



(51) International Patent Classification:

H04N 7/18 (2006.01) *G08B 13/196* (2006.01)
G09G 5/00 (2006.01)

(21) International Application Number:

PCT/MY2013/000164

(22) International Filing Date:

9 September 2013 (09.09.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

PI 2012004072 13 September 2012 (13.09.2012) MY

(71) Applicant: MIMOS Berhad [MY/MY]; Technology Park of Malaysia, Bukit Jalil, 57000 Kuala Lumpur (MY).

(72) Inventors: **SENAPEN, Mohanavelu Mudaliar A/L**; c/o Mimos Berhad, Technology Park of Malaysia, Bukit Jalil, 57000 Kuala Lumpur (MY). **LIM, Oon Leong**; c/o Mimos Berhad, Technology Park of Malaysia, Bukit Jalil, 57000 Kuala Lumpur (MY). **KARUPPIAH, Ettikan Kandasamy A/L**; c/o Mimos Berhad, Technology Park of Malaysia, Bukit Jalil, 57000 Kuala Lumpur (MY). **ONG, Hong Hoe**; c/o Mimos Berhad, Technology Park of Malaysia, Bukit Jalil, 57000 Kuala Lumpur (MY). **SOO, Saw Meng**; c/o Mimos Berhad, Technology Park of Malaysia, Bukit Jalil, 57000 Kuala Lumpur (MY).

(74) Agent: **HEMINGWAY, Christopher Paul**; Marks & Clerk (Malaysia) Sdn Bhd, Lot 5B Level 5, Tower 2 Etika Twins, 11 Jalan Pinang, 50450 Kuala Lumpur (MY).

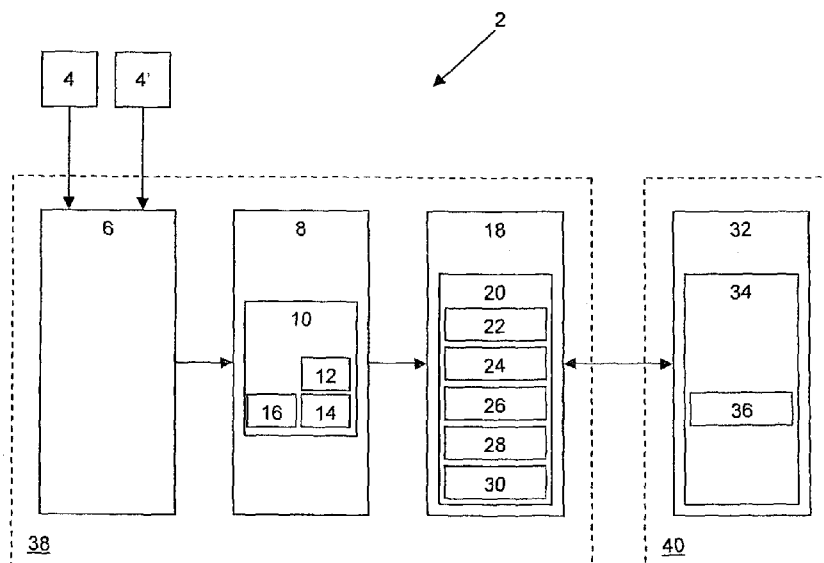
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

(54) Title: SYSTEM FOR IMPROVING IMAGE PROCESSING GOODPUT



(57) Abstract: An image processing system (2) comprising a plurality of imaging sources (4, 4') for capturing images and/or video; a central processing unit (38) for processing the images and/or video; at least one graphics processing unit (40) for further processing the images and/or video received from the central processing unit; wherein the central processing unit (38) includes a prioritization unit for assigning one or more priority values to the images and/or video based on the location (26, 26') of the imaging sources and/or the rating (24, 24') of the images and/or video, selectively concatenating (28) the images and/or video with priority values above a predetermined threshold, and transferring (136) the concatenated data to the at least one graphics processing unit (40).

SYSTEM FOR IMPROVING IMAGE PROCESSING GOODPUT

Field of Invention

The invention relates to a system for improving image processing goodput.

5

Background

In computer networks, goodput is the application level throughput, i.e. the number of useful information bits, delivered by the network to a certain destination, per unit of time. The useful information bits excludes protocol overhead bits as well as retransmitted data packets, and the goodput is assessed with regard to the amount of time from the first bit of the first packet is sent (or delivered) until the last bit of the last packet is delivered.

For example, if a file is transferred, the goodput that the user experiences corresponds to the file size in bits divided by the file transfer time. The goodput is always lower than the throughput (the gross bit rate that is transferred physically), which generally is lower than network access connection speed (the channel capacity or bandwidth).

In an imaging processing system, data is transferred between a central processing unit (CPU) and one or more graphics processing units (GPUs) via a bus such as PCI

express, but because some applications frequently transfer small amounts of data, a significant amount of bandwidth is wasted due to transfer overheads. In addition this problem is compounded by lack of data transfer coordination between the CPU and GPU.

5

Furthermore the data transfer between CPU and GPU is performed with the help of a queue buffer. Whenever congestion occurs on the PCI Express bus, data accumulates on either side (GPU or CPU) and if the congestion is not cleared in time, the buffer will overflow and may cause data to drop randomly. This behaviour is undesirable as it introduces unpredictability to the processing output. For example, in the case of video analytics, the accuracy of the video analytic and analytic detection probability will drop.

10

An aim of the invention is to improve the goodput in an image processing system.

15

Summary of Invention

In an aspect of the invention there is provided an image processing system comprising:

20

a plurality of imaging sources for capturing images and/or video;

a central processing unit for processing the images and/or video;

at least one graphics processing unit for further processing the images and/or video received from the central processing unit;

characterised in that the central processing unit includes a prioritization unit for assigning one or more priority values to the images and/or video based on the location of the imaging sources and/or the rating of the images and/or video, and selectively transferring the images and/or video with priority values above a
5 predetermined threshold to the at least one graphics processing unit.

Thus by prioritizing the images/video, the data passed to the GPU is filtered to reduce bandwidth wastage, thereby improving goodput.

10 In one embodiment the imaging sources comprise cameras.

In one embodiment the images and/or video with priority values above a predetermined threshold are concatenated in a buffer prior to being transferred to the GPU.

15

Typically the cameras stream images and/or video, and the images and/or video with the highest priority values received from each stream at a particular instant are concatenated first.

20 Typically the GPU segments the concatenated images and/or video into a non-concatenated form so that the images and/or video can be processed thereby.

The concatenation allows the data to be transferred from CPU to GPU more efficiently, thereby further improving goodput.

In one embodiment the images and/or video from the imaging sources are provided
5 as multiple streams for prioritization, but are concatenated as a single stream for transfer to a GPU. Typically the data is only transferred to the GPU when the buffer is full.

In one embodiment the CPU includes preprocessing means for filtering out images
10 and/or video with a rating below a predesignated threshold. Typically the rating of the images and/or video is determined by video analytics.

In one embodiment the load of the at least one GPU is estimated and provided as feedback to the preprocessing means to adjust the predetermined threshold as
15 appropriate. Thus if the GPU load is low, the threshold can be reduced. Conversely, if the GPU load is high, the threshold can be increased.

Typically the load is determined by measuring the GPU execution time per image.

20 In a further aspect of the invention there is provided a method of processing images comprising the steps of:

- capturing images and/or video from a plurality of imaging sources;
- processing the images and/or video via a central processing unit;

further processing the images and/or video received from the central processing unit via at least one graphics processing unit;

characterised in that one or more priority values are assigned to the images and/or video based on the location of the imaging sources and/or the rating of the images and/or video, and images and/or video with priority values above a predetermined threshold are selectively transferred to the at least one graphics processing unit.

In one embodiment the images and/or video with priority values above a predetermined threshold are concatenated in a buffer prior to being transferred to a graphics processing unit.

In one embodiment the concatenated images and/or video are segmented into a non-concatenated form so that they can be processed by the graphics processing unit.

Brief Description of Drawings

It will be convenient to further describe the present invention with respect to the accompanying drawings that illustrate possible arrangements of the invention. Other arrangements of the invention are possible, and consequently the particularity of the accompanying drawings is not to be understood as superseding the generality of the preceding description of the invention.

Figure 1 is a block diagram of the system according to an embodiment of the invention.

Figure 2 is a block diagram illustrating the overall flow of the system according to an
5 embodiment of the invention.

Figure 3 is a block diagram illustrating further detail of the preprocessing step.

Figure 4 is a block diagram illustrating further detail of the prioritization step.
10

Figure 5 is a block diagram illustrating further detail of the buffer aggregator step.

Figure 6 is a block diagram illustrating the buffer aggregator activity.

15 Figure 7 is a block diagram illustrating the segmentation step.

Figure 8 is a block diagram illustrating the GPU load estimation step.

Detailed Description

20

With reference to Figure 1, an overview of the system 2 according to an embodiment of the invention is illustrated.

In the central processing unit (CPU) 38, a network stream data receiving unit 6 receives network data from a plurality of cameras 4, 4'. The data is then passed to a video acquisition component 10 in a media data processing unit 8 to be converted to images in various formats such as MJPEG 12, JPEG 14 or MPEG4 16.

5

Data transfer to the graphics processing unit (GPU) 40 is controlled by a Multi-Factor Buffer Concatenation component 20 in a media preparation unit 18, which comprises Image Activity Preprocessor 22, Selection Prioritization 24, Camera Location Prioritization 26, Buffer Aggregator 28 and GPU Load Estimation 30, in
10 connection with Segmentation 36 in the GPU video processing unit 34.

With reference to Figure 2, the overall flow chart of the system is illustrated. The network frames are acquired 42, 42' from N (2 in this example) cameras 4, 4' by the network stream data receiving unit 6, and converted 44, 44' to images in the media
15 data processing unit 8.

In the Video Stream Preparation Unit 18 the images are preprocessed 22, 22' to filter out images that are less significant and images that pass this step will be inserted into the respective Preprocessing Output Buffer 46, 46'. Note that the video streams are
20 independent of each other before the Buffer Aggregator stage 28.

Images in the Preprocessing Output Buffer 46, 46' are retrieved for first round prioritization 24, 24' where images are assigned priority value based on their

importance (image rating). This is followed by second round of prioritization 26, 26' based on location to determine the final priority value for each image. At the end of prioritization, the images are tagged with their priority values and inserted into the respective Prioritization Output Buffer 48, 48'.

5

At the Buffer Aggregator 28, images with the highest priority value will be selected from the N prioritization output buffers 48, 48' and inserted into the transport buffer to be transferred 50 to the GPU video processing unit 34. This selection process of high priority images will continue until the transport buffer is full. After the transport
10 buffer is emptied, or in other words, the data has been transferred to the GPU video processing unit 34, the whole process is repeated.

At the GPU video processing unit 34, the transferred data must be segmented 36 or partitioned back to the original individual images for GPU consumption 52.

15

Finally, the GPU processing results are transferred 54 back to the CPU and the GPU load is estimated 30 and provided as feedback to the preprocessing step 22, 22' to tune the filtering criteria.

20 With reference to Figure 3, images are received from the camera 4 and the preprocessing step is initialised 56 by loading the configuration settings 58, reference image 60 and the analytics measurement rating table 62.

Once initialized, the normal analytics steps such as bitmap conversion 64, pixel data extraction 66 and analytic specific measurement 68 will take place. An example of analytic specific measurement is, in the case of motion detection, image difference. As the step name implies this measurement depends on the type of analytics. The measurement will then be retrieved 70 and mapped 72 to a rating value by looking it up in a configurable measurement rating table.

The rating is compared 74 against a threshold, and images with ratings less than a variable threshold will be discarded 76. