(54) Title: DETECTING VIDEO COPIES
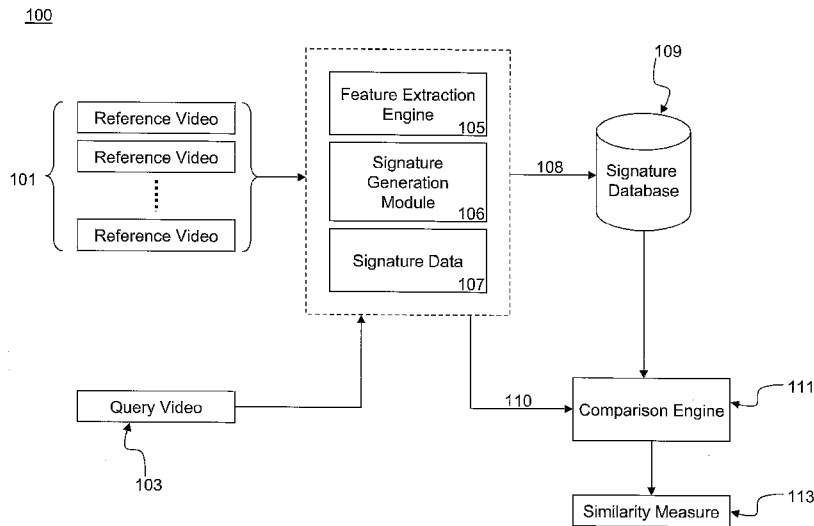
Figure 1

(57) Abstract: A computer-implemented method for detecting a copy of a reference video, comprises segmenting respective ones of
multiple frames of the reference video into multiple regions, determining sets of image features appearing in respective ones of the
multiple frames, determining a measure for the relative number of image features for a given region across the multiple frames, gen -
erating a spatio-temporal signature for the reference video using the determined measures, and comparing the signature for the refer-
ence video against a spatio-temporal signature of a query video to determine a likelihood of a match.

DETECTING VIDEO COPIES

BACKGROUND

Devices with digital video recording capability are readily available and easily affordable. In fact, multimedia technologies have advanced to the point where video recording capabilities are commonly included as part of electronic devices such as digital cameras, cell phones and personal digital assistants for example. Alongside the popularity and ease of capturing video content, there are now many options for sharing and consuming captured content. For example, it is straightforward to distribute video content from a computer onto social media websites for example, and more and more people are uploading and sharing video content in this and similar ways.

This situation creates issues relating to data management, such as database optimization for example. It is inefficient to store multiple copies of the same video in a database as it creates needless infrastructure expenses and complicates search and retrieval algorithms. Another issue relates to copyright infringement. There are ways to copy commercial content and redistribute it over the Internet for example. This can result in loss of revenue for a business, and it is typically not feasible to manually sift through hours of videos to determine if an illegal copy has been made or distributed.

SUMMARY

Detecting whether a video already exists in a database can allow more effective use of storage. In addition, automated video copy detection techniques can be used to detect copyright violations as well as to monitor usage. According to an example, there is provided a content-based video copy detection method, which generates signatures that capture spatial and temporal features of videos. Typically, a computed signature is compact, and can be created from individual video frames. For example, each video frame can be divided into multiple regions and discriminating or salient visual features for each region can be determined. In an example, a count of the visual features in each region can be used as a spatial

signature. To determine a spatial signature, the number of counts in each region can be sorted along a time line and an ordinal value assigned based on its temporal rank for example. Temporal and spatial signatures can be combined, and the resultant signature can be compared against signatures of different videos.

According to an example, there is provided a computer-implemented method for detecting a copy of a reference video, comprising segmenting respective ones of multiple frames of the reference video into multiple regions, determining sets of image features appearing in respective ones of the regions, determining a measure for the relative number of image features for a given region across the multiple frames, generating a spatio-temporal signature for the reference video using the determined measures, and comparing the signature for the reference video against a spatio-temporal signature of a query video to determine a likelihood of a match. A set of static pixels for the multiple frames with an intensity variance below a predetermined threshold intensity value is determined and disregard or otherwise excluded or ignored from further processing. Such static pixels can relate to objects in a video which have been inserted such as text and borders and the like. In an example, segmenting includes specifying a number of horizontal and vertical partitions to define a number of regions for the multiple frames. A set of static pixels for the multiple frames can be determined with a colour variance below a predetermined threshold value, and disregarded as above.

In an example, determining sets of image features can include using a scale- and rotation-invariant interest point detection method to determine an interest point in a region of a frame having a plurality of pixels, the interest point having a location in the region and an orientation. A spatial element of the spatio-temporal signature can represent the number of interest points for regions of the multiple frames, and a temporal element can represent the number of interest points for regions of the multiple frames after the number is sorted in each region along a time line of the multiple frames. The spatio-temporal signature can include spatial information from segmenting the frames into regions and temporal information from ranking

each region along a time line for the multiple frames. In an example, comparing the signature includes generating a spatio-temporal signature for the query video, and computing the distance between the spatio-temporal signature for the reference video and the spatio-temporal signature for the query video.

According to an example, there is provided apparatus for extracting a spatio-temporal signature from video data to detect a copy, comprising a segmentation module to segment respective ones of multiple frames of a reference video into multiple regions, a feature extraction engine to generate feature data representing sets of image features appearing in respective ones of the regions, a signature generation module to determine a measure for the relative number of image features for a given region across the multiple frames and to generate a spatio-temporal signature for the reference video using the determined measures, and a comparison engine to compare the signature for the reference video against a spatio-temporal signature of a query video to determine a likelihood of a match.

In an example, the comparison engine can generate a measure representing the similarity of the query video to the reference video. A reference video signature database to store the signature for the reference video can be provided. A static pixel identification engine to determine a set of static pixels for the multiple frames with an intensity or colour variance below a predetermined threshold value can be provided. In an example, the segmentation module can receive partition data representing a number of horizontal and vertical partitions to define a number of regions for the multiple frames. The feature extraction engine can determine sets of image features using a scale- and rotation-invariant interest point detection system configured to determine an interest point in a region of a frame having a plurality of pixels, the interest point having a location in the region and an orientation.

According to an example, there is provided a computer program embedded on a non-transitory tangible computer readable storage medium, the computer program including machine readable instructions that, when executed by a processor, implement a method for detecting a copy of a reference video, comprising

segmenting respective ones of multiple frames of the reference video into multiple regions, determining sets of image features appearing in respective ones of the regions, determining a measure for the relative number of image features for a given region across the multiple frames, generating a spatio-temporal signature for

5   the reference video using the determined measures, and comparing the signature for the reference video against a spatio-temporal signature of a query video to determine a likelihood of a match.

10   BRIEF DESCRIPTION OF THE DRAWINGS
An embodiment of the invention will now be described, by way of example only, and with reference to the accompanying drawings, in which:

Figure 1 is a schematic block diagram of a basic architecture for video copy
15   detection system according to an example;

Figure 2 is schematic diagram of four frames of a video, such as a reference or query video according to an example;

20   Figure 3 is a schematic block diagram of a method according to an example; and

Figure 4 is a schematic block diagram of an apparatus according to an example.

25   DETAILED DESCRIPTION
Figure 1 is a schematic block diagram of a system for video copy detection according to an example. A reference video 101 is an original video. A query video 103 is a video that is a candidate copy of a reference video 101. In an example, a feature extraction engine 105 is used to determine features for

30   reference and query videos by generating feature data representing sets of image features appearing in respective ones of regions of frames for the videos. A signature generation module 106 determines a measure for the relative number of image features for a given region across the multiple frames and generates

signature data 107 representing a spatio-temporal signature for the reference and query videos using the determined measures. The signature 108 for the reference video 101 is stored in a signature database 109. A signature 110 for a query video 103 is compared with a signature 108 using a comparison engine 111, and a

5   similarity measure 113 is generated which provides a measure representing the degree to which the query video and reference video are the same. In an example, signatures 108, 110 are combined signatures that capture spatial and temporal features of videos 101, 103.

10   In an example, the spatial part of a signature for a video is generated by dividing multiple video frames into regions. Local features are detected and counted in each region. These feature counts in the regions of a frame represent the spatial part of the signature for that frame. Figure 2 is schematic diagram of four frames of a video, such as a reference or query video. In the example of figure 2, each

15   frame (frames #1-4) have each been segmented or partitioned into four regions. That is, each frame has been partitioned using one vertical and one horizontal partition resulting in four regions per frame. The same partitioning is used across the frames to maintain symmetry between the regions for the frames. For each frame, the number of features within a region is determined. Any suitable

20   technique to determine features of an image can be used. In an example, a speeded-up robust feature detection method is used to determine features. Typically, such a method is a scale- and rotation-invariant interest point detector which uses integral images for image convolutions. In an example, the number of detected features for a region can be stored in storage device as will be described

25   below.

With reference to figure 2, the number of detected features for regions of the frames 1-4 can be seen. For respective regions across the frames, the number of features can be used to provide a ranking such that the higher the number of

30   detected features for that region the higher the ranking. For example, considering the upper left region for the frames in figure 2, it can be seen that this region in frame #1 would be ranked 4 as it has the lowest number of detected features of all

upper left regions (i.e. similar or corresponding regions) across the frames. Similarly, the corresponding region in frame #3 is ranked 1 for example.

The matrix S shows the spatial part of the signature. That is, feature counts from regions of the multiple frames are transferred into a matrix S. Rows represent the feature counts for a region across frames. Accordingly, a column represents feature counts for a frame. In order to add temporal information to the signature, the number of counts in each region is sorted along the time line of the frames, and assigned an ordinal value based on its temporal rank. In the example of figure 2, each row of the matrix λ shows the temporal part of the signature for one region. The element $\lambda_{i,j}$ represents the rank (or order) of region i in frame j. For example, region 1 in frame 1 has the smallest number of features in across all four frames, and thus has a rank of 4 in the λ matrix.

Therefore, according to an example the final combined signature of a video clip (matrix λ) includes two parts: spatial information generated from partitioning frames into regions, and temporal information generated by ranking each region along the time-line of the frames. In order to detect copies signatures can be compared using any similarity metric. In an example, an L1 (1-norm) distance is used in order to determine the similarity between videos.

Figure 3 is a schematic block diagram of a method according to an example. In block 301 a video 300 is divided into regions 302. More specifically, frames 303 of a portion of a video 300, which can be a reference or query video as described above, are segmented into regions 302. The number of regions 302 is a configurable parameter 304, which can be set by specifying a number of vertical 305 and horizontal 306 partitions for the frames of the video 300. For example, specifying 3 horizontal and 2 vertical partitions would divide respective frames into a grid with 3 rows and 2 columns for a total of 6 regions.

In block 307 the percentage of static pixels in each region is calculated. The presence of static pixels within the cropped image can indicate the presence of an

image, some text, a logo, or a background pattern superimposed onto the video. If a significant number of pixels within a region are masked then too much of the area may be occluded to obtain useful information. If this is the case, the region in question can be turned off or disregarded, and the process can proceed based on the remaining regions.

That is, according to an example, objects such as static borders, letter-box and pillar-box effects for example can be removed or otherwise disregarded. This can form a pre-processing step which is performed by determining how pixels change throughout the video clip under consideration. In an example, pixels with a variance 308 below a threshold value are likely to be edit effects which have been added to the video. These effects can include borders, logos and pattern insertions as well as letter-box and pillar-box effects from resizing. The variance can be calculated on each pixel as follows: a gray-scale value of pixel x in frame i is denoted x_i. Two quantities, M_k and Q_k are defined as follows:

$$M_k = \begin{cases} x_1, & k = 1 \\ M_{k-1} + \dfrac{x_k - M_k - 1}{k}, & k = 2, \ldots, n \end{cases}$$

$$Q_k = \begin{cases} 0, & k = 1 \\ Q_{k-1} + \dfrac{(k-1)(x_k - M_k - 1)^2}{k}, & k = 2, \ldots, n \end{cases}$$

For the nth frame, and once Q_n is calculated the variance is $\dfrac{Q_n}{n}$.

If all pixels in a row (or column) on the outside border of a frame have a variance below the predetermined threshold value, they are removed or otherwise disregarded from the image. The process is repeated until a row (or column) is encountered where at least one pixel shows variance above the threshold. The result is an image which is cropped of borders in which the pixels do not vary. The sub-image corresponding to the size of the cropped mask can be used for further processing. This will remove any pillar-box effects, letter-box effects, and borders

8

from cropping or shifting. In an example, the step of determining static pixels can be performed before or after a frame is segmented.

In block 309 the number of interest points in each region are determined. That is, the features for the frame are extracted, and a count for a region is incremented if an interest point resides in the region. As described above with reference to figure 2, the count of interest points in regions of frames for the video clip defines a spatial signature for the clip. In block 311 the temporal signature is generated. The temporal aspect of the signature is obtained by sorting the feature counts in each region along the time-line for the frames (denoted by the time axis in figure 2). In an example, the frame with the most interest points is assigned an ordinal value of 1. The frame with the next highest number of interest points is assigned a value of 2, and so on. The result is a matrix (for example) 313 where each row contains the ranking vector of a particular region.   The result in block 313 corresponds to the spatio-temporal signature for the video, such as a reference video for example.

More formally, for a video consisting of M frames and L regions, each region t_i would result in an M-dimension vector, $s_i = (f_{i,1}, f_{i,2}, ..., f_{i,M})$, where $f_{i,k}$ is the number of features counted in region i of frame k. The matrix $S_i = (s_1, s_2, ..., s_L)$ is used to produce the ranking matrix, $\lambda = (\lambda_1, \lambda_2, ..., \lambda_L)$. Each $\lambda_i = (r_1^i, r_2^i, ..., r_L^i)$, where r_k^i is the rank of the ith region of frame k.

For a video with M frames and L regions, the signature for the video consists of an $L \times M$ matrix. In order to calculate the distance between two signatures for a reference video and a query video, the L1 distance between them is used in an example. If the number of frames in the reference video is N and the number of frames in the query video is M, where $N \geq M$, each video is divided into L regions. A sliding window approach can then used where the distance between the query video and the first M frames of the reference video is calculated. The window of M frames is then slid over one frame and the distance between the

query video and M frames in the reference video starting at the second frame is determined. The minimum distance and the frame offset, p, for which this occurred is recorded as sliding proceeds. At the end of the reference video, the best match occurs at the minimum distance.

If λ_i is the ranking vector of the ith region, the distance between a query video V_q and a reference video V_r is calculated in an example as:

$$D\left(V_q, V_r\right) = p \, \mathbf{argmin} \left(D\left(V_q, V_r^p\right) \right).$$

where p is the frame offset in the reference video which achieved this minimum and represents the location of the best match between the query video and the reference video, and $D\left(V_q, V_r^p\right)$ is given by:

$$D\left(V_q, V_r^p\right) = \frac{1}{L}\sum_{i=1}^{L} d^p\left(\lambda_q^i, \lambda_r^i\right), \, where$$

$$d^p\left(\lambda_q^k, \lambda_r^i\right) = \frac{1}{C(M)}\sum_{j=1}^{M} \left| \lambda_q^k(j) - \lambda_r^j(k)(p+j) \right|.$$

Here, $C(M)$ is a normalizing factor which is a function of the size of the query. It represents the maximum possible distance between the reference video and the query video. This maximum distance occurs when the ranking of the reference video is exactly opposite to that of the query. There are two cases based on whether M is even or odd. In the case when M is even, it is twice the sum of the first $\frac{M}{2}$ odd integers. Similarly, when M is odd, $C(M)$ is twice the sum of the first $\frac{M-1}{2}$ even integers. Each of these sequences can be computed directly as follows:

$$C(M) = \left\{ \left(\frac{M}{2}\right)^2, \quad M \; even \right.$$

$$\left(\left|\frac{M}{2}\right|\right)\left(\left|\frac{M}{2}\right|+1\right), M \quad odd$$

In block 315 a query video 314 whose signature 316 has been determined as described above can be processed using comparison engine 111 in order to determine if it is a copy of the reference video. If the minimum distance between the signature 316 of the query video 314 and that of the reference video at offset p is below a threshold, then it is likely that the query video 314 is a copy of the reference video. In this case, an output such as similarity measure 113 can be provided indicating that a copy has been located starting in frame p of the reference video.

Accordingly, a method of an example proceeds by determining the distance between the set of frames from a query video and the first M frames in a reference video. It then shifts the comparison window and finds the distance between the query set and the reference set starting at the second frame of the reference set. This continues for a total of $N-M+1$ calculations. For every calculation, the M frames in the query are compared with M frames of the reference video for a total of $(N-M+1)M$ comparisons.

Figure 4 is a schematic block diagram of an apparatus according to an example suitable for implementing any of the systems, methods or processes described above. Apparatus 400 includes one or more processors, such as processor 401, providing an execution platform for executing machine readable instructions such as software. Commands and data from the processor 401 are communicated over a communication bus 399. The system 400 also includes a main memory 402, such as a Random Access Memory (RAM), where machine readable instructions may reside during runtime, and a secondary memory 405. The secondary memory 405 includes, for example, a hard disk drive 407 and/or a removable storage drive 430, representing a floppy diskette drive, a magnetic tape drive, a compact disk drive, etc., or a nonvolatile memory where a copy of the machine

readable instructions or software may be stored. The secondary memory 405 may also include ROM (read only memory), EPROM (erasable, programmable ROM), EEPROM (electrically erasable, programmable ROM). In addition to software, data representing any one or more of a reference video 101, query video 103 or

5   portions thereof including multiple frames for example, signature data 107 and similarity measures 113 may be stored in the main memory 402 and/or the secondary memory 405. The removable storage drive 430 reads from and/or writes to a removable storage unit 409 in a well-known manner.

10   A user can interface with the system 400 with one or more input devices 411, such as a keyboard, a mouse, a stylus, and the like in order to provide user input data . The display adaptor 415 interfaces with the communication bus 399 and the display 417 and receives display data from the processor 401 and converts the display data into display commands for the display 417. A network interface 419

15   is provided for communicating with other systems and devices via a network (not shown). The system can include a wireless interface 421 for communicating with wireless devices in the wireless community.

It will be apparent to one of ordinary skill in the art that one or more of the

20   components of the system 400 may not be included and/or other components may be added as is known in the art. The apparatus 400 shown in figure 4 is provided as an example of a possible platform that may be used, and other types of platforms may be used as is known in the art. One or more of the steps described above may be implemented as instructions embedded on a computer readable

25   medium and executed on the system 400. The steps may be embodied by a computer program, which may exist in a variety of forms both active and inactive. For example, they may exist as software program(s) comprised of program instructions in source code, object code, executable code or other formats for performing some of the steps. Any of the above may be embodied on a computer

30   readable medium, which include storage devices and signals, in compressed or uncompressed form. Examples of suitable computer readable storage devices include conventional computer system RAM (random access memory), ROM (read only memory), EPROM (erasable, programmable ROM), EEPROM

(electrically erasable, programmable ROM), and magnetic or optical disks or tapes. Examples of computer readable signals, whether modulated using a carrier or not, are signals that a computer system hosting or running a computer program may be configured to access, including signals downloaded through the Internet or other networks. Concrete examples of the foregoing include distribution of the programs on a CD ROM or via Internet download. In a sense, the Internet itself, as an abstract entity, is a computer readable medium. The same is true of computer networks in general. It is therefore to be understood that those functions enumerated above may be performed by any electronic device capable of executing the above-described functions.

According to an example, a feature extraction engine 105, signature generation module 106 and comparison engine 111 can reside in memory 402 and operate on data representing a reference video 101 or query video 103 to provide signature data 107 for comparison and for storage in database 109 for example. A database 109 can be provided on a HDD such as 405, or can be provided as a removable storage unit 409 for example. The database 109 can be remote from the apparatus 400 and can be connected thereto via the network interface for example.

CLAIMS

What is claimed is:

1.   A computer-implemented method for detecting a copy of a reference video,
comprising:

segmenting respective ones of multiple frames of the reference video into multiple
regions;

determining sets of image features appearing in respective ones of the multiple
frames;

determining a measure for the relative number of image features for a given region
across the multiple frames;

generating a spatio-temporal signature for the reference video using the
determined measures; and

comparing the signature for the reference video against a spatio-temporal
signature of a query video to determine a likelihood of a match.

2.   A computer-implemented method as claimed in claim 1, further comprising:
determining a set of static pixels for the multiple frames with an intensity variance
below a predetermined threshold intensity value; and

disregarding the static pixels.

3.   A computer-implemented method as claimed in claim 1, wherein
segmenting includes specifying a number of horizontal and vertical partitions to
define a number of regions for the multiple frames.

4.   A computer-implemented method as claimed in claim 1, further comprising:
determining a set of static pixels for the multiple frames with a colour variance
below a predetermined threshold value; and

disregarding the static pixels.

5.   A computer-implemented method as claimed in claim 1, wherein
determining sets of image features includes using a scale- and rotation-invariant
interest point detection method to determine an interest point in a region of a

frame having a plurality of pixels, the interest point having a location in the region and an orientation.

6.   A computer-implemented method as claimed in claim 1, wherein a spatial element of the spatio-temporal signature represents the number of interest points for regions of the multiple frames, and a temporal element represents the number of interest points for regions of the multiple frames after the number is sorted in each region along a time line of the multiple frames.

7.   A computer-implemented method as claimed in claim 1, wherein the spatio-temporal signature includes spatial information from segmenting the frames into regions and temporal information from ranking each region along a time line for the multiple frames.

8.   A computer-implemented method as claimed in claim 1, wherein comparing the signature includes:
generating a spatio-temporal signature for the query video; and
computing the distance between the spatio-temporal signature for the reference video and the spatio-temporal signature for the query video.

9.   Apparatus for extracting a spatio-temporal signature from video data to detect a copy, comprising:
a segmentation module to segment respective ones of multiple frames of a reference video into multiple regions;
a feature extraction engine to generate feature data representing sets of image features appearing in respective ones of the multiple frames;
a signature generation module to determine a measure for the relative number of image features for a given region across the multiple frames and to generate a spatio-temporal signature for the reference video using the determined measures; and
a comparison engine to compare the signature for the reference video against a spatio-temporal signature of a query video to determine a likelihood of a match.

10.   Apparatus as claimed in claim 9, the comparison engine further to generate a measure representing the similarity of the query video to the reference video.

11.   Apparatus as claimed in claim 9, further comprising:
a reference video signature database to store the signature for the reference video.

12.   Apparatus as claimed in claim 9, further comprising a static pixel identification engine to determine a set of static pixels for the multiple frames with an intensity or colour variance below a predetermined threshold value.

13.   Apparatus as claimed in claim 9, the segmentation module to receive partition data representing a number of horizontal and vertical partitions to define a number of regions for the multiple frames.

14.   Apparatus as claimed in claim 9, the feature extraction engine to determine sets of image features using a scale- and rotation-invariant interest point detection system configured to determine an interest point in a region of a frame having a plurality of pixels, the interest point having a location in the region and an orientation.

15.   A computer program embedded on a non-transitory tangible computer readable storage medium, the computer program including machine readable instructions that, when executed by a processor, implement a method for detecting a copy of a reference video, comprising:
segmenting respective ones of multiple frames of the reference video into multiple regions;
determining sets of image features appearing in respective ones of the regions;
determining a measure for the relative number of image features for a given region across the multiple frames;
generating a spatio-temporal signature for the reference video using the determined measures; and

