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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 September 2010 (10.09.2010)

PCT

(10) International Publication Number
WO 2010/099801 A1

(51) International Patent Classification:
H04N 7/64 (2006.01)

(21) International Application Number:
PCT/EP2009/001526

(22) International Filing Date:
4 March 2009 (04.03.2009)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **TELEFONAKTIEBOLAGET LM ERICSSON (publ)** [SE/SE]; S-164 83 Stockholm (SE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **RUSERT, Thomas** [DE/DE]; Adalbertsteinweg 175a, 52066 Aachen (DE). **KAMPMANN, Markus** [DE/DE]; Bremenberg 36a, 52072 Aachen (DE).

(74) Agent: **ZMUDA, Margarethe**; Ericsson GmbH, Ericsson Allee 1, 52134 Herzogenrath (DE).

(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, ST, 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, MK, 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).

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: PROCESSING OF MULTIMEDIA DATA

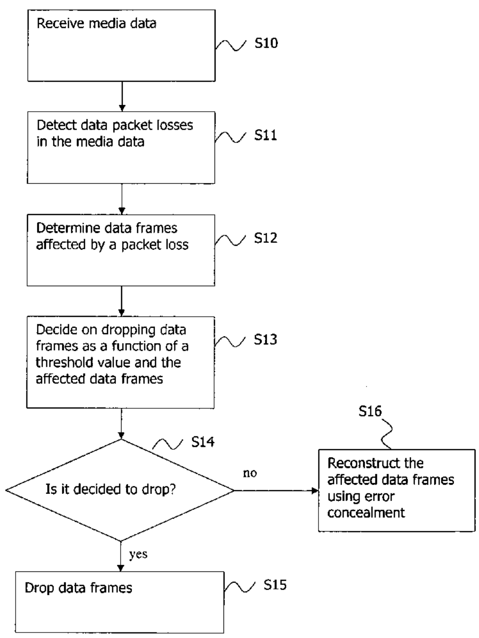

Fig.1

(57) Abstract: The invention relates to techniques for processing media data at a receiver device in a packet-switched network. It is proposed to detect data packet losses in the media data, and to determine data frames, like P-frames or I-frames, within a group of successive frames, like the GOP structure, affected by a packet loss. The number of the affected data frames is compared with a threshold value and if said number is equal or higher then the threshold value, then it is proposed to drop the affected data frames and to provide for playing-out data frames in form of a slide show.

WO 2010/099801 A1

Processing of multimedia data

Technical Field

The invention relates to techniques for processing media data on a receiver device in a packet-switched network, for example in an IP based network. The frames of the media data, in particular the frames of video data are processed before they are forwarded for presenting to the user.

The invention may be practiced with certain (TV) broadcast networks or with hybrid networks comprising a (TV) broadcast network and a mobile network, for example a DVB-H (Digital Video Broadcast-Handhelds) and a 3GPP mobile network. Basically, the invention may be practiced within any network environment in which video content may be distributed.

Background

Multimedia streaming is a major application in the internet and for 3G networks, and a packet-switched streaming service has been standardized in 3GPP allowing to stream multimedia data like audio and video data to handhelds. In another context, streaming is used for the realization of mobile TV or IPTV services.

In case of multimedia services like Mobile TV or IPTV, data packets containing media data are streamed from a server towards a terminal device. In the terminal device, for example a mobile terminal having implemented a video client adapted for the reception of mobile TV or IPTV, typically a buffer is provided for buffering the received streaming content. This avoids pauses or gaps in the presentation of the content to the user which may otherwise occur due to delays in the delivery of the content to the terminal. Due to limited storage resources in the mobile terminal, a buffer typically has a maximum allocatable size which is sufficient to temporarily store, e.g., few seconds of a video stream.

Thus at the client upon reception of data, said data in form of packets or frames is stored in a buffer in order to compensate variations of transmission delay or transmission rate. In the next step, the packets are taken from the buffer for the play out.

Usually the media data contains of a large amount of data. In order to save network resources the media data is compressed which effectively reduce the bandwidth required to transmit thereof. Dependent of the data, like audio video, image, different compression methods are applicable.

In case of video, a video compression schema reduces the quantity of data used to represent digital video images by combining image compression and motion compensation commonly known as predictive coding techniques.

A frame in a video stream is essentially a picture captured at a predetermined instant in time, the set of frames building the video stream. In typical video coding schemes, such as an MPEG coded video stream, a GOP (Group of Pictures) is a group of successive pictures within the video stream. Each MPEG coded video stream consists of successive GOPs. A GOP can contain the following frame types:

I frame (intra-coded frame): a frame corresponding to a fixed image which is independent of other frames. Each GOP begins with this type of frame.

P frame (predictive coded frame): contains motion compensated difference information relative to previously coded frames. Normally, P frames need much less storing space than I frames.

B frame (bidirectional predictive coded frame): contains motion compensated difference information relative to previously coded frames, where bi-directional interpolated prediction can be used. Normally, B frames need less storing space than I frames or P frames.

Video data being compressed at the sender side is transmitted via a communication network and is decompressed at the receiver side before the data is forwarded for presenting.

However during the transmission often packet loss occurs with the result that not all packets belonging to a media stream are available at the client buffer. Taking the GOP example, I-picture might be decoded since they have no dependencies on previously encoded pictures. In contrast, the P-pictures can be encoded using previously encoded I or P pictures as references, consequently in case of packet loss, an error free play out of the video stream is not possible anymore.

Thus, due to the predictive coding techniques, a loss of a video packet or video frame has also influence on the quality of the succeeding video frames of the stream, so that a video decoding error is propagated throughout the stream. This error propagation stops when the next I-frame is received at the client since an I-frame is not dependent on the preceding video frames.

There are different error control mechanisms used for reconstructing errors during a transmission. Some examples are forward error correction FEC, automatic retransmission request ARQ or error concealment method.

The error concealment methods are used at the decoder to reconstruct the lost video frames. A simple concealment technique is a repetition of the last successful received video frame before the lost frame during the play out. Additionally, motion information from the neighboring video frames could be used to make a more accurate reconstruction of the missing frame, and if only parts of a video frame are lost, then correctly received neighboring blocks could be additionally utilized to help compensating for the lost parts.

However, the error concealment techniques have their limitations. Normally, it is not possible to eliminate the error introduced by the missing video frames completely. Due to error propagation, these errors are propagated throughout the decoded video sequence until the next I-frame is received and decoded. These errors could reduce the overall video quality quite heavily. For example, if a P-frame is missed at the beginning of a GOP, the error is propagated throughout the complete GOP. This results in presenting the media data to the user with degraded quality or with interruptions.

Summary

There is a demand for a technique for processing media data before presenting said media data at the receiver side. In particular there is a demand to increase the experience level of the presented quality.

The invention is embodied in independent claims. Advantageous embodiments are described in the dependent claims.

The demand is satisfied with a method for processing media data at a receiver device in a packet-switched network. The method comprises the steps of receiving media data comprising sequence of data packets. Further said method comprises the steps of detecting data packet losses in the media data, and determining data frames within a group of successive frames of the media stream affected by a packet loss. Preferably a frame comprises at least one data packet. In the next step it is decided to drop data frames as a function of a threshold value for quality degradation of the presented media data and the determined data frames affected by the packet loss, and finally the dropping of the data frames based on the decision is performed.

The abovementioned demand is also satisfied by a device for processing media data on a receiver device in a packet-switched network. It is proposed that the device comprises a receiver entity for receiving media data comprising sequence of data packets. Further it is proposed to provide a detector for detecting data packet losses in the media data and a processor for determining data frames within a group of successive frames of the media stream affected by a packet loss. It is to be mentioned that a data frame comprises at least one data packet. Consequently, a packet loss of one data packet or more than one packet may affect a data frame. A decision unit is adapted to decide to drop data frames. The decision is based on an estimation being a function of a threshold value for quality degradation of presented media data and the determined data frames affected by the data packet loss. After a decision is taken, a dropping unit is adapted to drop the determined data frames affected by the packet loss are dropped. The receiver entity, the detector, the decision unit and the dropping unit are connected with each other in a way that allows exchange of information required to perform the embodiments of the present invention.

Further the device node is adapted to perform all steps as claimed in connection with the method which is to be performed in said node.

Brief Description of the Drawings

In the following, the invention will further be described with reference to exemplary embodiments illustrated in the figures, in which:

Fig. 1 is a flow diagram exemplarily illustrating an operation of the embodiment of the invention performed in the receiver device;

Fig. 2 schematically illustrates embodiments of the present invention;

Fig.3 schematically illustrates an embodiment of channel switching using a interleaving approach;

Fig. 4 schematically illustrates functional components of the receiver device in a network.

Detailed Description of Preferred Embodiments

In the following description, for purposes of explanation and not limitation, specific details are set forth, such as particular network environments and communication standards etc., in order to provide a thorough understanding of the current invention. It will be apparent to one skilled in the art that the current invention may be practiced in other embodiments that depart from these specific details. For example, the skilled artisan will appreciate that the current invention may be practised with any wireless network like for example UMTS, GSM or LTE networks. As another example, the invention may also be implemented in short-range wireless networks such as WLAN or Bluetooth systems or in wireline networks, for example in any IP-based networks, like IMS network.

The invention may be practiced with certain (TV) broadcast networks or with hybrid networks comprising a (TV) broadcast network and a mobile network, for example a DVB-H (Digital Video Broadcast-Handhelds) and a 3GPP mobile network. Basically, the invention may be practiced within any network environment in which video content may be distributed.

The data frame might be a predictive coded frame or an independent coded frmae. The predictive coded frames are frames that cannot be decoded independently. Accordingly, the predicative coded frames may be constituted by at least one of P-frames and/or B-frames. The independent coded frames are preferably frames such as I-frames that can be decoded independently.

The media data may comprise video data, audio data, or any other kind of (multi)media data, such as, for example, a combination of video and audio data. The content may be provided within the framework of a multimedia service such as a mobile TV or IPTV service.

The packet loss may occur due to loosing a data packet during a transmission. However it may be also a result of defecting data packets, for example during the transmission between the sender and the receiver.

The threshold value describes a level of quality degradation of the media data which is acceptable by the receiver. Thus, the threshold value might be a value which describes that if a measure measuring the affected frames is below the threshold value that it might be preferred to display all frames and to reconstruct the affected predictive frames, in order to fulfill the user's experience of the presented quality. However if the measure of the determined affected frames is equal or higher than the threshold value, then it is preferably to display only the independent coded frames and to drop the predictive coded frames in order to present a slide show to the user which is experienced as a better quality then a video with interruption or pictures with bad quality. In a preferred embodiment the threshold value corresponds to a number of affected data frames.

In one preferred embodiment the GOP structure is used as an example for the realisation of the group of successive frames of the media stream. The group of successive frames of the media stream may be a sequence of data packets/frames in which the media data is subdivided. In particular the group of successive frames of the media stream comprises data frames which are predictive encoded and data frames which are independent encoded.

In the following an embodiment of the present invention is presented in respect to Fig. 1 showing a flow chart with steps according to the present invention.

In the first step, S10, the data to be processed is received. The receiver device may be any device receiving the media data. Thus it might be a user device providing the data for presenting to the user. Further it may be a proxy, which performs the dropping of frames. In this case the quality of media data is enhanced and additionally the amount of data which is transmitted to the next hop is reduced leading to an efficient using of network resources.

In the next step, S11, it is detected whether a packet loss occurred and whether data frames are affected by the packet loss. There are different methods applicable for determining a packet loss. One embodiment might be that a video decoder by reconstructing the stream of pictures recognizes lost data packets. A data frame may comprise at least one data packet, depending on the frame size, usually it consists of a number of data packets. Consequently, loosing one data packets or a number of data packets may lead that only one frame is affected if all of the data packets belong to said data frame.

