the at least one feature connecting key frames in said data comprises at least one of a common face, image similarity and a time of capture of video frames.
9. The method of claim 1, comprising converting the key frames to a common format such that all key frames have the same width and height.
- 10 10. The method of claim 1, comprising temporally synchronizing all video sequences.
11. The method of claim 10, wherein video sequences are temporally synchronized using at least one of metadata associated with the video sequences and at least one of audio and video image feature matching.
- 15 12. A apparatus (800) comprising:
 - a memory (820) for storing at least control programs, instructions, software, video content, video sequences and data; and
 - a processor (810) for executing the control programs and instructions, said processor when executing said control programs configuring said apparatus (800) to:
 - 20 • Obtain data representative of a collection of video sequences, said data connecting key frames of said video sequences according to at least one feature;
 - 25 • Display a first arrangement of a subset of said key frames as crossed horizontal and vertical strips of key frames according to connections in said data;
 - 30 • Compute a second arrangement of said a subset of said key frames as crossed horizontal and vertical strips of key frames according to said first arrangement, the selection of a key frame of the first arrangement and connections in said data; and
 - Display the second arrangement.

13. The apparatus of claim 11, wherein said apparatus (800) comprises a video editor.
14. A machine-readable medium having one or more executable instructions stored thereon, which when executed by a digital processing system causes the digital processing system to perform a method (700) for intra and inter-video browsing, the method comprising:
- 5
- Obtaining data representative of a collection of video sequences, said data connecting key frames of said video sequences according to at least one feature;
- 10
- Displaying a first arrangement of a subset of said key frames as crossed horizontal and vertical strips of key frames according to connections in said data;
- 15
- Computing a second arrangement of said a subset of said key frames as crossed horizontal and vertical strips of key frames according to said first arrangement, the selection of a key frame of the first arrangement and connections in said data; and
 - Displaying the second arrangement.

