(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 November 2008 (27.11.2008)

PCT

(10) International Publication Number
WO 2008/143768 A1

(51) International Patent Classification:
G06F 17/30 (2006.01) *H04N 5/91* (2006.01)

(21) International Application Number:
PCT/US2008/005588

(22) International Filing Date: 1 May 2008 (01.05.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/930,905 17 May 2007 (17.05.2007) US

(71) Applicant *(for all designated States except US)*: **DOLBY LABORTORIES LICENSING CORPORATION** [US/US]; 100 Potrero Avenue, San Francisco, CA 94103-4813 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **RADHAKRISHNAN, Regunathan** [IN/US]; c/o DOLBY LABORATORIES LICENSING CORPORATION, 100 Potrero Avenue, San Francisco, CA 94103-4813 (US). **BAUER, Claus** [DE/US]; c/o DOLBY LABORATORIES LICENSING CORPORATION, 100 Potrero Avenue, San Francisco, CA 94103-4813 (US).

(74) Agents: **LATHROP, David N.** et al.; Gallagher & Lathrop, A Professional Corporation, 601 California Street, Suite 1111, San Francisco, CA 94108-2805 (US).

(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, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(54) **Title:** DERIVING VIDEO SIGNATURES THAT ARE INSENSITIVE TO PICTURE MODIFICATION AND FRAME-RATE CONVERSION

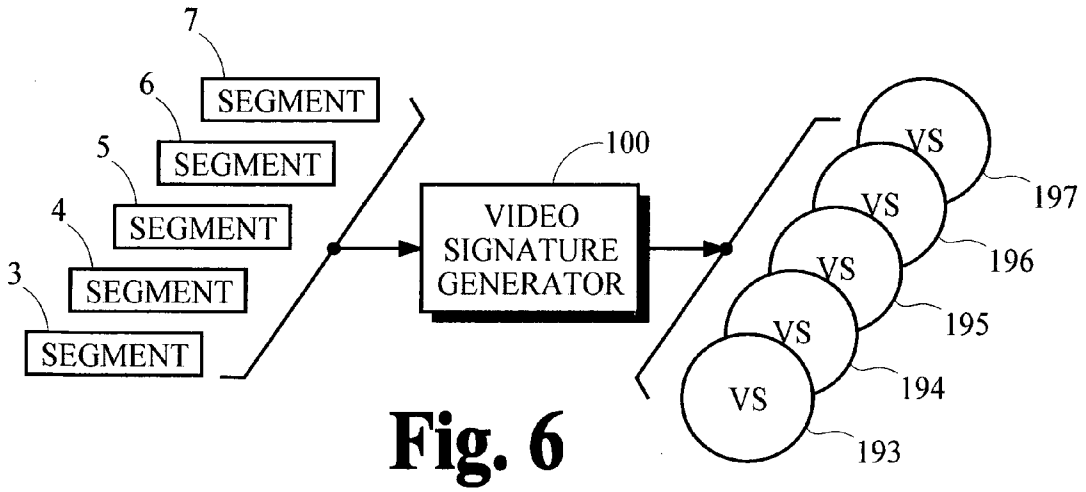


Fig. 6

(57) **Abstract:** A signature that can be used to identify video content in a series of video frames is generated by first calculating the average and variance of picture elements in a low-resolution composite image that represents a temporal and spatial composite of the video content in the series of frames. The signature is generated by applying a hash function to values derived from the average and variance composite representations. The video content of a signal can be represented by a set of signatures that are generated for multiple series of frames within the signal. A set of signatures can provide reliable identifications despite intentional and unintentional modifications to the content.

WO 2008/143768 A1

DESCRIPTION

Deriving Video Signatures That Are Insensitive to Picture Modification and Frame-Rate Conversion

TECHNICAL FIELD

The present invention pertains generally to the processing of video signals and pertains more specifically to processes that extract features from video signals to identify the signals. Throughout this disclosure, the terms "video signals" and "video content" refer to signals and content that represent images intended for visual perception.

BACKGROUND ART

Applications that attempt to detect authorized and unauthorized video content of a received signal often rely on processes that analyze the content of the received signal to generate some type of content identifier or signature. These applications use results from the analysis to determine whether the received content is a copy of some reference content. For many of these applications, it is important to obtain a reliable identification even when the content of the received signal has been modified unintentionally or intentionally so that it differs from the reference content but can still be recognized by a human observer as being substantially the same as the reference content. If the perceived difference between the reference content and the modified content is small, then preferably the signature-generation process should generate signatures from the reference and modified content that are very similar to one another.

Examples of unintentional modifications to signal content include the insertion or addition of noise to signals in transmission channels and on storage media. Examples of intentional modifications to video signals include luminance and color modifications such as contrast/brightness adjustments, gamma correction, luminance histogram equalization, color saturation adjustments and color correction for white balancing, include geometric modifications such as image cropping and resizing, image rotation and flipping, stretching, speck removal, blurring, sharpening and edge enhancement, and include coding techniques such as lossy compression and frame rate conversion.

DISCLOSURE OF INVENTION

It is an object of the present invention to provide identification processes that can be used to obtain a reliable identification of video content even if the content has been modified by mechanisms such as those mentioned above.

This object is achieved by the present invention that is described below.

The various features of the present invention and its preferred embodiments may be better understood by referring to the following discussion and the accompanying drawings in which like reference numerals refer to like elements in the several figures. The contents of the following discussion and the drawings are set forth as examples only and should not be understood to represent limitations upon the scope of the present invention.

BRIEF DESCRIPTION OF DRAWINGS

Figs. 1 and 2 are schematic block diagrams of a video signature generator that may be used to obtain a reliable identification of a video signal.

Fig. 3 is a schematic block diagram of a process performed in one implementation of an image pre-processor.

Fig. 4 is a schematic block diagram of a lower-resolution image obtained by a spatial-domain processor.

Fig. 5 is a schematic block diagram of video frames arranged in segments.

Fig. 6 is a schematic block diagram of a video signature generator processing segments of video content to generate a set of video signatures.

Fig. 7 is a schematic block diagram of a system that manages a signature data base for detection of copies of video content.

Fig. 8 is a schematic block diagram of a device that may be used to implement various aspects of the present invention.

MODES FOR CARRYING OUT THE INVENTION

A. Introduction

Various aspects of the present invention may be used advantageously in a system for identifying video content by analyzing segments of that content and generating a signature for each segment. The signatures generated for the segments in an interval of a signal form a signature set, which can be used as a reliable identification of the content in that interval. The following disclosure first describes processes that may be used to generate a signature for a single segment and then describes the generation and use of signature sets.

