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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 March 2008 (13.03.2008)

PCT

(10) International Publication Number
WO 2008/028334 A1

(51) International Patent Classification:
G06T 3/40 (2006.01)

(21) International Application Number:
PCT/CN2006/002261

(22) International Filing Date:
1 September 2006 (01.09.2006)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **THOMSON BROADBAND R & D (BEIJING) CO., LTD.** [CN/CN]; 8th Floor, Building A, Technology Fortune Center, No. 8 Xueqing Road, Haidian District, Beijing 100085 (CN).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **CHEN, Zhibo** [CN/CN]; 8th Floor, Building A, Technology Fortune Center, No. 8 Xueqing Road, Haidian District, Beijing 100085 (CN). **GU, Xiaodong** [CN/CN]; 8th Floor, Building A, Technology Fortune Center, No. 8 Xueqing Road, Haidian District, Beijing 100085 (CN). **CHEN, Quqing** [CN/CN]; 8th Floor, Building A, Technology Fortune Center, No. 8 Xueqing Road, Haidian District, Beijing 100085 (CN).

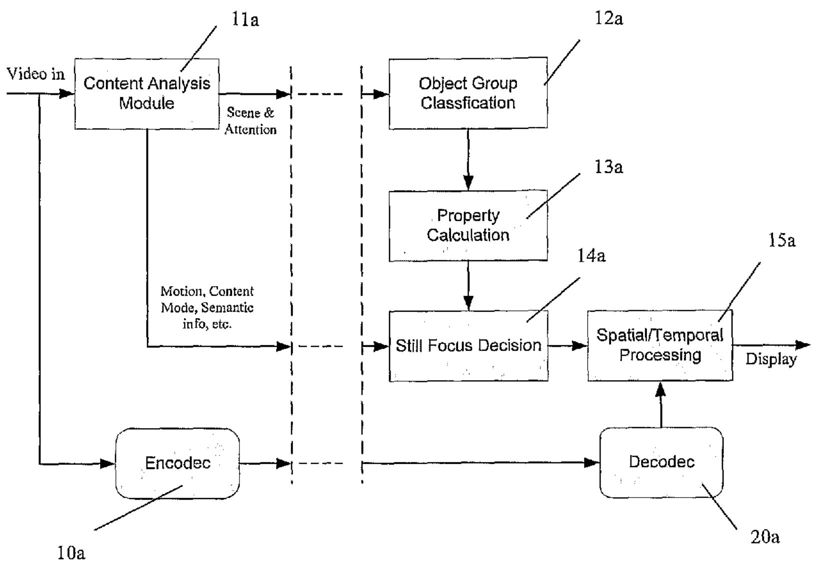
(74) Agent: **YU, Gang**; Kangxin Partners, P.C., Floor 16, Tower A, InDo Building, A48 Zhichun Road, Haidian District, Beijing 100098 (CN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, 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, HU, IE, IS, IT, LT, LU, LV, MC, NL, 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:** METHOD AND DEVICE FOR ADAPTIVE VIDEO PRESENTATION



(57) **Abstract:** The present invention relates to an adaptive video presentation method for automatically presenting a video with stream-embed information based on content analysis of the video on a display with a limited screen size. The method comprises steps of determining a salient object group containing at least one salient object based on perceptual interest value of macroblocks for each frame of said video, extracting a window having a minimum size containing the salient object group for a scene of the video, characterized in that it further comprises steps of comparing size of the extracted window with the display size; and presenting at least a selected area of the extracted window containing at least a part of the salient object group for the scene on the display in different operation modes based on the result of the comparison steps for different motion status for the scene of the video.

WO 2008/028334 A1

METHOD AND DEVICE FOR ADAPTIVE VIDEO PRESENTATION

FIELD OF THE INVENTION

The present invention relates to a method and a device for video presentation, and more particularly to a method and a device for adaptive video presentation on a display with limited screen size.

BACKGROUND OF THE INVENTION

In the multimedia application area, a variety of new mobile devices, such as Pocket PC, Smartphone, SPOT watch, Tablet PC, personal digital assistant devices, etc, are with the growing popularity in people's daily life. These devices are becoming more and more powerful in both numerical computing and data storage. Moreover, people have become enthusiastic to watch videos through these mobile devices.

However, low bandwidth connections and small displays are still the two serious obstacles that have under minded the usefulness of these devices in people's daily life. Though a few commercial video players such as Windows Media Player and PocketTV have been developed to enable users browse videos from the small-form factor devices, the limited bandwidth and small window sizes remain to be two critical obstacles. With the rapid and successful development of 2.5G and 3G wireless networks, the bandwidth factor is expected to less constraint in the near future. While at the same time the limitation on display size is likely to remain unchanged for a certain period of time.

There have been some existed works focused on the topic of display images on mobile devices. They can calculate and provide an optimal image viewing path based on the image attention model to simulate the human viewing behaviours. Since most of the valuable information is presented by videos, improving the experience of video viewing on small displays is very important to unleash the power of these mobile devices

One solution on providing a better user experience for viewing videos on the limited and heterogeneous screen sizes has been proposed by X. Fan et al in "Looking into Video Frames on Small Displays", ACM MM'03, 3003, which introduces three browsing modes: manual browsing method, full-automatic browsing method and semi-automatic browsing method.

However, in the proposed full-automatic browsing method, both direction and zoom controls are disabled. The resulting video stream uses more screen space to display the attention-getting regions while cropping out the other parts. Therefore this approach will have less difference with the conventional down-sampling scheme when video frames contain many separate focuses.

In the semi-automatic browsing method, human interaction is still required to switch the browsing focus when there are more than one important attention objects (AO). Therefore, the display focus was calculated after the user pressed the control button, and the artefact will appear when the focus is changed.

Therefore, the existed schemes couldn't provide a good solution for automatically browsing videos on devices with small display sizes and keeping a better tradeoff between video display quality and display size constraint, especially in multiple focuses cases.

SUMMARY OF THE INVENTION

The present invention provides an adaptive video presentation solution for full-automatically representing videos on display devices with small sizes according to metadata information based on content analysis in order to provide an optimal video viewing experience for users.

According to a first aspect of the present invention, the method comprises steps of determining a salient object group containing at least one salient object based on perceptual interest value of macroblocks for each frame of the video, extracting a window having a minimum size containing the salient object group for a scene of the video, characterized in that it further comprises steps of comparing size of the extracted window with the display size; and presenting at least a selected area of the extracted window containing at least a part of the salient object group for the scene on the display based on the result of the comparison step.

Advantageously, if size of the extracted window is equal or smaller than a predefined percentage of the display size, the extracted window is exposited on the display with an appropriate zoom-in operation; if size of the extracted window is larger than a predefined percentage of the display size and equal or smaller than the display size, the extracted window is exposited directly on the display; and if size of the extracted window is larger than the display size, the extracted window is exposited on the display with an appropriate zoom-out operation.

According to a second aspect of the present invention, the method further comprises steps of calculating a weighted average motion vector length MV_{act} of macroblocks inside frames for the scene of the video, comparing the weighted average motion vector length MV_{act} of the scene with a predefined threshold T_{motion}, determining whether the weighted average motion vector length MV_{act} of the scene is less than the predefined threshold T_{motion}, and presenting the extracted window for the scene of video on the display in a true motion exhibition mode corresponding to a fast motion status in case the average motion vector length MV_{act} of the scene is determined less than the predefined threshold T_{motion}; or else presenting the extracted window for the scene of the video on the display in a normal exhibition mode corresponding to a low motion status in case the average motion vector length MV_{act} of the scene is determined less than the predefined threshold T_{motion}.

Advantageously, the extracted window containing all the salient object groups for whole the scene of the video is presented on the display with a weighted average gravity point of the salient object group for the whole scene defined as a still focus centre of the extracted window, in such a way the true motion exhibition mode can be achieved in the fast motion status.

According to a third aspect of the present invention, the method, in the normal exhibition mode corresponding to the low motion status, further comprises steps of comparing length of scene of the video with a predefined minimum perceptual time, wherein if the length of scene is less than the predefined minimum perceptual time, it further comprises steps of determining number of the salient objects existing in the scene, presenting the extracted window on the display with a gravity flowing show operation in case only one salient object exists, or else presenting the extracted window directly on the display in case multiple salient objects exist, wherein if the length of scene is no less than the predefined minimum perceptual time, it further comprises a step of presenting the extracted window on the display with the gravity flowing show operation.

These, and other aspects, features and advantages of the present invention will be described or become apparent from the following detailed description of the preferred embodiments, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

Fig. 1 is a schematic view of a first embodiment of the system framework using the method in accordance with the present invention;

Fig. 2 is a schematic view of a second embodiment of the system framework using the method in accordance with the present invention;

Fig. 3 is a schematic view of a third embodiment of the system framework using the method in accordance with the present invention;

Fig. 4 is a schematic view of salient objects inside one frame;

Fig. 5 is a schematic view of salient object group inside one frame;

Fig. 6 is a flowchart of adaptive video presentation sample solution.

DETAIL DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention directs to a method and a device of adaptive video presentation (AVP) for better viewing experience with stream-embedded metadata base on content analysis information.

1. Introduction of basic Frameworks of AVP

According to the present invention, three types of AVP frameworks are proposed: decoder end solution, Joint encoder-decoder end solution, and encoder end solution, as shown in Figs. 1 to 3 respectively.

With reference to Fig. 1, the first type of the AVP framework solution leaves all the region processing and display work in the decoder end, where only a pre-analysis module 11a is provided at an encoder end 10a, while other functional blocks are provided at a decoder end 20a. The pre-analysis module 11a includes operations of scene change detection, attention area extraction and content/motion analysis. The other four functional blocks includes an object group classification (OGC) module 12a, which classifies objects/object groups based on the scene and attention mask information from pre-analysis module 11a; a property calculation (PC) module 13a, which calculates statistics (e.g. gravity point, semantic weight, but not limited to) property of each object/object group; a still focus decision (SFD) module 14a which decides candidate focus area in special image based on the statistics information derived from PC module (e.g. gravity points) 13a and other metadata information from pre-analysis module 11; a spatial-temporal processing module 15a, which does a spatial-temporal processing to guarantee the video smooth and acceptable and eliminates the artifacts.

Similarly, in the second type of AVP framework solution as illustrated in Fig. 2, a pre-analysis module 11b, an object group classification module 12b, a property calculation module 13b and a still focus decision module 14b are included in the encoder end 10b to generate candidate focus area, and a spatial/temporal processing module 15b is included in a decoder end 20b to do optimal display based on candidate focus area with consideration of temporal and spatial quality tradeoff.

In the third type of AVP framework solution as illustrated in Fig. 3, all the modules of pre-analysis 11c, object group classification 12c, property calculation 13c, still focus decision 14c and spatial/temporal processing 15c are included in an encoder end 10c to generate the last display are of each frame in a video sequence, an the metadata describing the suitable display area of each frame can be send as embed metadata to a decoder end 20c to help the adaptive display directly.

2. Introduction of Basic Definitions for Adaptive Video Presentation

To optimize the user's viewing experience on displays with small size display windows, a set of attention area blocks in video frames should be firstly considered to be displayed because the relative information have more positive contributions to the user's viewing experience. Some basic definitions for AVP are described as below.

A. Salient Object

A Salient Object is a set of attention area MacroBlocks (MB) connected each other, as shown by area of MBs with grey in Fig. 4. The salient objects are separated by the non-attention MBs, which is denoted by MBs with white.

A salient object O_i(i=1, 2, 3, …, N) is described by following parameters:

Name	Abbreviation	Description
Size	SZ_i	The number of attention MBs included in the salient object O_i
Contour	CO_i	The minimum rectangle shape which contains the salient object O_i
Rectangle Size of Object	RZO	Number of MacroBlocks included in the minimum rectangle shape contains the salient object Oi, as shown by shadow area in Fig. 4
Object Occupation Ratio	OOR	(Szi/RZO)x100%

B. Salient Object Group

Salient Object group contains all salient Objects in the current frame. It can be described by following parameters:

Name	Abbreviation	Description
Size	SZG	The number of attention MBs included in the salient object group
Contour	COG	The minimum rectangle shape which contains the salient object group
Rectangle Size of Object Group	RZG	Number of MacroBlocks included in the minimum rectangle shape contains the salient object group, as shown by shadow area in Fig. 5
Object Occupation Ratio	OOR	(Szi/RZO)x100%

C. Scene

Video consists of a collection of video frames and the segmentation results can be frame, shot, scene, and video with granularity from small to large. Shot is a sequence of frames recorded in a single-camera operation.

Scene is a collection of consecutive shots that have semantic similarity in object, person, space, and time. It's also defined to tell the switch of salient objects between two frames. The display scheme inside a scene should be definite and usually stays consistent.

One parameter defined for scene is:

Name	Abbreviation	Description
Length Of Scene	LOS	Number of frames in one scene

3. Introduction of Configuration Parameters and Operation Sets

Configuration parameters are necessary parameters on help making decisions of adaptive display mode selection, such as display or not, scaling down or not, summarizing or not, etc. There are four conditions defined to assist the video viewing path programming.

- Minimum Perceptual Time (MPT)

MPT is used as a threshold for the fixation duration when viewing a salient object. It's also used as the threshold of the weight of information value when watching a scene. If a salient object doesn't stay on the screen longer than a MPT threshold MPT_{so}, it may not be perceptible enough to let users catch the information. If a scene doesn't last longer than a threshold MPT_{sc}, only the most significant portion in it may be perceptible enough. MPT_{so} and MPT_{sc} can be selected according to different application scenarios and human visual property, which is usually set to 1/3 second and 2 second in our real application.

- Minimum Perceptual Size (MPS)

The MPS is used as a threshold of the minimum spatial area of a salient object. Normally, if the size of a salient object SO_i is less than a threshold MPT_{so}, the salient object Oi should become marked as non-attention object or be merged into its neighbourhood attention object. But the MPS threshold is not always correct since a salient object with smaller spatial area may carry the most important information, and it cannot be merged or unmarked. So, additional configuration parameter of Weight Information (be introduced below) can be used. Usually MPT_{so} can be set to 5 MacroBlocks or be set to 5%-10% of the largest salient object size.

- Weight of Salient Object (WSO)

This parameter is used to reserve the salient object that carries the most semantic important information and cannot be merged or unmarked. Normally, the value of each salient object's weight is set to 1. If a salient object obtains the most important information in the current and following frames (the frame number should be larger than a threshold T_{weight}, usually $T_{weight}=10$) in a scene, its size is re-calculated as:

$$SO_i=SO_i \times WSO_i$$

Where WSO_i can be defined by the semantic importance of each salient object, which is depend on the content mode, third part's appointed semantic information, specifically user's experience, etc. Furthermore, the gravity of the salient object group is re-calculated.

- Tolerance of Gravity Change (TGC)

Gravity points of either salient object or group are used as the center point of the display, the choosing process will be described in next two sections. IN order to avoiding the dithering effect of the display which caused by the slight change of the center point, the parameter TGC is introduced to ensure that the small change of position (in unit of MarcoBlock) of gravity point will not cause the change of the center point. Two kinds of TGC_H(Tolerance of Gravity Change in the horizontal direction) and TGC_V (Tolerance of Gravity Change in the vertical direction) can be defined according to the relationship between display screen and video frame size.

Operation Set includes all the possible operations needed for the requirement of Adaptive Video Presentation. Currently six operations are defined as Table-I shows:

Table-I Operations Set for Adaptive Video Presentation

Operations		Illustration
Low Motion Exhibition (($MV_{act}>=T_{motion}$)	Direct Exposit	Salient Objects or the salient object group is put on the display screen directly
	Gravity Flowing Show	The movement of DA (Display Area) should be controlled following the movement of gravity point of the OG (Object Group), usually TGC (Tolerance of Gravity Change) parameters should be used to keep a smooth display strategy
	Saliency Driven Pan	pan operation with the consideration of saliency distribution to display the salient area on limited display window, especially in case of large saliency object or multiple saliency objects exists
True Motion Exhibition ($MV_{act}<T_{motion}$)		present all the salient object groups for a scene of the video with the weighted average gravity point as a still focus centre of the display.

According to the present invention, the adaptive video presentation operations can be classified into two categories: low motion exhibition and true motion exhibition respectively corresponding to low motion status and high motion status, which can be distinguished by low motion status and high motion status can be distinguished by the weighted average motion vector length of all MacroBlocks inside one frame MV_{ACT}. Usually a threshold T_{MOTION} can be selected to do classification, if MV_{ACT} is less than T_{MOTION}, the low motion status is determined, or else the fast motion status is determined.

In the low motion status under the low motion exhibition mode when the weighted average motion vector length for frames MV_{ACT} less then a predefined threshold T_{MOTION}, there are three different exhibition operations can be used, i.e. direct exposit, gravity flowing show and saliency driven pan. Among these three operations, the first one is to directly exposit the salient objects or salient object groups on the display, the second one is called as gravity flowing show, which control the movement of display area by following the movement of the gravity point of the salient object group, usually tolerance of gravity change (TGC) parameters are used to keep a smooth display strategy; the third one is basically a pan operation with the consideration of saliency distribution to display the salient area on limited display window, especially in case of large saliency object or multiple saliency objects exist.

In the high motion status under the high motion status when the weighted average motion vector length for frames MV_{ACT} equal or larger then a predefined threshold

T_{MOTION}, the true motion exhibition is introduced to display the salient objects or salient object group. In the high motion status, the gravity point of the OG (Object Group) is changed frequently or rapidly. In case for a scene of the video, the gravity point moves forwards and backwards, a weighted average gravity point for the scene of the video will be used as a still focus centre of the display, the middle of the forward gravity point and backward one is approximately shown in the center of the display, therefore the presentation of the video can be viewed with the true motion exhibition mode on the display, i.e. the viewer can see the OG moving forwards and backwards on the display window. In case for a scene of the video, the gravity point moves along a certain direction rapidly, then the weighted average gravity point for the scene of the video will be determined as the still focus centre of the display, the viewer can see the OG moving from one side to the other side of the display window.

4. Decisions of Adaptive Video Presentation Operation

The video can be treated as an information gravity point flowing plane, in which different salient objects have different weights of importance of the information, and the MBs have the same characteristics inside each salient object. Therefore, it's the gravity point but not the center point of the salient object or group should be the center of the display.

It can be imagined that there exits a kink of density distribution for the video content. The small display should focus on the area centralized by the gravity point of the group or a salient object, or progressively display the area by using the panning operation, which depends on the density distribution of the information.

The STP (spatial-temporal processing) module is the most important module in the AVP framework. Optimal spatial-temporal operations will be taken in the module to guarantee a smooth and acceptable video viewing experience.

Table-II demonstrates a sample of decisions of AVP operations, and of course some other types of combination can be considered due to the detail requirement of real application. In Table-II, DS means display size of the corresponding display device.

Table-II Sample of decision of AVP operations

Conditions	AVP Operations	
	Low Motion Status ($MV_{ACT} < T_{MOTION}$)	Fast Motion Status ($MV_{ACT} >= T_{MOTION}$)
RZG=<DS/n (e.g. n=2,3 …)	Direct Exposit + Zoom in	True Motion Exhibition + Zoom in
DS/n<RZG=<DS (e.g. n=2,3 …)	Direct Exposit	True Motion Exhibition
RZG>=DS	Zoom out	True Motion Exhibition + Zoom out
LOS<MPT		
One salient object	Gravity Flowing shown (with TGC limitation) + Zoom out	
Multiple salient object	Direct Exposit (Pan is forbidden)	

	LOS>m*MPT (e.g. m=2,3 …)	Gravity Flowing Show (with TGC limitation) + Saliency Driven Pan + Zoom out	
	Others	Gravity Flowing Show (with TGC limitation) + Saliency Driven Pan	

Fig. 6 demonstrates the flowchart of one exemplary scheme for decisions of the adaptive video presentation solution in accordance with the present invention.

In step 100, the motion status of the scene of the video is determined by comparing the weighted average motion vector length for frames MV_{ACT} with the predefined threshold T_{MOTION}. In case the MV_{ACT} is less than the predefined threshold T_{MOTION}, then the next step goes to step 200, or else to step 400. In step 200, it will determine whether a minimum rectangle size of the object group RZG is equal or less than the DS/n, where n=2, 3 … , and preferably n=2. If it is determined in step 200 that the RZG is equal or less than the DS/n, then it goes to step 210, which an extracted window with RZG are directly exposited on the display with an appropriate zoom-in operation. If the RZG is equal or larger then the DS, then in step 220, it will determine that if the RZG is equal or larger then the DS, if the RZG is less than the DS but larger than the DS/n, then in step 230 the extracted window with the RZG will be directly exposited on the display, if the RZG is larger than the DS, then in step 240, it will determine whether the length of scene LOS is less then the minimum perceptual time MPT. Then in step 250, it will determine whether the salient object group contains only one salient object. In a condition that only one salient object exists and the LOS is less than the MPT, the video will be presented on the display in a gravity flowing shown operation with appropriate zoom-out operation, in step 260. In step 270, case multiple salient objects exist and the LOS is less than the MPT, the video will be directly exposited on the display, since in this condition the pan operation is forbidden to avoid frequent changing of the presentation operation so as to smooth the viewing experience. In step 280, it will whether the LOS is larger than m times of MPT, where m=2, 3 …. If the LOS is larger than m times of MPT, the video will be presented on the display in a gravity flowing show operation along with a saliency driven pan operation and appropriate zoom-out operation, as in step 290. If the LOS is larger than MPT but not larger than m times of MPT, then the video will be presented in the gravity flowing show operation along with the saliency driven pan operation without zoom-out. Similarly, the decision scheme of true motion exhibition mode is made through steps 400 to 440.

Although the embodiment which incorporates the teachings of the present invention has been shown and described in detail herein, those skilled in the art can readily devise many other varied embodiments that still incorporate these teachings. Having described preferred embodiments for a system and method for allowing multi-user gaming at wireless hotspots wherein gaming users can save and transport gaming characters for use in future games and future gaming experiences (which are intended to be illustrative and not limiting), it is noted that modifications and variations can be made by persons skilled in the art in light of the above teachings. It is therefore to be understood that changes may be made in the particular embodiments of the invention disclosed which are within the scope and spirit of the invention as outlined by the appended claims. Having thus described the invention with the details and particularity required by the patent laws, what is claimed and desired protected by Letters Patent is set forth in the appended claims.

CLAIMS

1. An adaptive video presentation method for automatically presenting a video with stream-embed information based on content analysis of the video on a display with a limited screen size, comprising steps of

determining a salient object group containing at least one salient object based on perceptual interest value of macroblocks for each frame of the video,

extracting a window having a minimum size containing the salient object group for a scene of the video;

characterized in that it further comprises steps of

comparing size of the extracted window with the display size; and

presenting at least one selected area of the extracted window containing at least one part of the salient object group for the scene on the display based on the result of the comparison step.

2. The method as claimed in any one of the preceding claims, characterized in that it further comprises a step of

calculating a weighted average motion vector length MV_{act} of macroblocks inside frames for the scene of the video;

comparing the weighted average motion vector length MV_{act} of the scene with a predefined threshold T_{motion};

determining whether the weighted average motion vector length MV_{act} of the scene is less than the predefined threshold T_{motion}; and

presenting the extracted window for the scene of video on the display in a true motion exhibition mode corresponding to a fast motion status in case the average motion vector length MV_{act} of the scene is determined less than the predefined threshold T_{motion}; or else

presenting the extracted window for the scene of the video on the display in a normal exhibition mode corresponding to a low motion status in case the average motion vector length MV_{act} of the scene is determined larger than or equal to the predefined threshold T_{motion}.

3. The method as claimed in claim 2, characterized in that the extracted window containing all the salient object groups for whole the scene of the video is presented on the display with a weighted average gravity point of the salient object group for the whole scene defined as a still focus centre of the extracted window under the true motion exhibition mode in the fast motion status.

4. The method as claimed in one of claims 1 to 3, characterized in that if size of the extracted window is equal to or smaller than a predefined percentage of the display size, the extracted window is exposited on the display with an appropriate zoom-in operation.

5. The method as claimed in one of claims 1 to 3, characterized in that if size of the extracted window is larger than a predefined percentage of the display size and equal or smaller than the display size, the extracted window is exposited directly on the display.

6. The method as claimed in one of claims 1 to 3, characterized in that if size of the extracted window is larger than the display size, the extracted window is

exposited on the display with an appropriate zoom-out operation.

7. The method as claimed in claim 2, characterized in that in the normal exhibition mode corresponding to the low motion status, the method further comprises steps of

comparing length of scene of the video with a predefined minimum perceptual time;

wherein if the length of scene is less than the predefined minimum perceptual time, it further comprises steps of

determining number of the salient objects existing in the scene;

presenting the extracted window on the display with a gravity flowing show operation in case only one salient object exists, or else

presenting the extracted window directly on the display in case multiple salient objects exist;

wherein if the length of scene is no less than the predefined minimum perceptual time, it further comprises a step of

presenting the extracted window on the display with the gravity flowing show operation.

8. The method as claimed in claim 7, characterized in that the extracted window is presented on the display with a saliency driven pan operation if the length of scene is no less than the predefined minimum perceptual time.

9. The method as claimed in claim 7, characterized in that the gravity flowing show operation is performed

along with a limitation of tolerance of gravity change TGC.

10. The method as claimed in one of claims 7 to 9, characterized in that the extracted window is presented on the display with an appropriate zoom-out operation.

11. A device for adaptive video presentation which automatically presents a video with stream-embed information based on content analysis of the video on a display with limited screen size, comprising

means for determining a salient object group containing at least one salient object based on perceptual interest value of macroblocks for each frame of the video,

means for extracting a window having a minimum size containing the salient object group for a scene of the video;

characterized in that the device further comprises

means for comparing size of the extracted window with the display size; and

means for presenting at least a selected area of the extracted window containing at least a part of the salient object group for the scene on the display based on the result of the comparison step.