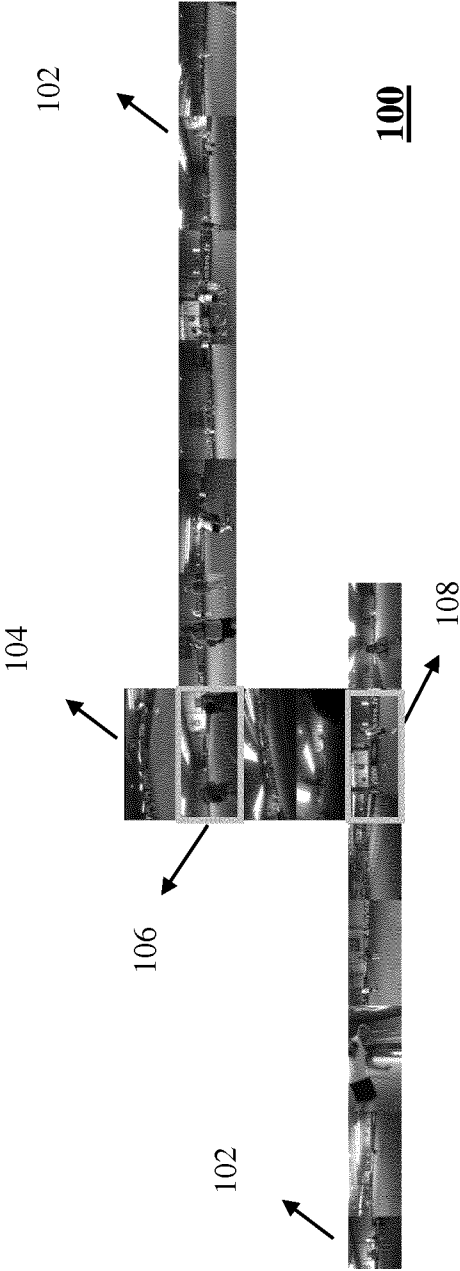


FIG. 1

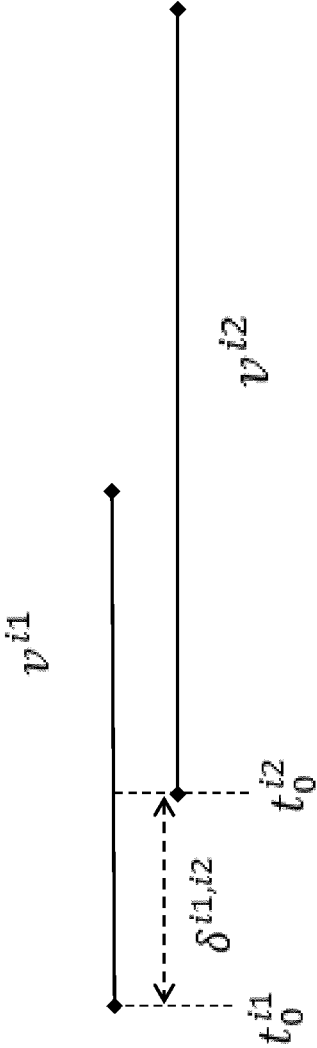


FIG. 2

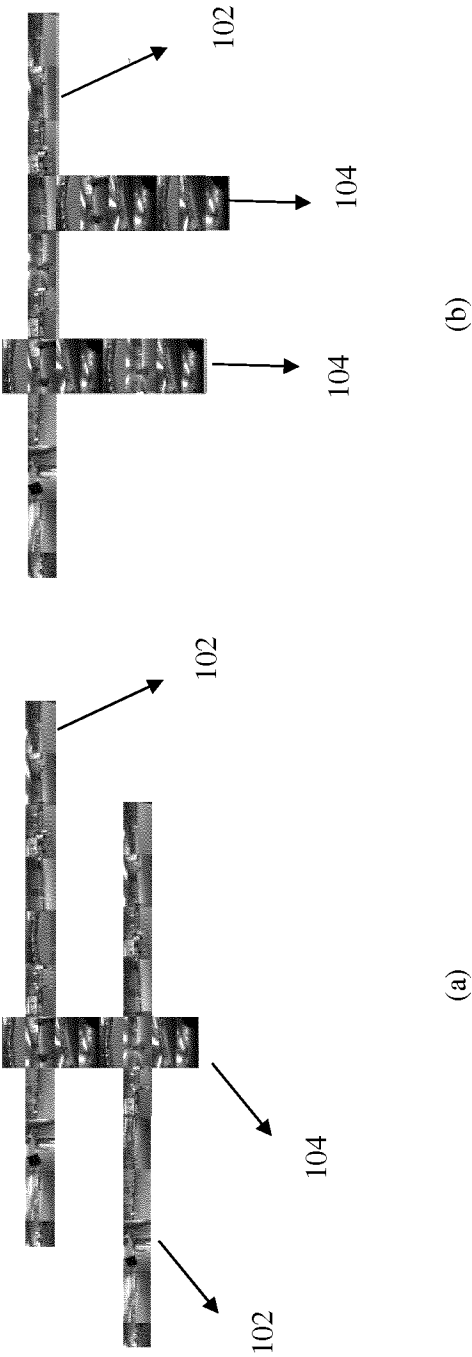


FIG. 3

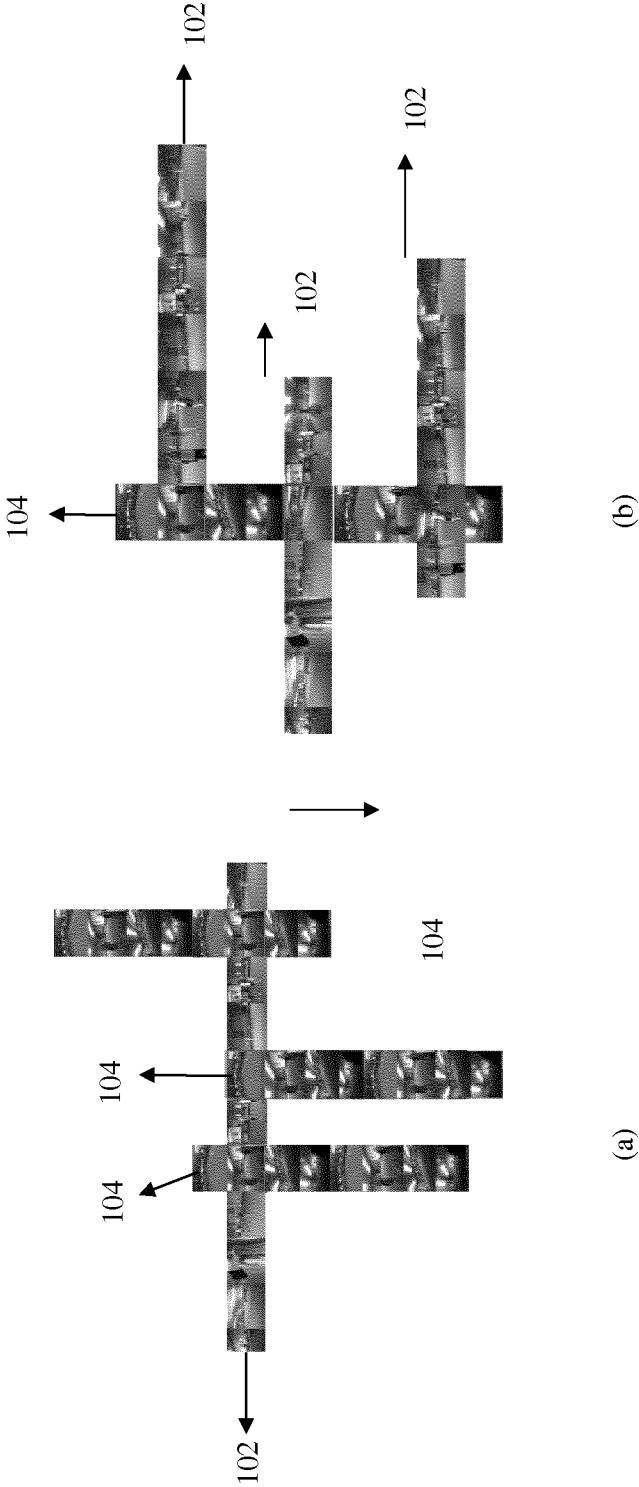


FIG. 4

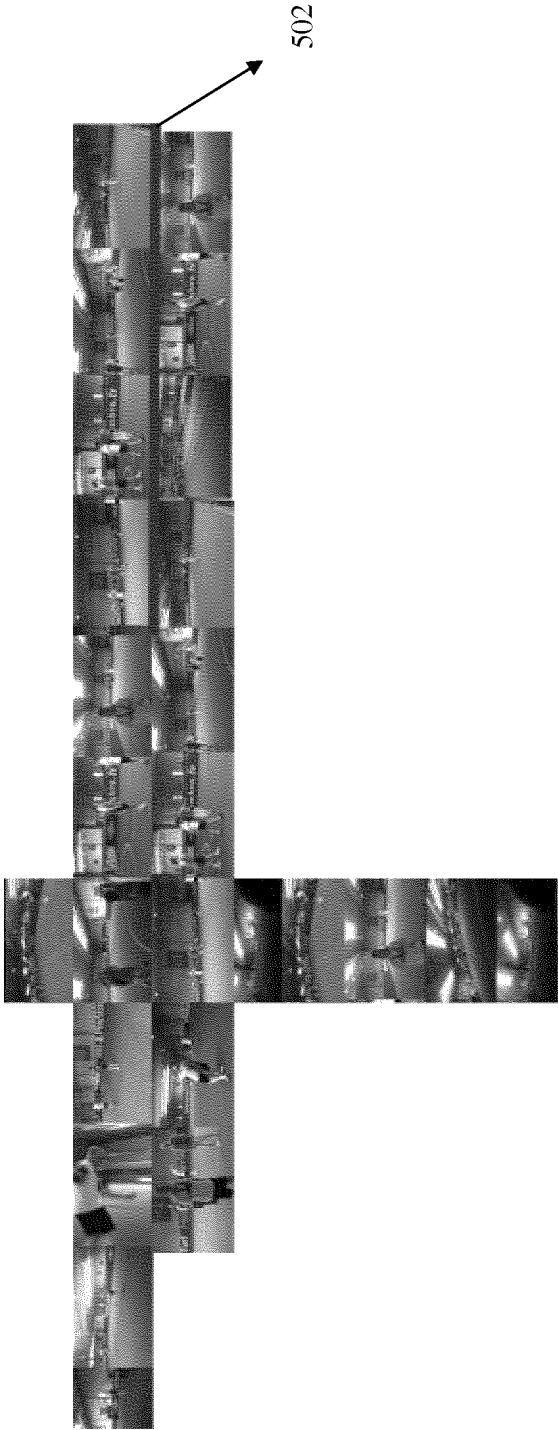


FIG. 5

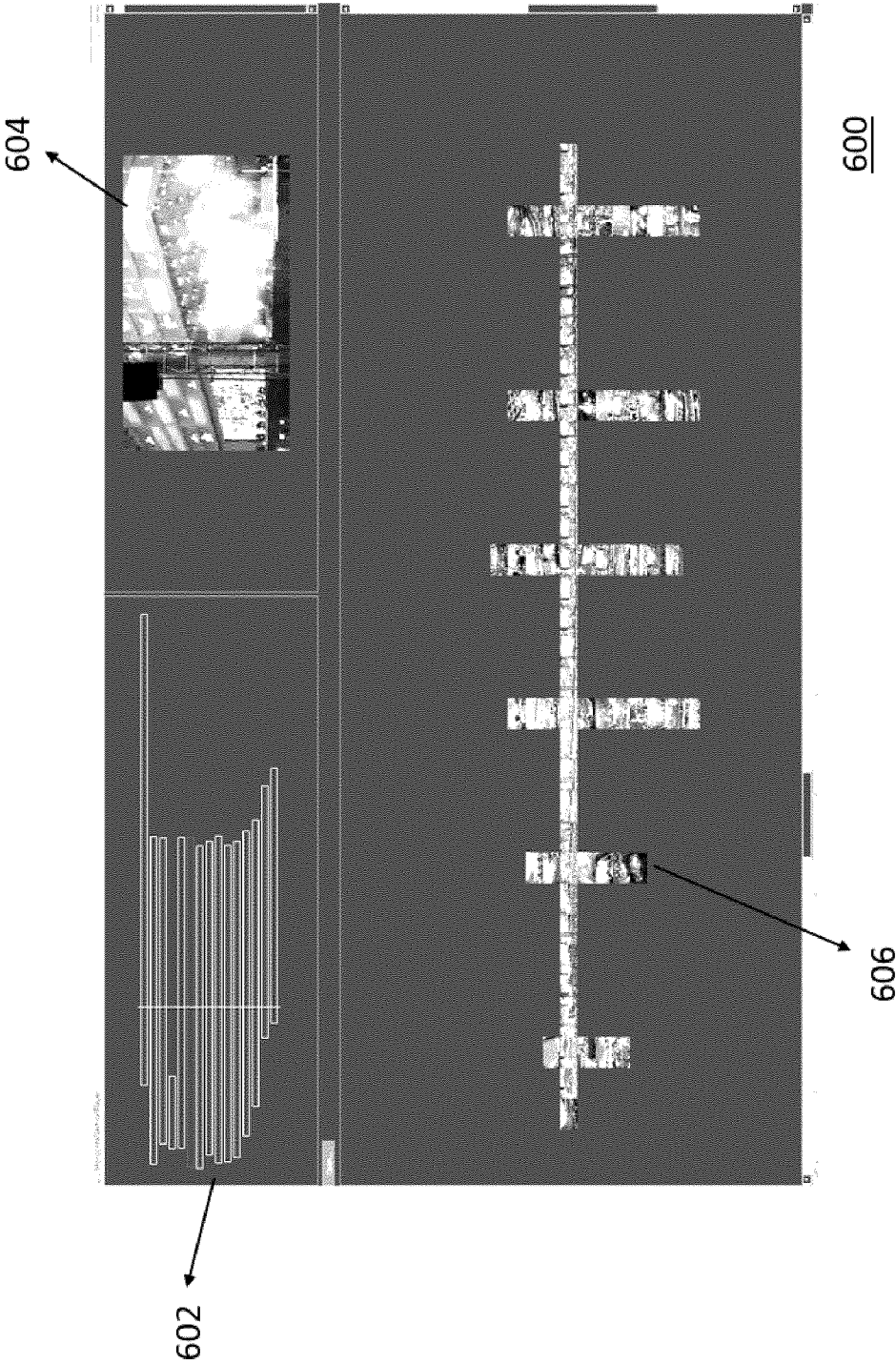


FIG. 6

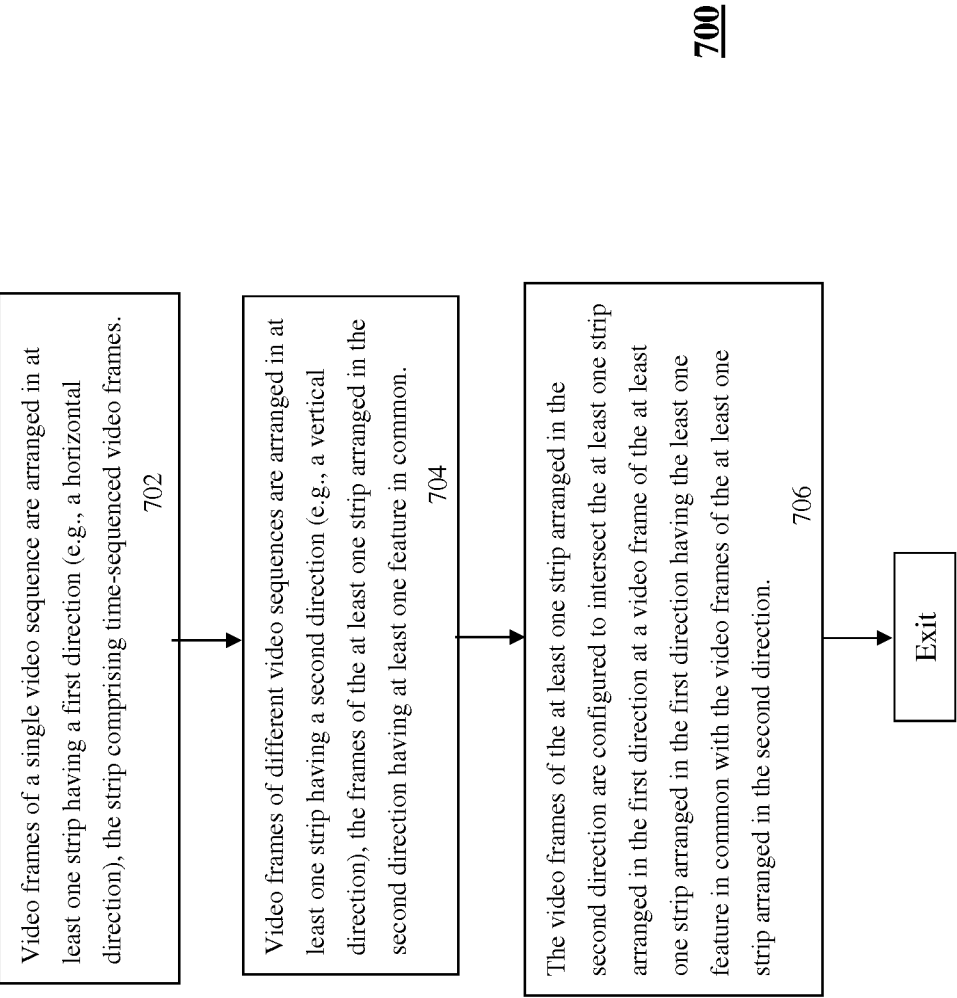


FIG. 7

800

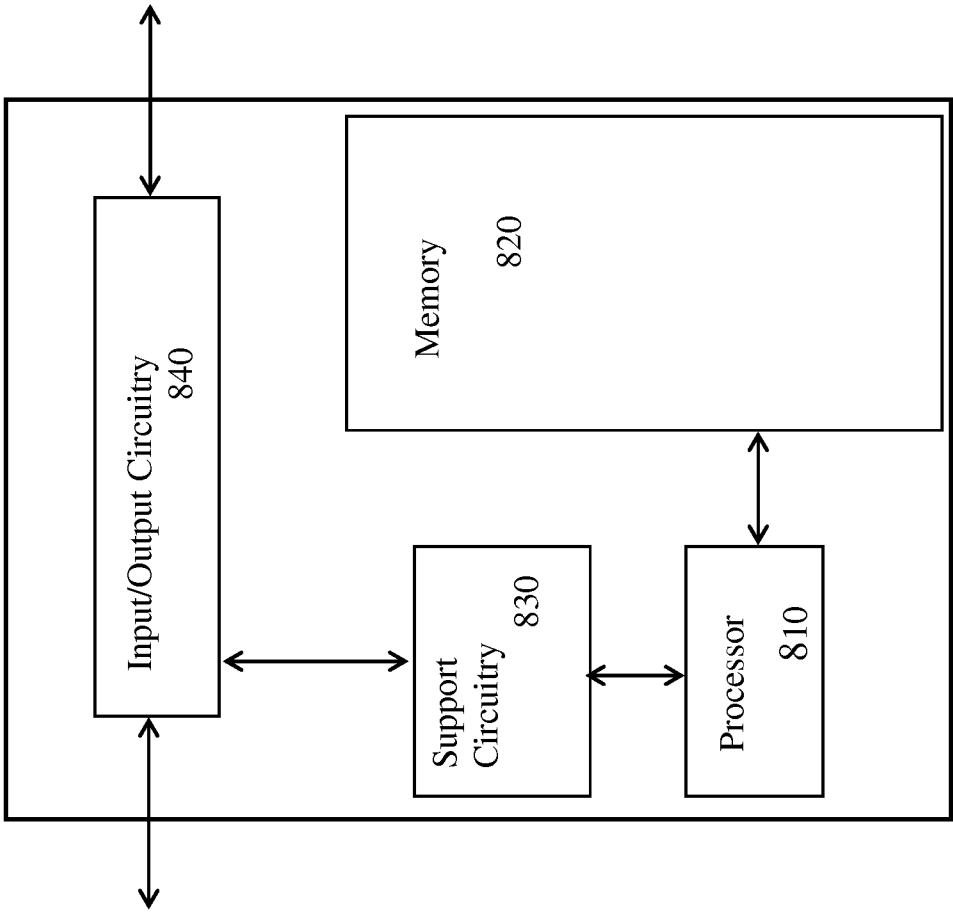


FIG. 8