Fig. 1 is a schematic block diagram of a video signature generator 100 that analyzes the video content in a signal segment 3 to generate a video signature 193 that identifies or represents that content. In the example shown, the segment 3 includes a

series of video frames 3a to 3d. If the video signal conveys audio content as well as video content, an audio signature that represents the audio content may be obtained by processing the audio content in a variety of ways including those disclosed in U.S. provisional patent application no. 60/872,090 entitled "Extracting Features of Video and Audio Signal Content to Provide a Reliable Identification of the Signals" by Regunathan Radhakrishnan, et al., filed on November 30, 2006, which is incorporated herein by reference in its entirety.

B. Video Signature Extractor

One implementation of the video signal generator 100 is illustrated in Fig. 2. In this implementation, an image pre-processor 110 obtains a series of format-independent images for the pictures conveyed in the frames 3a, 3b, 3c, 3d, a spatial-domain processor 130 down-samples the format-independent images to generate a series of lower-resolution representations of the format-independent images, a temporal-domain processor 150 generates values that represent a composite of the series of lower-resolution representations, and a video signature processor 170 applies a hash function to the composite values to generate the video signature 193 that represents and identifies the content of the segment 3. The processing that is performed by the processors 110, 130, 150 and 170 may be implemented in a variety of ways. Preferred implementations of these processes are described below.

1. Image Pre-Processor

For one exemplary implementation, each video frame 3a, 3b, 3c, 3d in the segment 3 conveys a picture that is represented by an array of pixels D. The image pre-processor 110 derives a format-independent image of the picture for each frame. The format-independent image is represented by an array of pixels F. The derivation of the format-independent image may be done in a variety of ways. A few examples are described below.

In one application, the video signature generator 100 generates signatures for television video signals that convey video content in a variety of formats including progressive-scan and interlaced-scan with the standard-definition (SD) resolution of 480 x 640 pixels and the high-definition (HD) resolution of 1080 x 1920 pixels. The image pre-processor 110 converts the picture in each frame into a format-independent image that has a format common to all signal formats of interest. In preferred implementations, the pixels F in the format-independent images are obtained by down-sampling the pixels D in

the frame to reduce sensitivity to modifications that can occur when frames of video are converted between different formats.

In one example, the resolution of the format-independent image is chosen to have a resolution of 120 x 160 pixels, which is a convenient choice for television signals conveying images in HD and SD resolutions for both progressive-scan interlaced-scan formats. The image pre-processor 110 converts SD-format video content into format-independent images by down-sampling the pixels in each frame picture by a factor of four. The image pre-processor 110 converts HD-format video content into format-independent images by cropping each frame picture to remove 240 pixels from the left-hand edge and 240 pixels from right-hand edge to obtain an interim image with a resolution of 1080 x 1440 pixels and down-sampling the pixels in the interim image by a factor of nine.

If a video signal conveys content in an interlaced-scan format in which frames of video are arranged in two fields, the signal may be converted into a progressive-scan format before obtaining the format-independent image. Alternatively, greater independence from the choice of scan format can be achieved by obtaining the format-independent image from only one of the fields in an interlaced-scan frame. For example, the format-independent image can be obtained from only the first field in each frame or from only the second field in each frame. Video content in the other field can be ignored. This process avoids the need to convert to a progressive-scan format before obtaining the format-independent image.

If appropriate cropping and down sampling is used, the resultant image is essentially independent of the frame picture format so that the subsequent signature generation process is insensitive to different formats and to modifications that occur from conversions between formats. This approach increases the likelihood that a video signature generated from a series of format-independent images will correctly identify the video content in a series of frame pictures even if those pictures have been subjected to format conversion.

Preferably, the format-independent image excludes picture areas that are likely to be affected by intentional modifications. For video applications such as television, for example, this may be achieved by cropping to exclude corners and edges of the image where logos or other graphical objects may be inserted into the video content.

Fig. 3 provides a schematic illustration of the results obtained by a process 112 performed by the image pre-processor 110 that includes the cropping and down-sampling operations described above. The picture in the frame 3a within the segment 3 is cropped to extract the pixels D in a central portion of the picture. The pixels D in this central portion are down-sampled to obtain the pixels F in the format-independent image 5a. A format-independent image 5a, 5b, 5c, 5d in a series of images 5 is obtained for each frame 3a, 3b, 3c, 3d in the segment 3. The process 112 may be expressed as:

$$\{F_m\} = IP[\{D_m\}] \text{ for } 0 \le m < M \quad (1)$$

where $\{F_m\}$ = the set of pixels in the format-independent image for frame m;

$IP[\ ]$ = the image pre-processor operations applied to the picture in frame m;

$\{D_m\}$ = the set of pixels in the picture for frame m; and

M = the number of frames in the segment.

The cropping operation that resizes a picture for format conversion may be combined with or performed separately from the cropping operation that excludes areas of a picture that may be affected by intentional modification such as the insertion of logos. The cropping operations may be performed before or after the down-sampling operations. For example, the format-independent image may be obtained by cropping video content and subsequently down sampling the cropped images, it can be obtained by down sampling the video content and subsequently cropping the down-sampled images, and it can be obtained by a down-sampling operation performed between the two cropping operations mentioned above.

If each video frame conveys a color image comprising pixels represented by red, green and blue (RGB) values, for example, a separate format-independent image may be obtained for each of the red, green, and blue values in each frame. Preferably, one format-independent image is obtained for each frame from the luminance or brightness of pixels that is derived from the red, green, and blue values in the frame. If each video frame conveys a monochromatic image, the format-independent image may be obtained from the intensities of the individual pixels in that frame.

2. Spatial-Domain Processor

In an exemplary implementation, the spatial-domain processor 130 obtains a down-sampled lower-resolution representation of the format-independent images by grouping the pixels F in each of the format-independent images into regions that are GX

pixels wide and GY pixels high. A lower-resolution image with picture elements E is derived from the intensities of the pixels F in a respective format-independent image by calculating the average intensity of the pixels in each region. Each lower-resolution image has a resolution of K x L elements. This is illustrated schematically in Fig. 4. The picture elements E may be obtained by performing a process that implements the following expression:

$$E_m(k,l) = \frac{1}{GX \cdot GY} \sum_{i=k \cdot GX}^{(k+1) \cdot GX - 1} \sum_{j=l \cdot GY}^{(l+1) \cdot GY - 1} F_m(i,j) \text{ for } 0 \le k < K; 0 \le l < L; 0 \le m < M \qquad (2)$$

where $E_m(k,l)$ = a picture element in the lower-resolution image for frame m;

GX = the width of pixel groups expressed in numbers of pixels F;

GY = the height of pixel groups expressed in numbers of pixels F;

K = the horizontal resolution of the lower-resolution image;

L = the vertical resolution of the lower-resolution image; and

$F_m(i,j)$ = a pixel in the format-independent image for frame m.