In a further embodiment a buffer receiving and buffering data packets before forwarding them for play out may be adapted to detect loosing of data packets and to provide a corresponding report to the video decoder. Preferably the data packets and the frames are numbered for the purpose of recognizing loses. Thus, it is possible to detect looses. However, preferably the video decoder takes the decision whether a frame is affected or not.

Returning to Fig.1, in the next step, S12, the affected data frames are to be determined. There are different possible embodiments for implementing this step. In the following some preferable examples are given.

The step of determining the data frames affected by the packet may comprise assessing a number of the affected data frames. When considering the GOP structure it is proposed to count how many of data frames are affected by a packet loss.

Further the step of determining the data frames affected by the packet may comprise the step of assessing what kind of data frames is affected. For example if an independent coded frame like an I-frame is affected, than it is proposed to consider as well said I-frame as affected as also all predictive coded frame, P-frames or B-frames being dependent on this I-frame.

According to a further embodiment for realisation the step of determining it is proposed to assess a position of the data frame affected by the packet loss within a group of successive frames of the media stream. Assuming, that the affected data frames are predictive coded farmes, usually those predictive coded frames which are at the end of the group of successive frames do not influence the dropping decision since an affected predictive coded frame just before an independent coded frame does not lead to error propagation. Thus, it is proposed in this embodiment not to drop this kind of P-frames even if the number of the affected P-frames exceeds the threshold value.

In a further embodiment for realisation the step of determining it is proposed to assess the amount of intra blocks in a data frame affected by the packet loss within a group of successive frames of the media stream. Video compression typically operates on square-shaped groups of neighboring pixels, often called blocks. Within a data frame, there may be intra blocks of pixels which are coded independent of other blocks, wherein the number of such blocks influence the independent decode-ability of a data frame. In this case it is proposed that the number of intra blocks influence the decision of dropping the data frames in a way that preferably no dropping of any data frames is performed even if the number of the affected data frames exceeds the threshold value.

As a further embodiment for realisation the step of determining it is proposed determine the affected data frames based on the size of the data frame affected by the packet loss. In case of a large data packet, for example in case of a video frame encoded with a high bit rate, the loss of said data packet is rated as more severe. In case of video data a loss of a large data packet may cause a cut of a whole scene within this video frame. In this case it is proposed that the size of the affected data frames influence the dropping decision in a way that number of affected data frames is reduced or the threshold value is increased. When considering the GOP structure, this would mean that the number of affected data frames, like for example P-frames needed for deciding on dropping said frames within considered GOP is reduced.

In case of a small lost packet, like for example a video frame encoded with a lower bit rate, the loss case is rated as less severe since this may mean that no major changes of the video content has happened. In this case, the number of affected data frames within a GOP which is to be compared with the threshold value for quality degradation is increased.

Preferably the determination step based on the size of the data frame affected by the packet loss is preceded by a step of evaluating the size of the data frame affected by the packet loss. In some embodiments the size of the lost packets is not directly known at the receiver side. However, there are techniques available for determining these values at the receiver. For example, a message called RTCP (Real Time Transport Control Protocol) Sender Report (SR) sent from the server to the receiver in regular time intervals includes the number of octets or amount of data in bytes as well as the number of data packets transmitted by the server. Using this information and comparing it with the number of indeed received octets and packets allows the determination of the sizes of lost packets.

The determination of the threshold value describing the acceptance of quality degradation of the media data might be performed in any suitable and preferable way. In one embodiment it is proposed to provide this threshold value as a result of a priori performed measurements. The threshold value might be an absolute value or a percental value. It may also be a dynamic value being provide-able to the receiver during a transmission.

Now returning to Fig.1, in the next step, S13, a decision is to be taken whether to drop some data frames. In the following some embodiments for implementing this step are presented.

In a preferred embodiment for the step of deciding on dropping data frames it is proposed to compare the determined data frames affected by the packet loss with the threshold value. As described in the abovementioned embodiments the determination step has an influence on the dropping decision.

According to one embodiment of the present invention it is proposed that the data frames might be affected directly or indirectly by a packet loss. It may be that at least one packet in a data frame is deemed as lost, for example due to packet loosing or packet defecting, in this case a data frame is affected directly. In contrast thereto, it may occur that there are data frames being determined for dropping although they are not affected directly. For example when there is some dependency between these frames and the data frames being directly affected by a packet loss. In the frame of the decision step, the decision is taken which data frames are to be dropped considering the number of either directly or/and indirectly affected data frames.

According to one embodiment the number of affected data frames is compared with the threshold value. In one embodiment it is proposed that if the number of affected predictive coded frames, like for example the P-frames, is equal or above the threshold to drop all predictive coded frames located between independent coded frames, S14 and S15.

In another embodiment it is proposed to decide to drop only a set of predictive coded frames in the group of successive frames. This case may occur when for example some predictive coded frames are received correctly and the following frames in a GOP are affected by a packet loss. Herein it might be decided to play out the correctly received frames and to drop the affected frames. Preferably the decision is taken, when in the following GOP the predictive coded frames are received correctly. The dropping of the affected frames would lead to freezing the picture until playing out the next GOP.

According to a further embodiment, the step of deciding on dropping data frames S13 may comprise a step of considering predictive coded frames affected by the packet loss and located in a further group of successive frames. In this embodiment it is proposed to check also further group of successive frames, like in a preferred embodiment the neighbouring group of successive frames and to base the dropping decision on the outcome. For example when considering the GOP structure if in the neighbouring GOP all P-frames have been received correctly and in the considered GOP some P-frames are missing at the end of the GOP structure, then preferably it is to decide to not to drop the P-frames, S14, S16, even if the number of the determined P-frames is higher then the threshold value, then P-frames at the end of GOP do not propagate errors. Adequately, if the number of the affected P-frames in a neighboured GOP is high but in the considered GOP the number of affected P-frames is below the threshold value, then it may be decided to drop all P-Frames, S14, S15, in the considered GOP in order to generate a constantly slide show effect.