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2016/080220

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F17/30
ADD.

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)
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 practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	C SNOEK ET AL: "The MediaMill TRECVID 2005 Semantic Video Search Engine", PROCEEDINGS OF THE 3RD TRECVID WORKSHOP, 30 November 2005 (2005-11-30), XP055348682, Gaithersburg, USA Section 5.4.3 and Figure 10 -----	1-14
A	EP 2 447 946 A2 (SONY CORP [JP]) 2 May 2012 (2012-05-02) paragraphs [0039] - [0071] -----	1-14
A	US 2012/293687 A1 (KARN KEITH STOLL [US] ET AL) 22 November 2012 (2012-11-22) paragraphs [0083] - [0086] -----	1-14
A	US 2002/056095 A1 (UEHARA YUSUKE [JP] ET AL) 9 May 2002 (2002-05-09) paragraphs [0011] - [0019] -----	1-14



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

22 February 2017

Date of mailing of the international search report

02/03/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Targon, Valerio

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2016/080220

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
EP 2447946	A2	02-05-2012	CN	102456379 A		16-05-2012
			EP	2447946 A2		02-05-2012
			JP	5678576 B2		04-03-2015
			JP	2012095118 A		17-05-2012
			US	2012110509 A1		03-05-2012

US 2012293687	A1	22-11-2012	BR 11	2013029493 A2		24-01-2017
			CN	103535023 A		22-01-2014
			EP	2710795 A1		26-03-2014
			JP	5877895 B2		08-03-2016
			JP	2014520424 A		21-08-2014
			KR	20140010989 A		27-01-2014
			US	2012293687 A1		22-11-2012
			US	2014105500 A1		17-04-2014
			WO	2012158588 A1		22-11-2012

US 2002056095	A1	09-05-2002	JP	4587416 B2		24-11-2010
			JP	2001309269 A		02-11-2001
			US	2002056095 A1		09-05-2002