The horizontal size GX of the groups is chosen such that $K \cdot GX = RH$ and the vertical size GY of the groups is chosen such that $L \cdot GY = RV$ where RH and RV are the horizontal and vertical resolutions of the format-independent image, respectively. For the exemplary implementation discussed above that generates elements in a down-sampled format-independent image with a resolution of 120 x 160 pixels, one suitable size for the groups is 8 x 8, which provides a lower-resolution image with a resolution of 120/8 x 160/8 = 15 x 20 picture elements.

Alternatively, the grouping performed by the spatial-domain processor 130 can be combined with or performed prior to processing performed by the image pre-processor 110.

By using the lower-resolution picture elements E to generate a video signature rather than the higher-resolution pixels F, the generated video signature is less sensitive to processes that change details of video signal content but preserve average intensity.

3. Temporal-Domain Processor

In an exemplary implementation of the temporal-domain processor 150, values that represent a composite of the series of lower-resolution images are obtained from the temporal averages and variances of respective picture elements E.

The temporal average $Z(k,l)$ of each respective picture element $E(k,l)$ may be calculated from the following expression:

$$Z(k,l)=\frac{1}{M}\sum_{m=0}^{M-1} E_m(k,l) \text{ for } 0\le k<K; 0\le l<L \quad (3a)$$

Alternatively, the video content of selected frames within the segment 3 may be given greater importance by calculating the temporal averages from a weighted sum of the picture elements as shown in the following expression:

$$Z(k,l)=\frac{1}{M}\sum_{m=0}^{M-1} w_m \cdot E_m(k,l) \text{ for } 0\le k<K; 0\le l<L \quad (3b)$$

where w_m = the weighting factor for picture elements in the lower-resolution image derived from the video content of frame m.

If desired, the time-domain process represented by expression 3a or 3b may be performed prior to the spatial-domain process represented by expression 2.

The value $Z(k,l)$ represents an average intensity for each picture element $E(k,l)$ over both time and space; therefore, these average values do not convey much information about any motion that may be represented by the video content of the segment 3. A representation of motion may be obtained by calculating the variance of each picture element $E(k,l)$.

If the average value $Z(k,l)$ for each picture element $E(k,l)$ is calculated as shown in expression 3a, the variance $V(k,l)$ of each respective picture element $E(k,l)$ may be calculated from the following expression:

$$V(k,l)=\frac{1}{M}\sum_{m=0}^{M-1} \left(E_m(k,l)-Z(k,l)\right)^2 \text{ for } 0\le k<K; 0\le l<L \quad (4a)$$

If the average value for each picture element is calculated as shown in expression 3b, the variance $V(k,l)$ of each respective picture element $E(k,l)$ may be calculated from the following expression:

$$V(k,l)=\frac{1}{M}\sum_{m=0}^{M-1} \left(w_m \cdot \left[E_m(k,l)-Z(k,l)\right]\right)^2 \text{ for } 0\le k<K; 0\le l<L \quad (4b)$$

In a preferred implementation, the values that represent a composite of the series of lower-resolution images are the values of elements in two rank matrices $\boldsymbol{Z}_r$ and $\boldsymbol{V}_r$ that are derived from the temporal average and variance arrays $\boldsymbol{Z}$ and $\boldsymbol{V}$, respectively. The value of each element in the rank matrices represents the rank order of its respective element in the associated arrays. For example, if the element $Z(2,3)$ is the fourth largest element in the average value array $\boldsymbol{Z}$, the value of the corresponding element $Z_r(2,3)$ in

the rank matrix $\boldsymbol{Z}_r$ is equal to 4. For this preferred implementation, the composite values QZ and QV may be expressed as:

$$QZ(k,l) = Z_r(k,l) \text{ for } 0 \le k < K; 0 \le l < L \quad (5)$$

$$QV(k,l) = V_r(k,l) \text{ for } 0 \le k < K; 0 \le l < L \quad (6)$$

The use of rank matrices is optional. In an alternate implementation, the values that represent a composite of the series of lower-resolution images are the values of the elements in the temporal average and variance arrays $\boldsymbol{Z}$ and $\boldsymbol{V}$. For this alternate implementation, the composite values QZ and QV may be expressed as:

$$QZ(k,l) = Z(k,l) \text{ for } 0 \le k < K; 0 \le l < L \quad (7)$$

$$QV(k,l) = V(k,l) \text{ for } 0 \le k < K; 0 \le l < L \quad (8)$$

4. Video Signature Processor

The video signature processor 170 applies a hash function to K x L arrays of the composite values QZ and QV to generate two sets of hash bits. A combination of these two sets of hash bits constitute the video signature that identifies the content of the segment 3. Preferably, the hash function is relatively insensitive to changes in the composite values and more sensitive to changes in any hash key that may be used. Unlike a typical cryptographic hash function whose output changes significantly with a change to even a single bit of its input, a preferred hash function for this application provides an output that undergoes only small changes for small changes in the input composite values. This allows the generated video signature to change only slightly with small changes to video content.

One suitable hash function uses a set of N_Z base matrices to generate a set of N_Z hash bits for the QZ composite values, and uses a set of N_V base matrices to generate a set of N_V hash bits for the QV composite values. Each of the base matrices is a K x L array of elements. These elements represent a set of vectors that preferably are orthogonal or nearly orthogonal to one another. In the implementation described below, the elements of the base matrices are generated by a random-number generator under the assumption that these elements represent a set of vectors that are nearly orthogonal to one another.

The matrix elements $pz_n(k,l)$ of each base matrix $\boldsymbol{PZ}_n$ for use with the composite values QZ may be generated from the following expression:

$$pz_n(k,l) = \text{RGN} - \bar{p}_n \quad \text{for } 1 \le n \le N_Z,\ 0 \le k < K,\ 0 \le l < L \quad (9)$$

where RNG = the output of a random-number generator; and

$\bar{p}_n$ = the average value of the numbers generated by RNG for each matrix.

The matrix elements $pv_n(k,l)$ of each base matrix $\boldsymbol{PV}_n$ for use with the composite values QV may be generated from the following expression:

$$pv_n(k,l) = \text{RGN} - \bar{p}_n \quad \text{for } 1 \le n \le N_V,\ 0 \le k < K,\ 0 \le l < L \tag{10}$$

The generator RNG generates random or pseudo-random values that are uniformly distributed in the range [0,1]. The initial state of the generator may be initialized by a hash key, which allows the hash function and the generated video signature to be cryptographically more secure.

One set of hash bits BZ_n is obtained by first projecting the composite values QZ onto each of the N_Z base matrices, which may be expressed as:

$$HZ_n = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} QZ(k,l) \cdot pz_n(k,l) \quad \text{for } 1 \le n \le N_Z \tag{11}$$

where HZ_n = the projection of the composite values QZ onto the base matrix $\boldsymbol{PZ}_n$.

The set of hash bits BZ_n is then obtained by comparing each projection to the median value of all projections and setting the hash bit to a first value if the projection is equal to or exceeds the threshold and setting the hash bit to a second value if the projection is less than the threshold. One example of this process may be expressed as:

$$BZ_n = \text{sgn}\left(HZ_n - \bar{H}_Z\right) \tag{12}$$

where $\text{sgn}(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \ge 0 \end{cases}$ and

$\bar{H}_Z$ = the median value of all projections HZ_n.

Another set of hash bits BV_n is obtained in a similar manner as shown in the following expressions:

$$HV_n = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} QV(k,l) \cdot pv_n(k,l) \quad \text{for } 1 \le n \le N_V \tag{13}$$

$$BV_n = \text{sgn}\left(HV_n - \bar{H}_V\right) \tag{14}$$

where HV_n = the projection of the composite values QV onto the base matrix $\boldsymbol{PV}_n$; and

$\bar{H}_V$ = the median value of all projections HV_n.

The video signature is obtained from a concatenation of the two sets of hash bits, which forms a value that has a total bit length equal to $N_Z + N_V$. The values for N_Z and

N_V may be set to provide the desired total bit length as well as weight the relative contribution of the composite values QZ and QV to the final video signature. In one application mentioned above that generates video signatures for television signals, N_Z and N_V are both set equal to eighteen.

C. Applications

1. Signature Sets

A signature generated by the video signature generator 100 represents the video content of the segment from which the signature was generated. A reliable identification of the video content in an interval of a signal much longer than a segment can be obtained by generating a set of signatures for the segments included in that interval.

The diagram shown in Fig. 5 is a schematic illustration of an interval of a signal that includes several segments of video frames. Five segments are shown. The first segment 3 of the signal includes video frames 3a to 3d. Each subsequent segment 4, 5, 6, 7 includes video frames 4a to 4d, 5a to 5d, 6a to 6d and 7a to 7d, respectively. A set of signatures can be generated for these segments by using the video signal generator 100 to process the contents of the video frames in each segment as described above.

Each segment contains an integral number of video frames. Preferably, the series of frames in each segment conveys video content for an interval of time that is equal to a nominal length L or within one frame period of the nominal length L. The term "frame period" refers to the duration of the video content conveyed by one frame. The nominal start times t# for successive segments are separated from one another by an offset ΔT. This offset may be set equal to the frame period of the lowest frame rate of signals to be processed by the video signature generator 100. For example, if the lowest rate to be processed is twelve frames per second, the offset ΔT may be set equal to 1/12 sec. or about 83.3 msec.

The nominal length L may be chosen to balance competing interests of decreasing the sensitivity of the subsequently-generated video signature to content modifications such as frame-rate conversion and increasing the temporal resolution of the representation provided by the video signature. Empirical studies have shown that a nominal segment length L that corresponds to about two seconds of video content provides good results for many applications.

The specific values mentioned for the segment length L and the offset amount ΔT are only examples. If the offset ΔT is not equal to an integer number of frame

periods, the offset between the actual start times of successive segments can vary as shown in the figure by the different offset amounts $\Delta 1$ and $\Delta 2$. If desired, the length of the offset between actual start times may kept within one frame period of the nominal offset ΔT.

Fig. 6 is a schematic block diagram showing a set of video signatures 193 to 197 that are generated from the video content of segments 3 to 7, respectively. Referring to Figs. 5 and 6, the video signature generator 100 obtains the video content of the segment 3 starting at the nominal start time t1 and processes this video content to generate the video signature 193. The video signature generator 100 then obtains the video content of the segment 4 starting at the nominal start time t2 and processes this video content to generate the video signature 194. The generator continues by processing the video content in segments 5, 6 and 7, which begin at nominal start times t3, t4 and t5, respectively, to generate the video signatures 195, 196 and 197. Signatures may be generated for essentially any number of segments that may be desired.

The nominal start times do not need to correspond to any particular time data that may accompany the video content. In principle, the alignment between the nominal start times and the video content is arbitrary. For example, in one implementation the nominal start times are expressed as relative offsets from the beginning of a signal to be processed. Each segment begins with the video frame conveying video content having a start time that is closest to its respective nominal start time. Alternatively, each segment could begin with the video frame that spans the nominal start time for that segment. Essentially any alignment between beginning frame and nominal start time may be used.

2. Detection of Copies

The signature sets generated from segments of video content can be used to identify the content even when that content has been modified by a variety of processes including those mentioned above. The ability to determine reliably whether specified video content is a copy of a reference content, even when modified, can be used in a variety of ways including the following:

- Detection of unauthorized copies: Networks of peer-to-peer servers can facilitate the distribution of content but they can also increase the difficulty of detecting unauthorized or pirated copies of proprietary content because many copies of the content can exist among the peer-to-peer servers. A facility can automatically determine if any unauthorized copies exist in the network by

generating signature sets for all content available from the network and checking these signature sets against a data base of reference signature sets.

- Confirmation of broadcast: Businesses that contract with broadcast networks to distribute specified video content can confirm the terms of the contract were met by generating signature sets from signals received by a broadcast receiver and comparing these signature sets to reference signature sets for the specified content.
- Identification of reception: Businesses that provide ratings for broadcast networks can identify content that is received by a receiver by generating signature sets from the received signals and comparing those signature sets against reference signature sets.

Fig. 7 is a schematic block diagram of a system that may be used to implement a variety of applications such as those mentioned in the preceding list. The video signature generator 100 generates reference video signature sets from reference streams of video content received from the path 31. The generated reference video signature sets are stored in the signature data base 180. The reference signature sets may be stored with other information that may facilitate implementation of the application. For example, the reference signature sets may be stored with the underlying content itself or with information about the content such as the content owner, content licensing terms, title of the content or a textual description of the content. Each reference signature set has a data base search key. This key may be derived in any manner that may be desired. Preferably, the key is based on or derived from the signatures in the associated reference signature set.

Any specified video content may be checked against reference content represented by one or more signature sets stored in the signature data base. The content to be checked is referred to herein as the test content. The identity of the test video content may be checked by having the video signature generator 101 generate one or more test video signature sets from the test video content received from the path 33 and passing the test video signature sets to the video search engine 185. The video search engine 185 attempts to find reference video signature sets in the signature data base 180 that are exact or close matches to the test video signature sets.

In one implementation, the video search engine 185 receives one or more test signature sets from the video signature generator 101. Each test signature set includes an

ordered series of test signatures S_{TEST} in the order in which they were generated from the test content. The video search engine 185 receives reference signature sets from the signature data base 180 via the path 182. Each reference signature set includes an ordered series of reference signatures S_{REF} in the order in which they were generated from the corresponding reference content. The video search engine 185 determines the similarity between test content and a particular reference content by calculating a measure of dissimilarity DSM between the test signature set for the test content and the reference signature set for the particular reference content. This measure of dissimilarity DSM is derived from the Hamming distances between corresponding signatures in the series of signatures for the test signature set and the reference signature set for the particular reference content. This measure may be calculated in a number of ways including either of the following expressions:

$$DSM = \sum_{s=1}^{M} \left(\mathrm{HD}\left[S_{REF}(s), S_{TEST}(s) \right] \right) \tag{15a}$$

$$DSM = \sqrt{\sum_{s=1}^{M} \left(\mathrm{HD}\left[S_{REF}(s), S_{TEST}(s) \right] \right)^2} \tag{15b}$$

where DSM = the calculated measure of dissimilarity;

HD[x,y] = the Hamming distance between signatures x and y;

$S_{REF}(s)$ = the s-th signature in the series of reference signatures; and

$S_{TEST}(s)$ = the s-th signature in the series of test signatures.

The video search engine 185 searches the signature data base 180 for the reference signature set that yields the smallest measure of dissimilarity with the test signature set. The reference content associated with this reference signature set is the most likely candidate in the data base to share a common origin with the test content. If the measure of dissimilarity is less than some classification threshold, the test content associated with the test signature set is deemed to share a common origin with or be a copy of the reference content that is associated with the matching reference signature set. Empirical results suggest that good results can be obtained for a variety of video content using if the series of signatures in each signature set represent about two seconds of video content.

For ease of explanation in the following discussion, test content and some specified reference content are said to be "matching" if the test content shares a common origin with the specified reference content.

The value that is chosen for the classification threshold mentioned above affects the likelihood that test and reference content will be correctly recognized as either matching or not matching each other. It also affects the likelihood that an incorrect decision is made. The probability of an "incorrect negative decision" that matching content will be incorrectly classified as content that does not match increases as the value of the classification threshold decreases. Conversely, the probability of an "incorrect positive decision" that non-matching content will be incorrectly classified as content that does match increases as the value of the classification threshold increases.

The classification threshold may be set in any way that may be desired. One method that may be used to set the value of the classification threshold obtains the original video content that is represented by a reference signature set in the data base 180 and creates a number of copies of this original content. The copies are modified in a variety of ways such as by frame-rate conversion and any of the other intentional and unintentional modifications described above. The method generates a test signature set for each copy and calculates a first set of measures of dissimilarity DSM between the test signature sets and the reference signature set. The method also calculates a second set of measures of dissimilarity DSM between the test signature sets and the signature sets for other video content that do not share a common origin with the original content. The range of values in the two sets may not overlap. If they do overlap, the amount of overlap is typically a very small portion of the range of values in each set. The classification threshold is set to a value within the overlap or between the two ranges if they do not overlap. This threshold value may be adjusted according to the needs of the application to balance the risk of incurring either incorrect positive or incorrect negative decisions.

D. Implementation

Devices that incorporate various aspects of the present invention may be implemented in a variety of ways including software for execution by a computer or some other device that includes more specialized components such as digital signal processor (DSP) circuitry coupled to components similar to those found in a general-purpose computer. Fig. 8 is a schematic block diagram of a device 70 that may be used to implement aspects of the present invention. The processor 72 provides computing resources. RAM 73 is system random access memory (RAM) used by the processor 72 for processing. ROM 74 represents some form of persistent storage such as read only memory (ROM) for storing programs needed to operate the device 70 and possibly for carrying out various aspects of

the present invention. I/O control 75 represents interface circuitry to receive and transmit signals by way of the communication channels 76, 77. In the embodiment shown, all major system components connect to the bus 71, which may represent more than one physical or logical bus; however, a bus architecture is not required to implement the present invention.

In embodiments implemented by a general purpose computer system, additional components may be included for interfacing to devices such as a keyboard or mouse and a display, and for controlling a storage device 78 having a storage medium such as magnetic tape or disk, or an optical medium. The storage medium may be used to record programs of instructions for operating systems, utilities and applications, and may include programs that implement various aspects of the present invention.

The functions required to practice various aspects of the present invention can be performed by components that are implemented in a wide variety of ways including discrete logic components, integrated circuits, one or more ASICs and/or program-controlled processors. The manner in which these components are implemented is not important to the present invention.

Software implementations of the present invention may be conveyed by a variety of machine readable media such as baseband or modulated communication paths throughout the spectrum including from supersonic to ultraviolet frequencies, or storage media that convey information using essentially any recording technology including magnetic tape, cards or disk, optical cards or disc, and detectable markings on media including paper.

CLAIMS

1. A method that comprises:

receiving a series of frames in a video signal segment conveying video content representing a series of pictures with a specified resolution;

processing the video content conveyed by the segment to calculate composite values expressing average and variance of picture elements in a series of low-resolution images, each low-resolution image representing at least a portion of a respective picture in the series of pictures with a second resolution that is lower than the specified resolution; and

generating a signature that identifies the video content of the video signal segment by applying a hash function to a set of values obtained from the composite values.

2. The method of claim 1, wherein each value in the set of values is equal to a respective composite value.

3. The method of claim 1 that comprises obtaining the set of values from rank matrices for the composite values.

4. The method of any one of claims 1 through 3 that comprises generating the series of low-resolution images by cropping and down-sampling pixels representing the series of picture and generating the picture elements from averages of intensities for groups of pixels.

5. The method of any one of claims 1 through 4 that comprises:

generating a series of format-independent images by cropping and down-sampling pixels representing the series of picture; and

generating the picture elements from averages of intensities for groups of pixels in each format-independent image.

6. The method according to any one of claims 1 through 5, wherein an application of the hash function to the set of values obtained from the composite values comprises

projecting the set of values onto a set of random vectors, and wherein each component of the signature is derived from the projection of the set of values onto a respective random vector.

7. The method according to claim 6, wherein:

each respective vector in the set of random vectors has vector elements with values that are obtained from a difference between uniformly distributed random variables within a range from zero to one and an average of the uniformly distributed random variables for all vector elements of the respective vector;

the projection of the intermediate values onto a respective random vectors is obtained from an inner product of the intermediate values with the vector elements of the respective vector; and

each component of the signature has either a first value if the projection of the intermediate values onto a corresponding random vector is greater than a threshold or has a second value if the projection of the intermediate values onto the corresponding random vector is less than a threshold, wherein the threshold is equal to a median of the projections of intermediate values onto the set of random vectors.