In a further embodiment, the step of deciding which predictive coded frames are to be dropped, S13 comprises a step of dropping predictive coded frames in a prediction hierarchy level and all other predictive coded frames depending thereon. In general, hierarchical predicted coding is one of the existing ways of coding video. Video data contains spatial and temporal redundancy. Similarities can thus be encoded by merely registering differences within a frame (intra-frame prediction) and/or between frames (inter-frame prediction). This embodiment refers to the temporal coding. In general it might be said that there are key pictures on a first layer (first temporal prediction hierarchy level) and from these key pictures, frames in a next layer (second temporal prediction hierarchy level) are predicted. Thus in case of a hierarchical prediction structure if the number of affected predicted frames of a certain temporal prediction hierarchy level exceeds a threshold, all frames of this prediction hierarchy level may be dropped, S15, as well as all other frames in the GOP that depend on those dropped frames are dropped.

In a further embodiment it is proposed to take by the dropping decision, S13, into consideration the predictive coded frames of a first layer in a scalable video coding structure and all other predictive coded frames depending thereon. Scalable media is typically ordered into hierarchical layers of data, where a video signal can be encoded into a base layer and one or more enhancement layers. A base layer can contain an individual representation of a coded media stream such as a video sequence. Enhancement layers can contain additional data relative to previous layers in the layer hierarchy. The quality of the decoded media stream improves as enhancement layers are added to the base layer. An enhancement layer enhances the temporal resolution (i.e., the frame rate), the spatial resolution (e.g. resolution of a frame), or the fidelity (quality) (e.g. by reducing noise in the frame). Each layer, together with all of its dependent layers, is one representation of the video signal at a certain spatial resolution, temporal resolution and/or quality level. Therefore, the term scalable coding structure is used herein to describe a scalable layer together with all of its dependent layers. Scalability has already been present in the video coding standards MPEG-2 Video, H.263, and MPEG-4 Visual. According to the present embodiment of a scalable video codec with a layered coding structure it is proposed that if the number of affected predictive frames of a certain layer exceeds a threshold, all frames of this layer as well as all frames of dependent layers within the certain GOP are dropped.

In step S15 in Fig. 1 it is proposed to drop the data frames being determined as affected and on which the decision has been taken to drop them.

In one embodiment it is proposed to drop the determined data frames and to show the the non-dropped frames. The presentation of some of the data frames to the users results in a presentation of a slide show. Thus instead of showing video with as badly experienced quality, a slide show of pictures is presented to the user. In one embodiment it is proposed to drop all predictive coded frames and to display the independent coded frames. Additionally or/and alternatively it may be decided to provide predictive frames for presenting. This may be the case when for example temporal prediction hierarchies are used. Assuming that for example, a number of frames corresponding to a Nth temporal prediction hierarchy level are not available, in that case it may be decided to drop the frames in hierarchy level N and hierarchy levels greater than N (i.e. frames that depend on those in level N), and to present the frames that correspond to hierarchy levels lower than N.

In step S16 in Fig.1, it is proposed not to drop any affected data packets but to try to reconstruct said data packets using any the preferably reconstructions method, like for example the error concealment method.

In the following some further embodiments of the present invention are presented.

In Fig.2a, the embodiment of dropping all predictive coded frames and displaying the independent coded frames is depicted. The receiver, Receiver, receives a group of successive data frames of the media stream, GOP00, comprising the data frames, I_{00}, P_{01}, P_{02}, P_{03}, P_{04} and I_{11}. During the determination step for determining affected data frames, it is recognised that the predictive coded frame, P_{02} is affected. In order to avoid the propagation of the errors it is decided to drop all P-frames within this GOP and to display only the I-frames, thus the frames I_{00} and I_{11} are displayed, Display. The step of displaying only independent coded frames, I-frame result in a slide show presented to the user. The block artifacts in the complete GOP due to error propagation are avoided resulting in a better visual quality of the presented video stream. Thus according to this embodiment it is proposed not to display the predictive frames independent whether they have been received error free or whether they are affected by a packet loss.

In a further embodiment, presented in respect to Fig.2b, the case is shown, where an independent coded frame, I-frame I_{10} in a a group of successive data frames, GOP11, is affected. In this case, consequently all data frames dependent on the I-frame, P_{11} P_{14}, are also seen as affected. Herein, it is proposed to play out the last correct received P-frame, P_{07} of the preceding received GOP, GOP01 and to drop the affected data frames. The next data frame which will be played out would be the correctly received frame of a next GOP, I_{21}. Thus, for presenting at the Display, the data frames P_{07} and I_{21} are provided.

In the following some examples are presented, in which the present invention may be applied.

In the so called thinning case, it is proposed to drop packets at the server if the link bandwidth is not high enough for the transmission of all packets. In particular it is proposed to drop these data packets which do not affect the video quality very much, thus for example P-frames at the end of a GOP . However, if additional packet loss occurs during the transmission the situation can change and a high video quality loss could happen. In this situation, the above described embodiments of the present invention may be used to increase the video quality. In this case it is proposed to present slide show instead of video with not acceptable quality.

In the following a further embodiment of the present invention by implementing fast channel switching is presented in respect to Fig.3. This figure illustrates a channel switching into a new channel, wherein the channel switch happens into GOP0 at time point, sw. GOP1 is the following GOP sequence in the new channel.

Fast channel switching is a general term describing possibility to switch fast between channels, like for example TV channels. In particular the demand for providing fast channel switching is an issue in wireless communication networks, since the users expect the same switching experience as when watching TV at home.

There are different approaches realising the fast channel switching. One of the approaches is based on the interleaving in a GOP structure. As already discussed, the GOP always begins with an I-frame and afterwards several P-frames follow, in each case with some frames distance. As above mentioned the I-frames contain the full image, they do not require any additional information to reconstruct the image. Therefore any errors in the streams are corrected by the next I-frame (an error in the I-frame propagates until the next I-frame). Errors in the P-frames could propagate until the next I-frame. The more I-frames the MPEG stream has, the more it is reliable. However, increasing the number of I-frames increases the stream size.

With the interleaving, it is proposed to reorganise the video data in a manner that enables decoding and display of an independent picture by placing it at the end of a GOP while transmitting. Thus, the order of video frames within a GOP is changed during the transmission; in particular it is proposed that the I-frame appears last in a GOP.

Fig.3 presents the interleaving approach while applying a preferred embodiment. In Fig.3a , the coding step of a GOP structure GOP0 at the coder side is presented. Fig.3b depicts the sequence of frame transmission. In Fig.3c, the way of displaying the received frames is depicted.

In Fig.3a, the way of coding of GOP0 at the sender/coder side is shown. As already explained, an I-frame, I_{01} is independent encoded and contains the whole information of a picture. The following P-frames, P_{01} to P_{07}, are predictive encoded. The sequence of coding is that at first the I-frame is coded and then the P-frames. This is also the usual sequence of transmitting the frames. In contrast thereto, in the interleaving approach it is proposed to send at first the predictive coded frames followed by the I-frame. In Fig.3b, the interleaved transmission of the GOP structure is depicted.

Coming now to the receiver side, Fig.3c, assuming that frames P_{01} to P_{03} are received correctly and a channel switch is performed at the time point sw; in the interleaving case, this would mean that the I_{00} frame and the P_{01} to P_{03} frames would be presented and after the channel switch, sw, a break in presenting would occur until the I_{10} pictures of GOP1 of the new channel are taken from a buffer for play out. In order to overcome this disadvantage, it is proposed, according to the present embodiment, to drop the frames P_{01} to P_{03} and to present only the I_{00} frame and as next the frame I_{10}

as slides. In respect to the embodiment of the present invention, the frames P_{01} to P_{03} are not directly but indirectly affected by the packet loss since the following frames P_{04} to P_{07} are affected in a sense that they will not be presented. The dropping of the packet is depicted with the aid of the dashed lines. Compared to the case without dropping, where a short video sequence followed by a large gap is shown, the advantage of the present embodiment is an increase of the subjective video quality.

Fig. 4 schematically illustrates a packet switched network 400 with a receiver device 40. The packet switched network 400 may be an UMTS network and the receiver device 40 may be mobile devices or any proxy performing the method in the packet switched network. The receiver device either provides the end result directly to the user interface or the reduced number of packets is provided from the proxy to the link in order to reduce the amount of data of the media stream which is to be transmitted to the next hop.

According to Fig.4 the receiver device 41 has a receiver unit which is adapted to receive the media data. The media data is provided to the detector 42 which is adapted to detect a packet loss in the received data and to determine the affected data frames. In a preferred embodiment the video decoder is responsible for detecting the lost data packets. A frame may comprise at least one data packet, depending on the frame size, usually it consists of a number of data packets. Consequently, loosing one data packets or a number of data packets may lead that only one frame is affected if all of the data packets belong to said frame.

In a further embodiment a buffer receiving data packets may be adapted to detect loosing of data packets and to provide a corresponding report to the video decoder. Preferably the data packets and the frames are numbered for the purpose of recognizing loses. Thus, it is possible to detect looses. However, preferably the video decoder takes the decision whether a data frame is affected or not.

In the next step, the detector provides corresponding information to the decision unit 43 which decides whether to drop the directly or indirectly affected data frames and to display only the data frames, which are not considered as affected or to apply error concealment on the affected frames. In case the dropping decision is taken, the dropping unit 44 performs the dropping of the data frames and a provision unit, not shown in the figure, provides the non dropped data frames to the end device, which might be an end user or the next hop.

While the current invention has been described in relation to its preferred embodiments, it is to be understood that this description is for illustrative purposes only. Accordingly, it is intended that the invention be limited only by the scope of the claims appended hereto.

Claims

1. A method for processing media data on a receiver device in a packet-switched network, the method comprising the steps of

- receiving media data comprising sequence of data packets,
- detecting data packet losses in the media data,
- determining data frames within a group of successive data frames of the media stream affected by a packet loss wherein a data frame comprises at least one data packet,
- deciding to drop data frames as a function of a threshold value for quality degradation of presented media data and the determined data frames affected by the data packet loss, and
- dropping the data frames based on the decision.

2. The method according to claim 1,
wherein the step of determining comprises assessing a number of data frames affected by the packet loss within a group of successive frames of the media stream.

3. The method according to claim 1,
wherein the step of determining comprises assessing a position of the data frame affected by the packet loss within a group of successive frames of the media stream.

4. The method according to claim 1,
wherein the step of determining comprises assessing amount of intra blocks in the data frame affected by the packet loss within a group of successive frames of the media stream.

5. The method according to claim 1,
wherein the step of determining affected data frames is based on the size of the data frame affected by the packet loss.

6. The method according to claim 5,
wherein the determination step based on the size of the data frame affected by the packet loss is preceded by a step of evaluating the size of the data frame affected by the packet loss.

7. The method according to one of the preceding claims 1 to 6 wherein the level of acceptance of quality degradation is a threshold value provided as a measure, representing a number of data frames which may be accepted as lost.

8. The method according to one of the preceding claims 1 to 7 wherein the step of deciding on dropping data frames comprises a step of comparing the determined data frames affected by the packet loss with the threshold value.

9. The method according to claim 1 wherein the step of dropping data frames comprises a step of dropping all predictive coded frames located between independent coded frames.

10. The method according to claim 1 wherein the step of dropping data frames comprises a step of dropping a set of predictive coded frames in the group of successive frames.

11. The method according to claim 1, 9 or 10 wherein the step of dropping data frames comprises a step of considering predictive coded frames affected by the packet loss and located in a further group of successive frames.

12. The method according to one of the claims 1, 9, 10 or 11 wherein the step of dropping predictive coded frames comprises a step of dropping predictive coded frames in a prediction hierarchy level and all other predictive coded frames depending thereon.

13. The method according to one of the claims 1, 9, 10 or 11 wherein the step of dropping predictive coded frames comprises a step of dropping predictive coded frames of a layer in a layered coding structure and all other predictive coded frames depending thereon.

14. The method according to claim 1, 9 or 10 wherein the step of dropping data frames comprises a step of dropping at least one independent coded frame.

15. The method according to one of the claims 1-14

wherein the step of dropping predictive coded frames is followed by the step of providing not-dropped data frames for presenting.

16. The method according to claim 15,
wherein the not-dropped data frames are independent coded frames.

17. The method according to claim 15
wherein the not-dropped data frames are predicted coded frames which have not been dropped.

18. A device for processing media data on a receiver device in a packet-switched network, the device comprises

- receiver entity for receiving media data comprising sequence of data packets, and
- detector for detecting data packet losses in the media data, and
- processor for determining data frames within a group of successive frames of the media stream affected by a packet loss wherein a data frame comprises at least one data packet, and
- decision unit for deciding to drop data frames as a function of a threshold value for quality degradation of presented media data and the determined data frames affected by the data packet loss, and
- dropping unit for dropping data frames based on the decision.

19. A device according to claim 18 wherein
further a provision unit (35) is provided which is adapted to provide non-dropped frames for presenting on the user device.

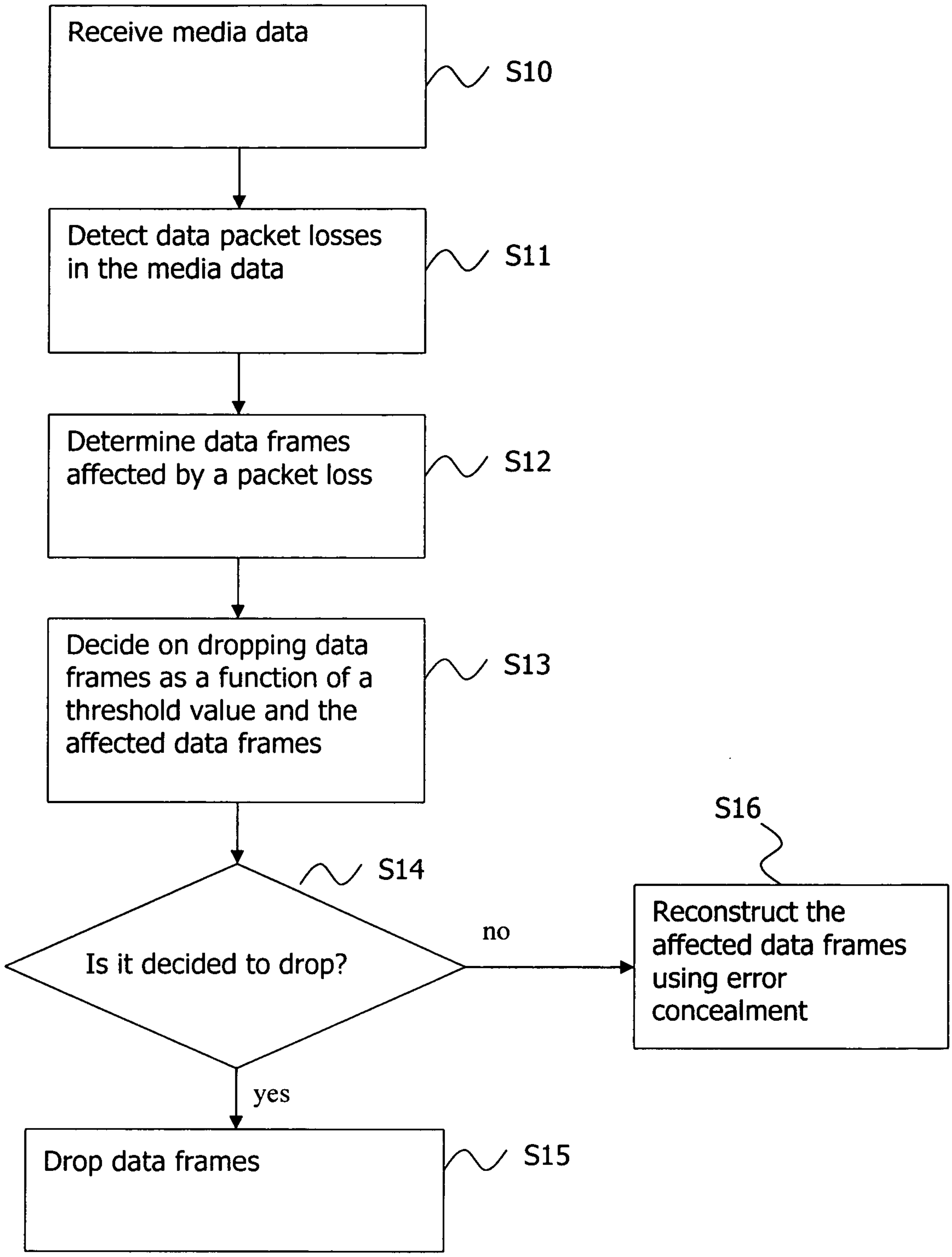
Receive media data
S10
Detect data packet losses in the media data
S11
Determine data frames affected by a packet loss
S12
Decide on dropping data frames as a function of a threshold value and the affected data frames
S13
S16
S14
Is it decided to drop?
no
Reconstruct the affected data frames using error concealment
yes
Drop data frames
S15

Fig.1

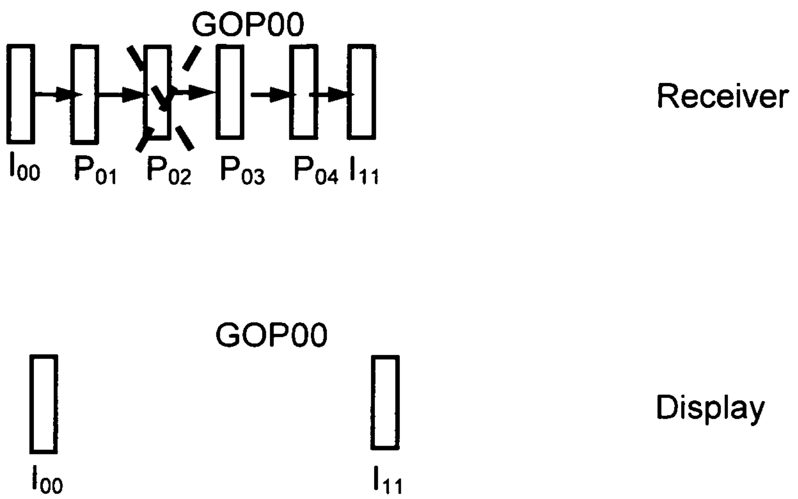
GOP00
Receiver
I00 P01 P02 P03 P04 I11
GOP00
Display
I00
I11

Fig.2a

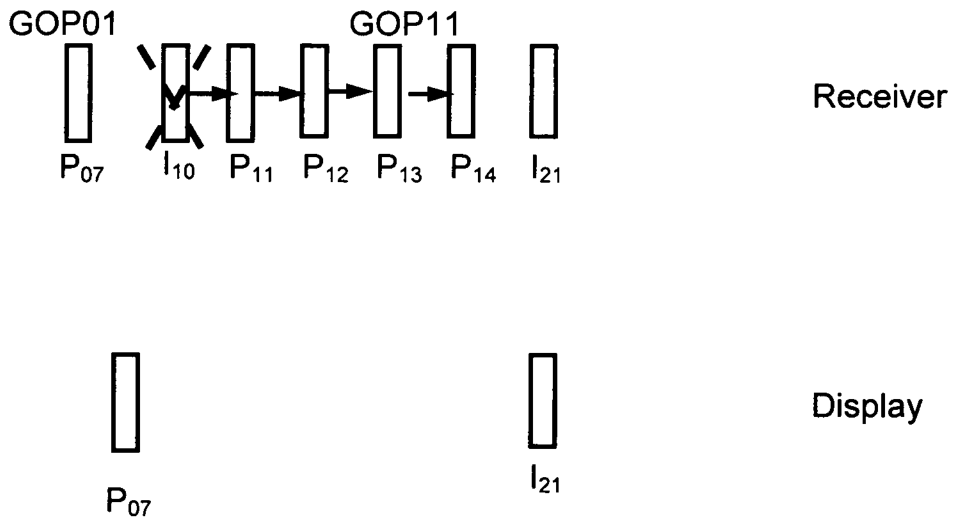
GOP01
GOP11
Receiver
P07 I10 P11 P12 P13 P14 I21
Display
P07
I21

Fig.2b

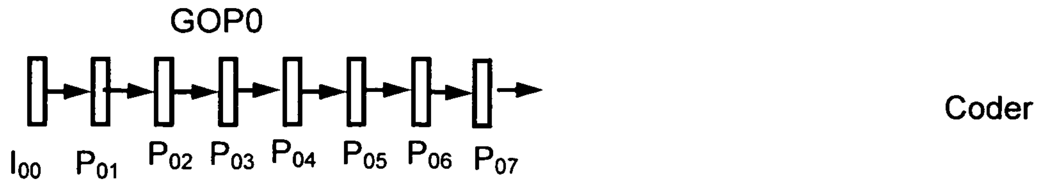

Fig.3a

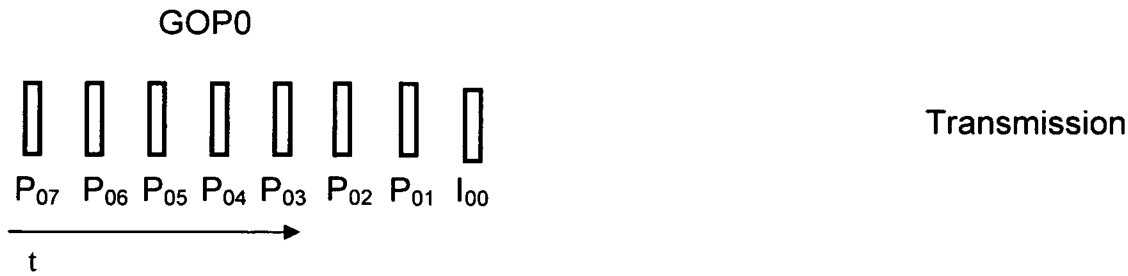

Fig.3b

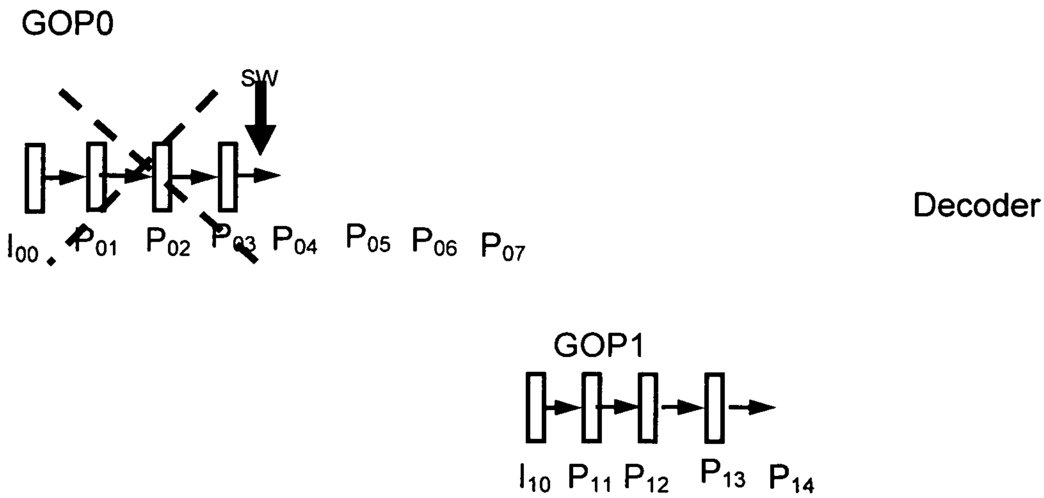

Fig.3c

Fig.3

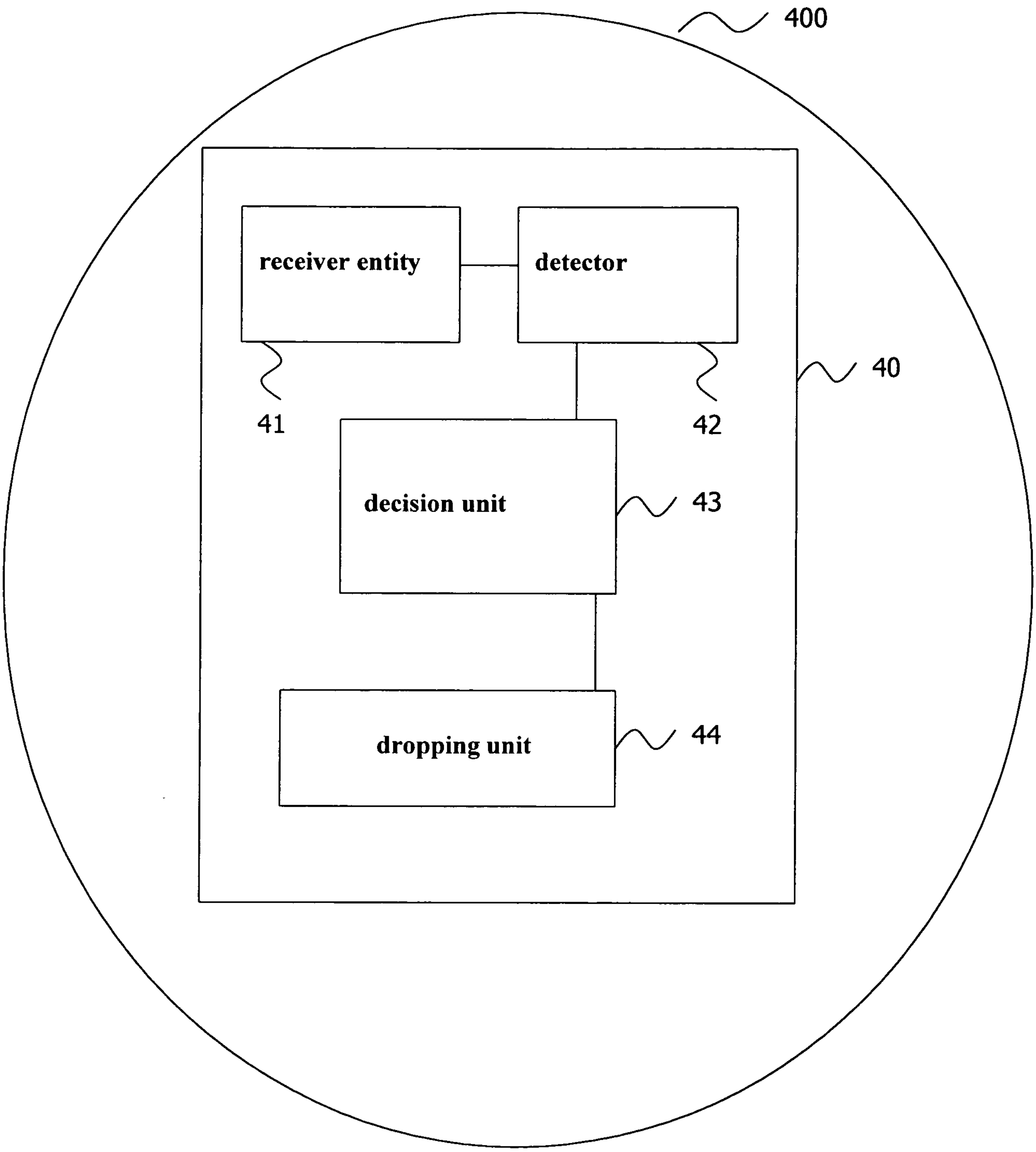
400
receiver entity
detector
41
42
40
decision unit
43
dropping unit
44

Fig.4

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/001526

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/64

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

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.
X Y	US 2007/291836 A1 (SHI FANG [US] ET AL) 20 December 2007 (2007-12-20) [0008], [0010], [0011], [0041], [0045], [0059]-[0061], [0067], [0078], [0080], [0081], [0083], [0088], [0092], [0103], [0104], [0106]-[0108] figures 3a, 4, 5, 7, 8 -----	1-3,7-19 4-6
Y	WO 2006/075302 A1 (KONINKL PHILIPS ELECTRONICS NV [NL]; VAN DER STOK PETRUS D V [NL]; STE) 20 July 2006 (2006-07-20) [0021], [0031], [0034] -----	4-6
A	WO 2008/112441 A1 (MICROSOFT CORP [US]) 18 September 2008 (2008-09-18) [0003], [0004], [0007], [0019], [0020], [0041], [0042] ----- -/--	1-19

[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
2 December 2009	11/12/2009

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, Fax: (+31-70) 340-3016	La, Valérie

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/001526

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 1 137 005 A2 (IBM [US]) 26 September 2001 (2001-09-26) [0007], [0011], [0015]	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2009/001526

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 2007291836	A1	20-12-2007	CA	2645014 A1	29-05-2008
			EP	2080384 A2	22-07-2009
			JP	2009537081 T	22-10-2009
			KR	20090006163 A	14-01-2009
			WO	2008063687 A2	29-05-2008
WO 2006075302	A1	20-07-2006	CN	101107865 A	16-01-2008
			JP	2008527889 T	24-07-2008
			KR	20070104608 A	26-10-2007
			US	2008137728 A1	12-06-2008
WO 2008112441	A1	18-09-2008	US	2008225735 A1	18-09-2008
EP 1137005	A2	26-09-2001	JP	4287068 B2	01-07-2009
			JP	2001346205 A	14-12-2001
			US	6662329 B1	09-12-2003