comparing the signature for the reference video against a spatio-temporal signature of a query video to determine a likelihood of a match.
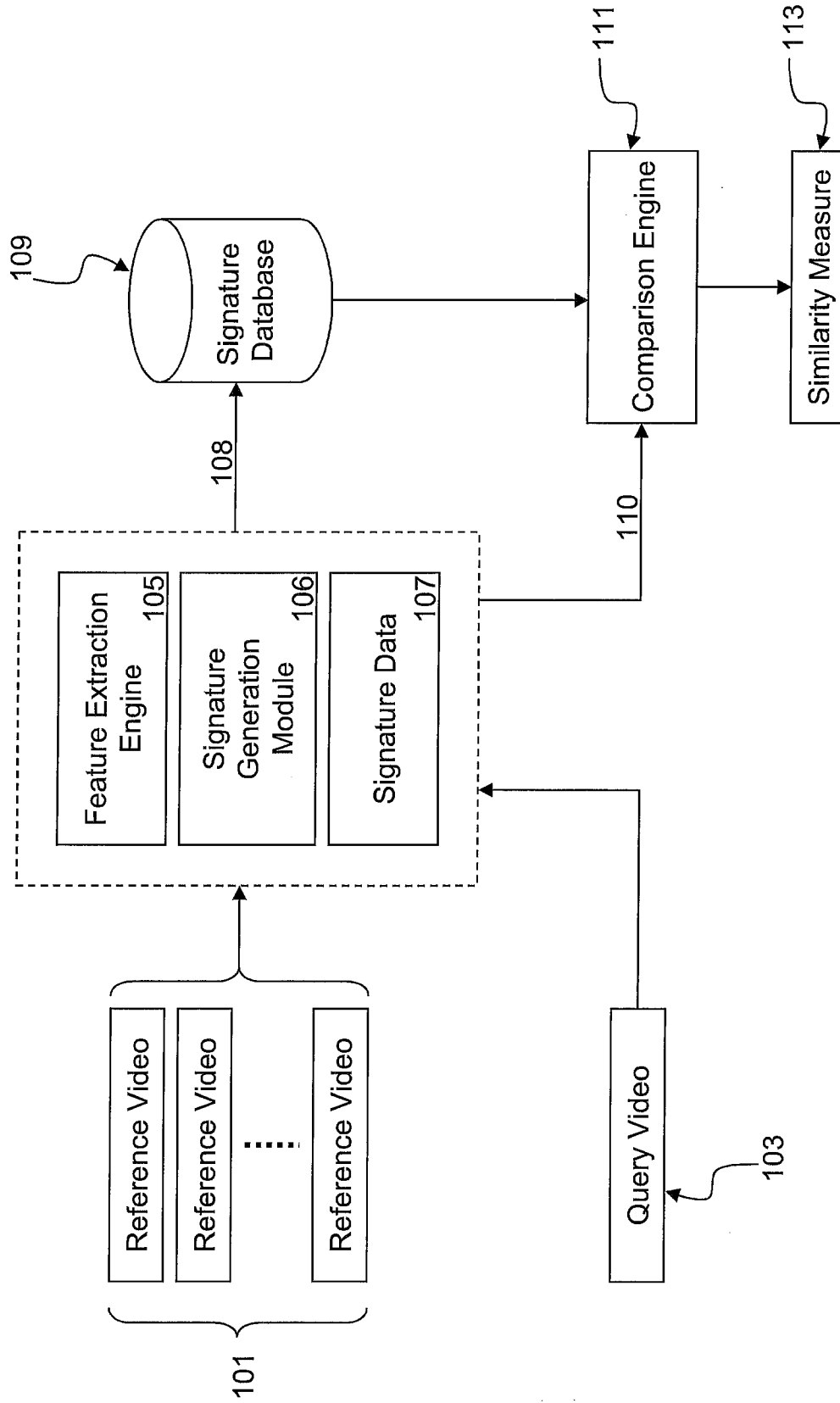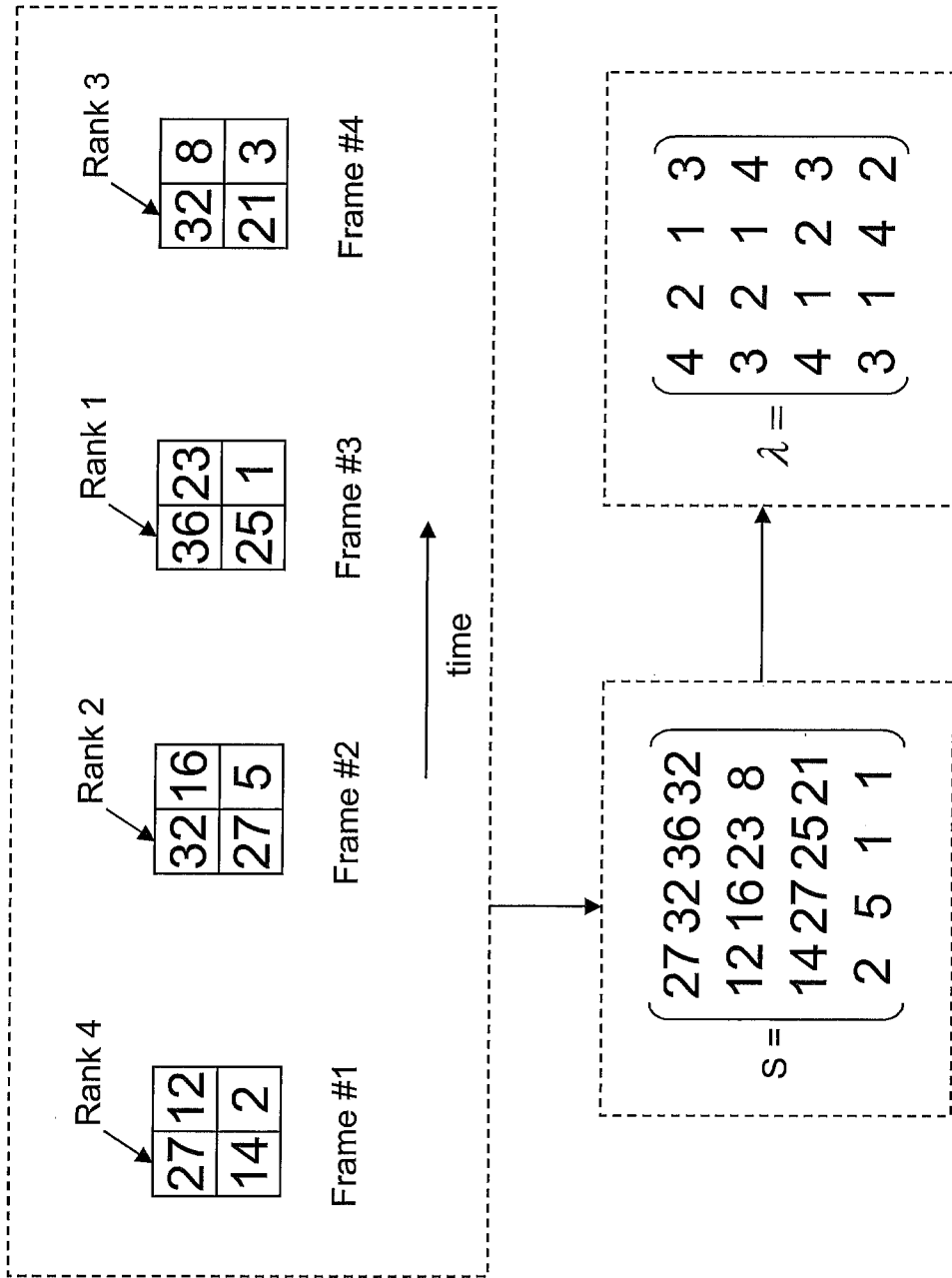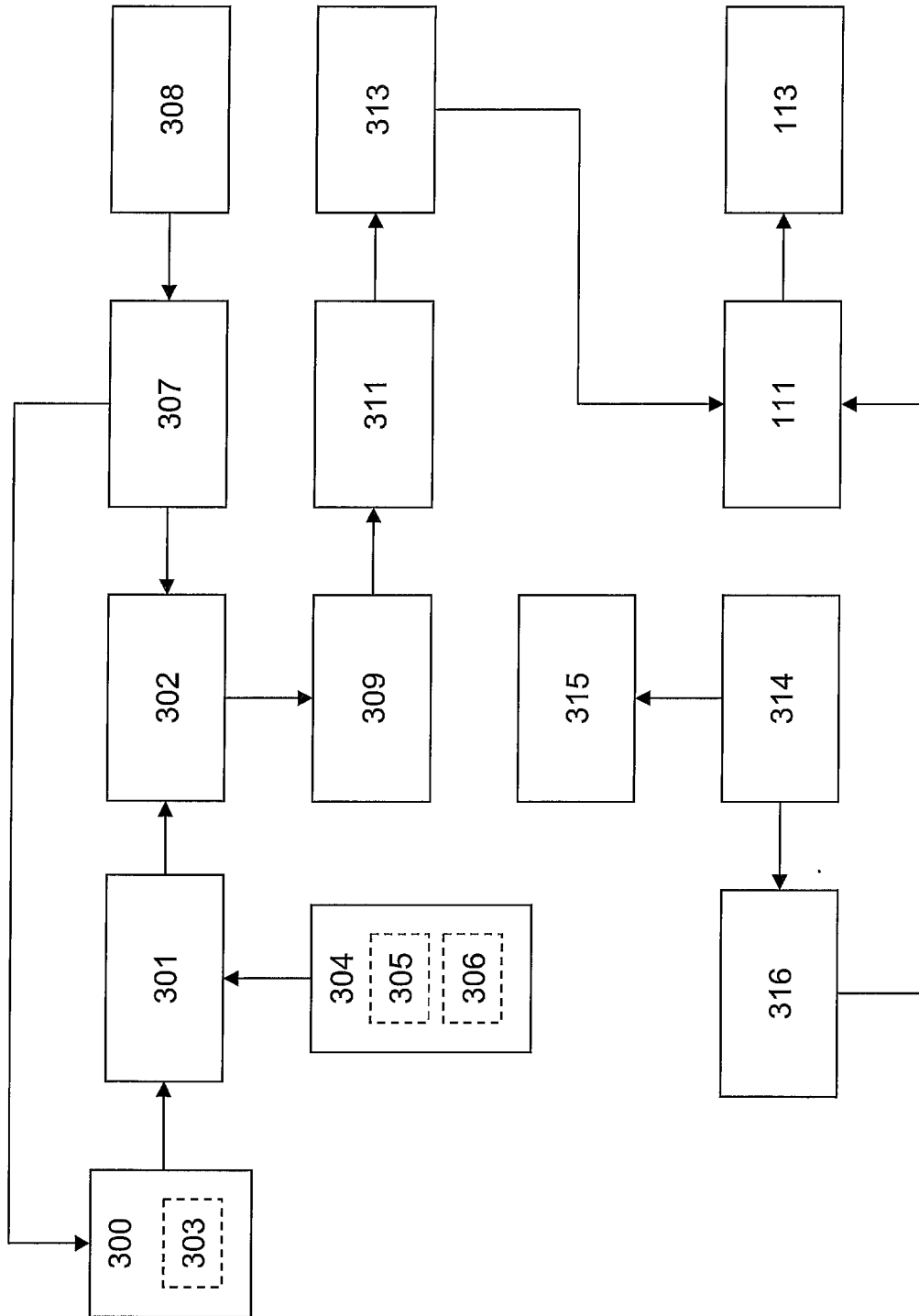
Figure 1

Figure 2

Figure 3

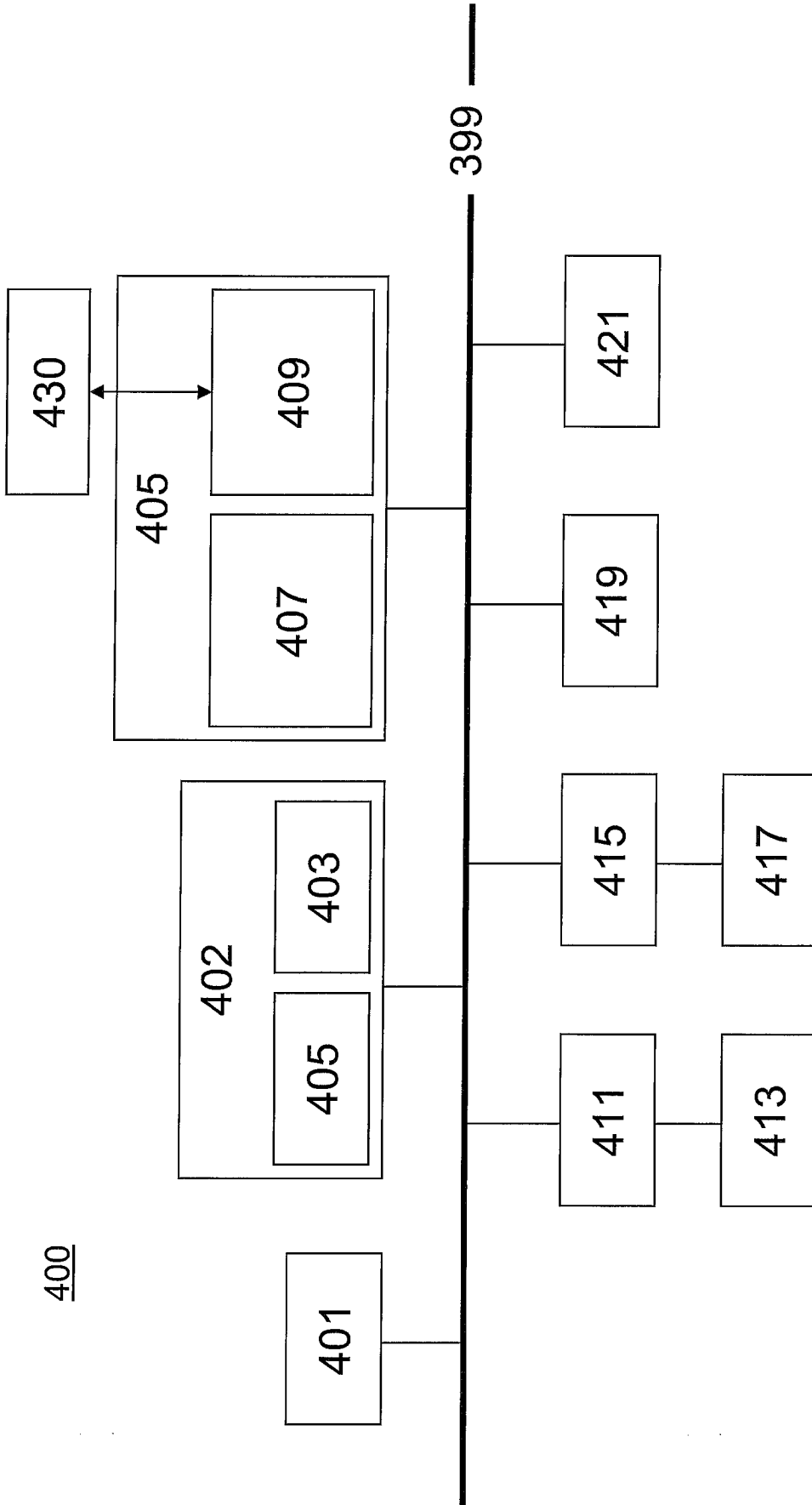Figure 4

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F17/30   G06K9/00
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 G06K

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	US 2007/253594 A1 (LU JIAN [US] ET AL) 1 November 2007 (2007-11-01)	1,3, 6-11,13, 15
Y	paragraph [0004] - paragraph [0017] paragraph [0024] - paragraph [0057]; figures 1,2 paragraph [0058] - paragraph [0093]; claims; figures 3-6 -----	2,4,12
X	EP 1 494 135 A2 (SEIKO EPSON CORP [JP]) 5 January 2005 (2005-01-05) paragraph [0002] - paragraph [0066]; claims; figures ----- -/--	1,3,5-9, 13-15

| 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 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	Date of mailing of the international search report
21 September 2012	02/10/2012

| 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 Bakker, Jeroen |

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MUHAMMAD USMAN ET AL: "Real Time Video Copy Detection under the Environments of Video Degradation and Editing", ADVANCED COMMUNICATION TECHNOLOGY, 2008. ICACT 2008. 10TH INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 17 February 2008 (2008-02-17), pages 1583-1588, XP031245411, ISBN: 978-89-5519-136-3	2,4,12
A	the whole document -----	1,9,15
X	KIM C ET AL: "Spatiotemporal Sequence Matching for Efficient Video Copy Detection", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 15, no. 1, 1 January 2005 (2005-01-01), pages 127-132, XP011124674, ISSN: 1051-8215, DOI: 10.1109/TCSVT.2004.836751 the whole document -----	1,3,6-9, 13-15
A	MOHAN R: "Video sequence matching", ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 1998. PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON SEATTLE, WA, USA 12-15 MAY 1998, NEW YORK, NY, USA,IEEE, US, vol. 6, 12 May 1998 (1998-05-12), pages 3697-3700, XP010279641, DOI: 10.1109/ICASSP.1998.679686 ISBN: 978-0-7803-4428-0 page 3697 - page 3699 -----	1-15
A	WO 2007/148264 A1 (KONINKL PHILIPS ELECTRONICS NV [NL]; HAITSMA JAAP A [NL]; BHARGAVA VIK) 27 December 2007 (2007-12-27) page 2, line 30 - page 5, line 22 claims; figures -----	2,4,12

-/--

3

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	Naghmeh Khodabakhshi ET AL: "Copy Detection of 3D Videos", MMSys'12, 24 February 2012 (2012-02-24), pages 131-142, XP55034653, Chapel Hill, North Carolina, USA Retrieved from the Internet: URL:http://delivery.acm.org/10.1145/216000 0/2155578/p131-khodabakhshi.pdf?ip=145.64. 134.242&acc=ACTIVE SERVICE&CFID=135513402&CFTOKEN=52890846&__ acm__=1343996770_4a39085f82bec6ca730e65100 8c56e9b [retrieved on 2012-08-03] the whole document -----	1-15

3

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report			Publication date	Patent family member(s)			Publication date
US 2007253594	A1		01-11-2007	CN	101351986	A	21-01-2009
				JP	4932903	B2	16-05-2012
				JP	2009535920	A	01-10-2009
				US	2007253594	A1	01-11-2007
				WO	2007127590	A2	08-11-2007
EP 1494135	A2		05-01-2005	CN	1574019	A	02-02-2005
				EP	1494135	A2	05-01-2005
				JP	2005020742	A	20-01-2005
				KR	20050000307	A	03-01-2005
				KR	20060117890	A	17-11-2006
				US	2004258397	A1	23-12-2004
WO 2007148264	A1		27-12-2007	CN	101473657	A	01-07-2009
				EP	2036354	A1	18-03-2009
				JP	2009542081	A	26-11-2009
				US	2009324199	A1	31-12-2009
				WO	2007148264	A1	27-12-2007