8. The method of any one of claims 1 through 7 that comprises:

receiving a plurality of video signal segments;

calculating the composite values for each segment in the plurality of video signal segments; and

generating a plurality of signatures that identify the video content of the plurality of video signal segments to obtain a signature set, a respective signature in the plurality of signatures generated for each segment in the plurality of video signal segments.

9. The method of claim 8, wherein:

the segments of the video signal convey video content in frames representing a series of pictures at a frame rate with a frame period between adjacent frames; and

the method comprises forming the segments of the video signal with offsets between pairs of successive segments that are variable and within one frame period of being equal to a nominal time interval.

10. The method of claim 8 that comprises:

obtaining a first signature set with a plurality of signatures that identify the video content of segments of a first video signal, the segments of the first video signal conveying video content in frames representing a series of pictures at a first frame rate with a first period between adjacent frames;

obtaining a second signature set with a plurality of signatures that identify the video content of segments of a second video signal, the segments of the second video signal conveying video content in frames representing a series of pictures at a second frame rate with a second period between adjacent frames; and

deriving a measure of dissimilarity between the first signature set and the second signature set to determine whether the first video signal and the second video signal convey video content representing respective series of pictures originating from a common source of pictures.

11. The method of claim 10, wherein pairs of successive segments convey video content starting at respective times that are offset from one another by intervals of time that are greater than or equal to the first period and are greater than or equal to the second period.

12. The method of claim 11 that comprises:

forming the segments of the first video signal with offsets between pairs of successive segments that are variable and within one first period of being equal to a nominal time interval; and

forming the segments of the second video signal with offsets between pairs of successive segments that are variable and within one second period of being equal to the nominal time interval.

13. The method of claim 10, wherein the first frame rate differs from the second frame rate, and the method comprises:

forming the segments of the first video signal with a first number of frames that represent an interval of video content that is within one first period of being equal to a nominal segment length; and

forming the segments of the second video signal with a second number of frames that represent an interval of video content that is within one second period of being equal to the nominal segment length, the second number being unequal to the first number.

14. An apparatus that comprises means for performing steps of the methods of any one of claims 1 through 13.

15. A medium conveying a program of instructions that is executable by a device to perform the method of any one of claims 1 through 13.

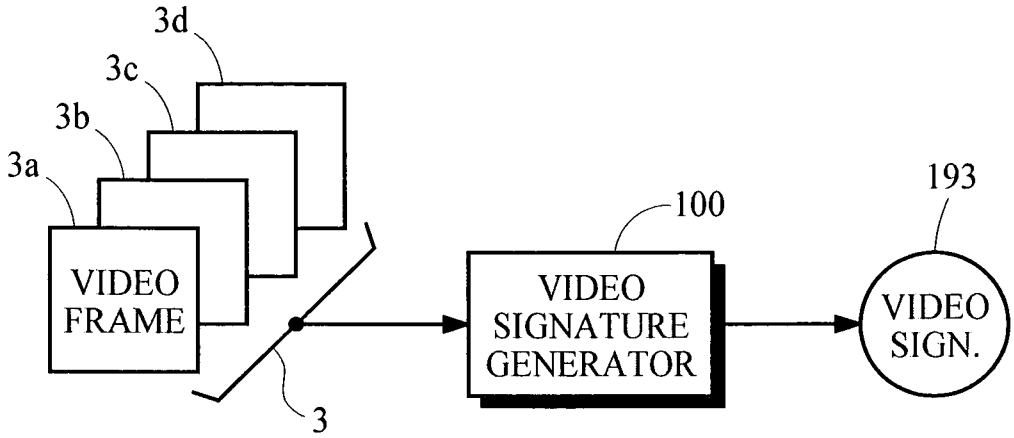


Fig. 1

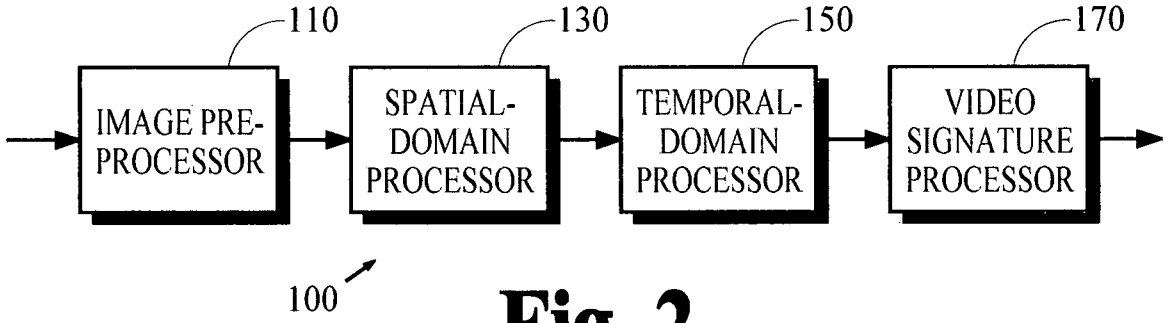


Fig. 2

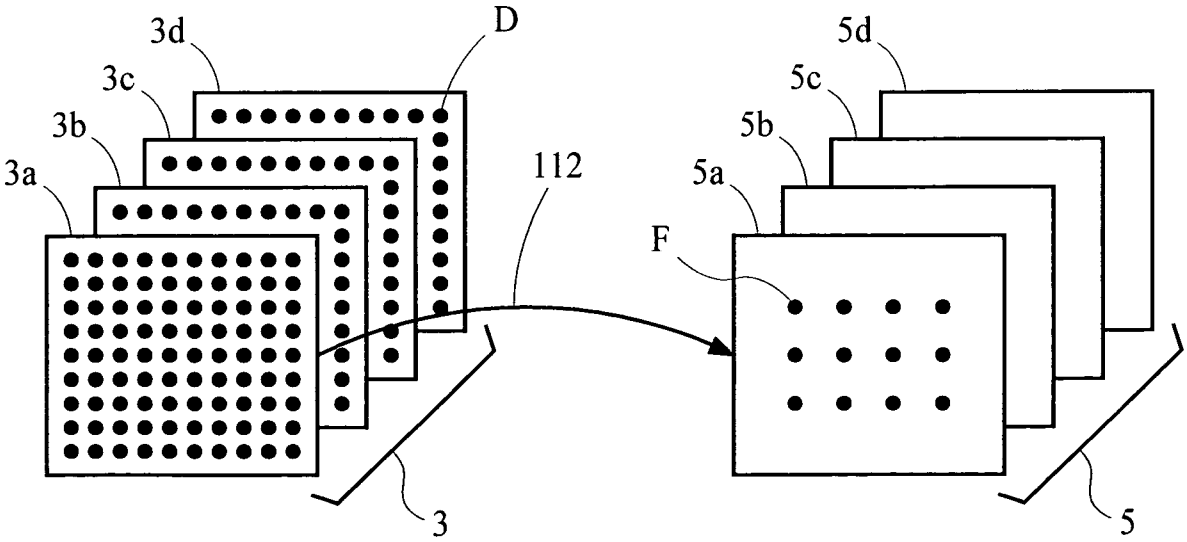


Fig. 3

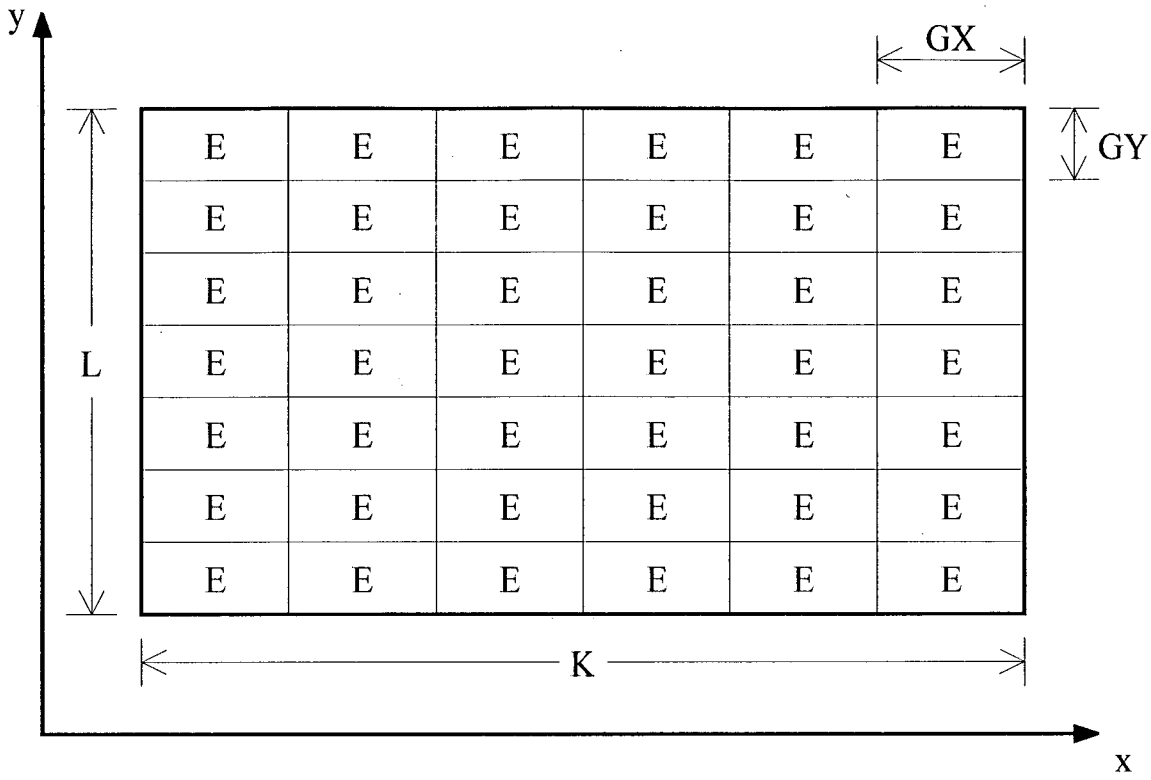

y
GX
GY
E
L
K
x

Fig. 4

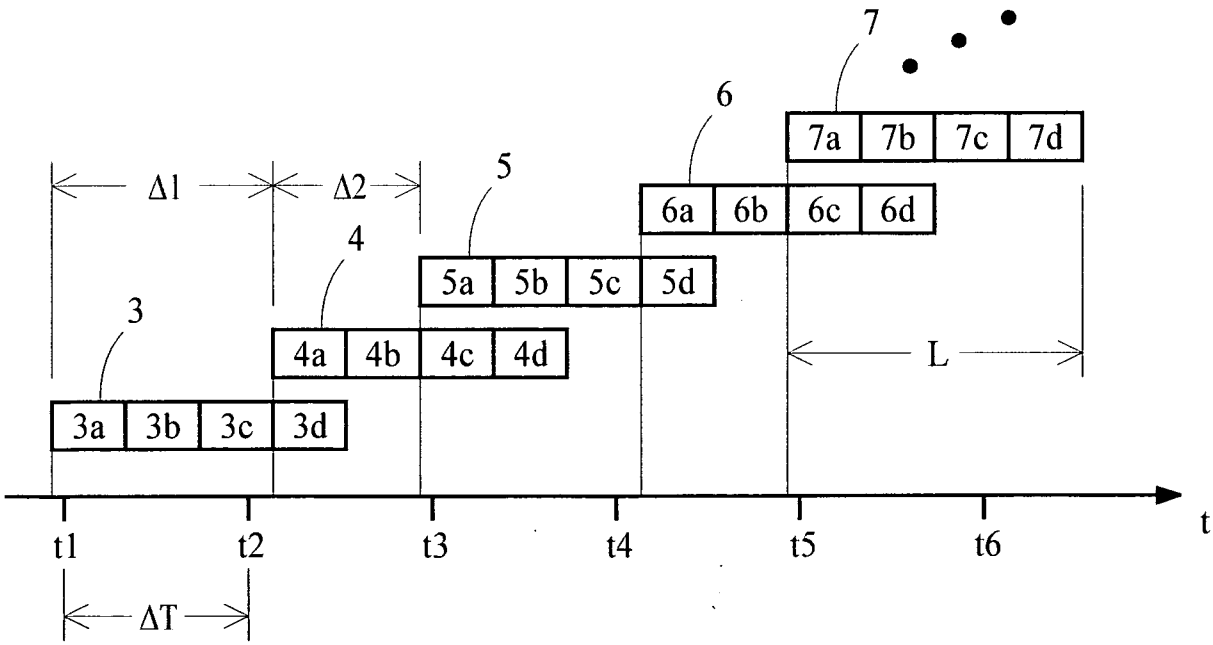

7
6
5
4
3
7a 7b 7c 7d
6a 6b 6c 6d
5a 5b 5c 5d
4a 4b 4c 4d
3a 3b 3c 3d
Δ1
Δ2
L
t
t1 t2 t3 t4 t5 t6
ΔT

Fig. 5

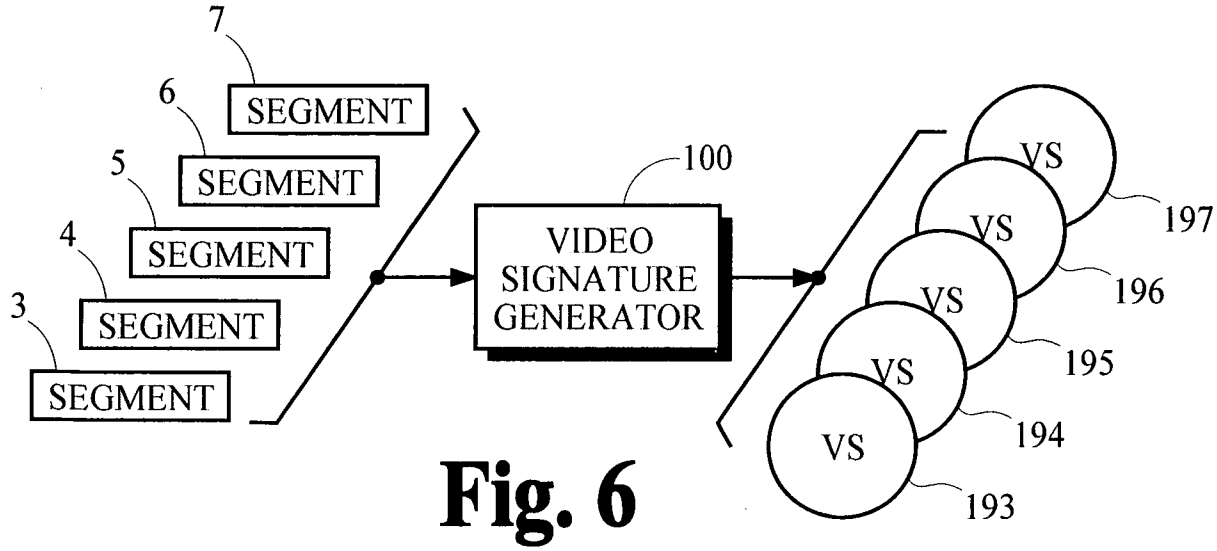


Fig. 6

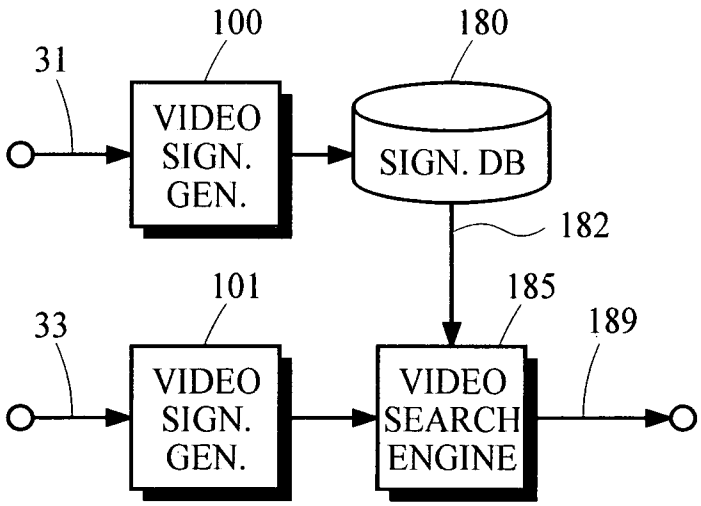


Fig. 7

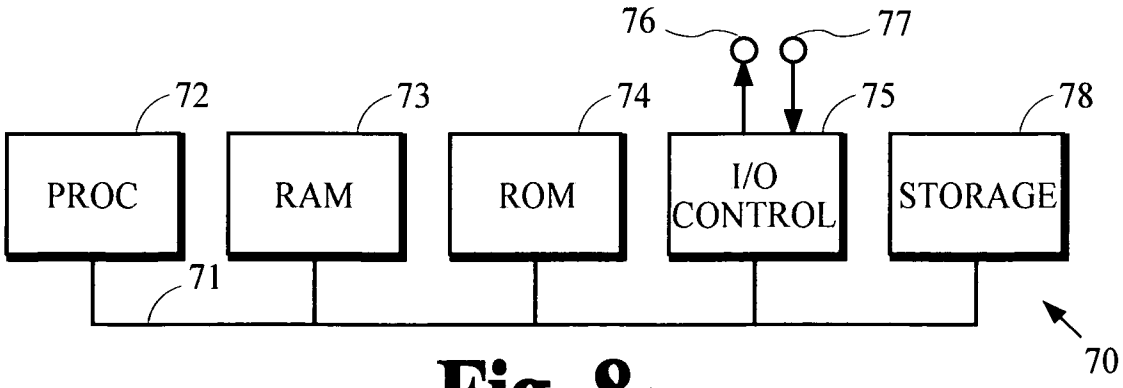


Fig. 8

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2008/005588

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F17/30 H04N5/91

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)
H04N G06F G06T G11B

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2006/184963 A1 (SNIJDER FREDDY F [NL] ET AL) 17 August 2006 (2006-08-17) paragraph [0001] - paragraph [0008] paragraph [0021] - paragraph [0043]; claims 1-22	1-15
Y	US 2004/240562 A1 (BARGERON DAVID M [US] ET AL) 2 December 2004 (2004-12-02) paragraph [0035] - paragraph [0096] claims 1-61	1-15
Y	US 5 465 353 A (HULL JONATHAN J [US] ET AL) 7 November 1995 (1995-11-07) the whole document	1-15

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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 document 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	Date of mailing of the international search report
16 July 2008	30/07/2008

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

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2008/005588

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 870 754 A (DIMITROVA NEVENKA [US] ET AL) 9 February 1999 (1999-02-09) column 2, line 65 - column 22, line 25; claims 1-21; figures 1-12	1-15
A	US 2005/175224 A1 (VENKATESAN RAMARATHNAM [US] ET AL VENKATESAN RAMARATHNAM [US] ET AL) 11 August 2005 (2005-08-11) paragraph [0010] - paragraph [0160]; claims 1-29; figures 1-5	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2008/005588

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006184963 A1	17-08-2006	NONE	
US 2004240562 A1	02-12-2004	CN 1574953 A	02-02-2005
		EP 1482734 A2	01-12-2004
		JP 2004357302 A	16-12-2004
		KR 20040103351 A	08-12-2004
US 5465353 A	07-11-1995	JP 7282088 A	27-10-1995
US 5870754 A	09-02-1999	DE 69737824 T2	06-03-2008
		EP 0842478 A1	20-05-1998
		WO 9740454 A1	30-10-1997
		JP 11509025 T	03-08-1999
US 2005175224 A1	11-08-2005	AU 2004240154 A1	25-08-2005
		BR 0405606 A	27-09-2005
		CA 2491826 A1	11-08-2005
		CN 1655500 A	17-08-2005
		EP 1569063 A2	31-08-2005
		JP 2005227756 A	25-08-2005
		KR 20050081159 A	18-08-2005
		MX PA05000524 A	18-10-2005