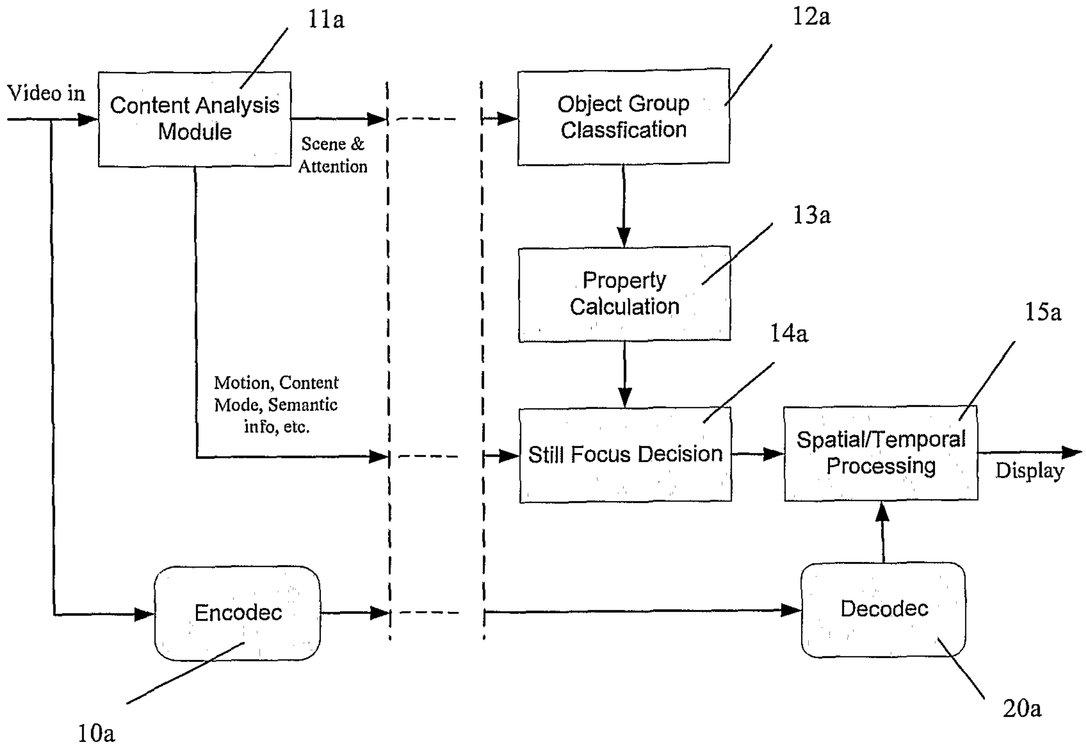

11a
12a
Video in
Content Analysis Module
Scene & Attention
Object Group Classfication
13a
Property Calculation
15a
14a
Motion, Content Mode, Semantic info, etc.
Still Focus Decision
Spatial/Temporal Processing
Display
Encodec
Decodec
20a
10a

FIG. 1

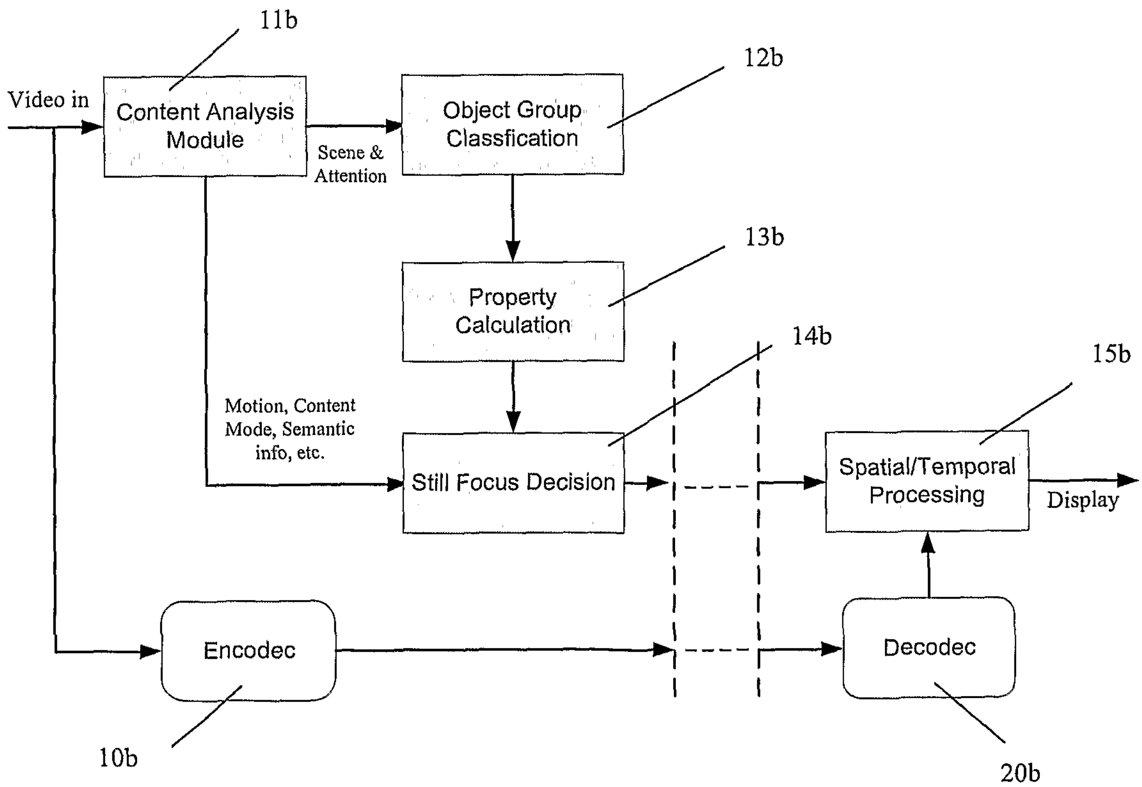

11b
12b
Video in
Content Analysis Module
Scene & Attention
Object Group Classfication
13b
Property Calculation
14b
15b
Motion, Content Mode, Semantic info, etc.
Still Focus Decision
Spatial/Temporal Processing
Display
Encodec
Decodec
10b
20b

FIG. 2

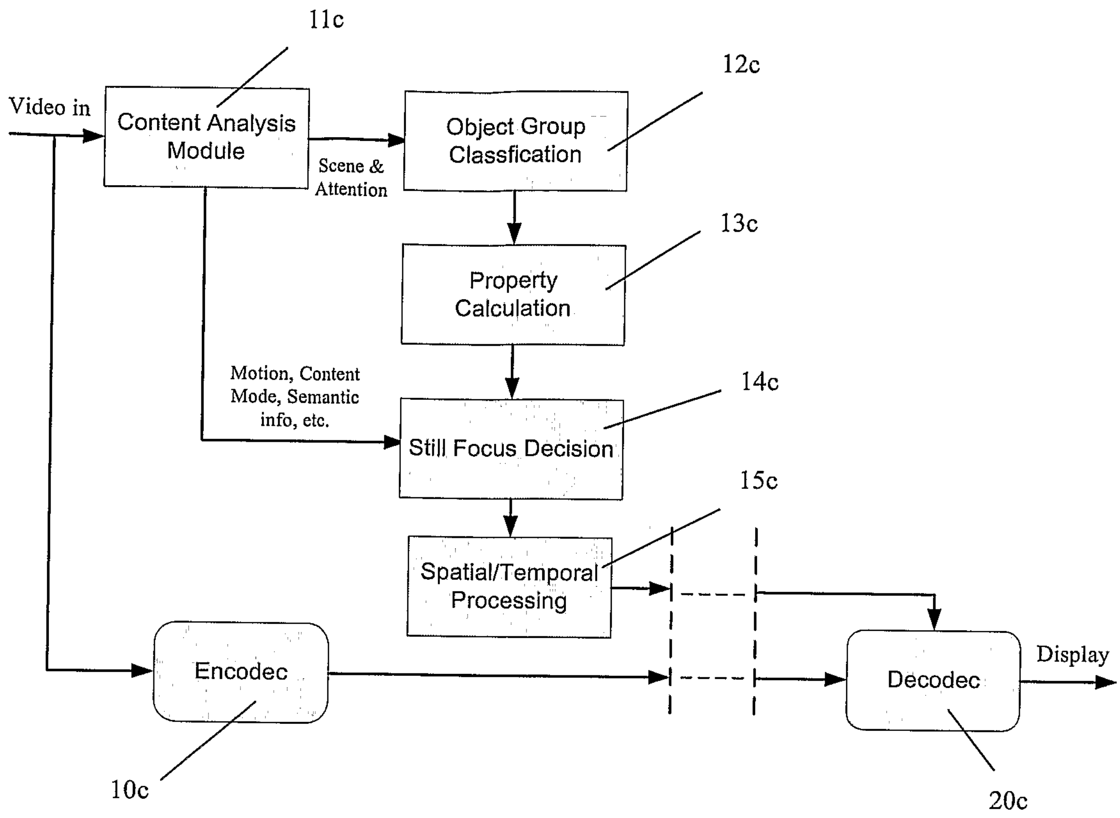


FIG. 3

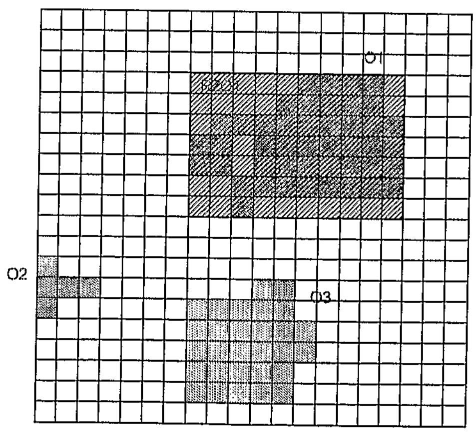


FIG. 4

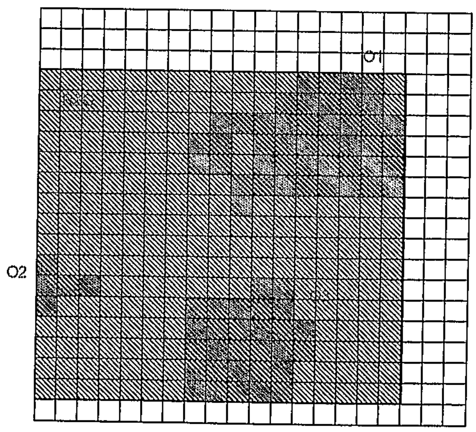


FIG. 5

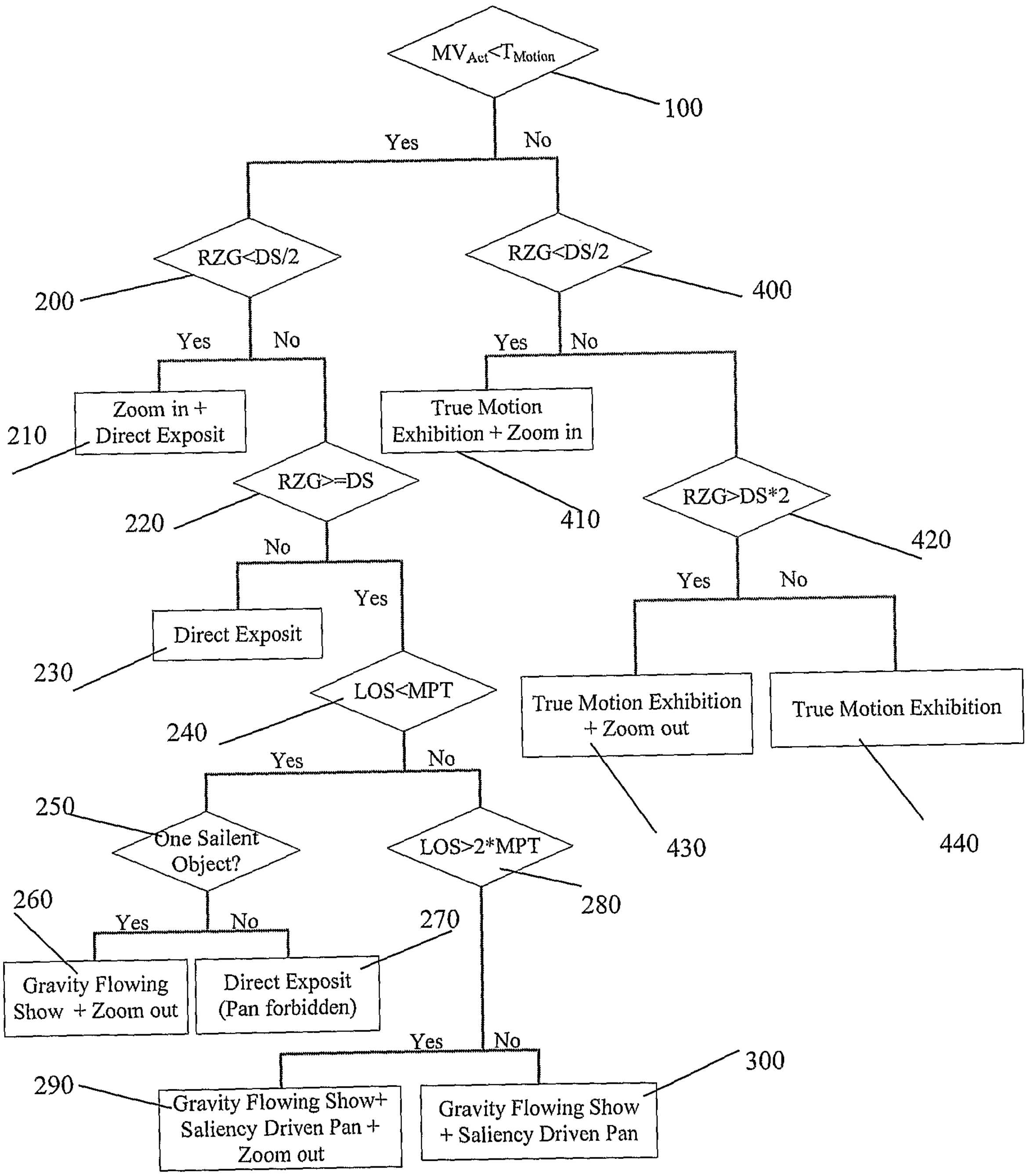

MV_{Act}<T_{Motion}
100
Yes
No
RZG<DS/2
200
RZG<DS/2
400
Yes
No
Yes
No
Zoom in + Direct Exposit
210
True Motion Exhibition + Zoom in
410
RZG>=DS
220
RZG>DS*2
420
No
Yes
Yes
No
Direct Exposit
230
LOS<MPT
240
True Motion Exhibition + Zoom out
430
True Motion Exhibition
440
Yes
No
250
One Sailent Object?
LOS>2*MPT
280
260
Yes
No
270
Gravity Flowing Show + Zoom out
Direct Exposit (Pan forbidden)
Yes
No
290
Gravity Flowing Show+ Saliency Driven Pan + Zoom out
Gravity Flowing Show + Saliency Driven Pan
300

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2006/002261

A. CLASSIFICATION OF SUBJECT MATTER

G06T 3/40 (2006.01) i

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)

IPC G06T 3/- H04N 7/-

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)

WPI; EPODOC; PAJ; CNPAT: video, image, present+, display, size, adaptive+, auto+, metadata, salient, compar+ Scal+, frame, scene, convert+, encode, decode, mobile, terminal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN1529499A (UNIV SHANGHAI JIAOTONG) 15 Sep.2004 (15.09.2004) Whole document	1-11
A	WO2004090812A1 (KONINKL PHILIPS ELECTRONICS NV) 21 Oct.2004 (21.10.2004) Whole document	1-11
A	US2005226538A1 (KONINKL PHILIPS ELECTRONICS NV) 13 Oct.2005 (13.10.2005) Whole document	1-11

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&"document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 May.2007 (29.05.2007)	14 · JUN 2007 (14 · 06 · 2007)

Name and mailing address of the ISA/CN
The State Intellectual Property Office, the P.R.China
6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China
100088
Facsimile No. 86-10-62019451

Authorized officer
WU, Shuang
印 吴爽
Telephone No. (86-10) 62084676

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2006/002261

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN1529499A	15.09.2004	CN1233161C	21.12.2005
WO2004090812A1	21.10.2004	KR20050119689A	21.12.2005
		EP1616300A	18.01.2006
		CN1771516A	10.05.2006
		US2006181643A	07.08.2006
		JP2006523409T	12.10.2006
US2005226538A1	13.10.2005	WO03102903A	11.12.2003
		AU2003228063A	19.12.2003
		EP1514236A	16.03.2005
		CN1659591A	24.08.2005
		JP2005528643T	22.09.2005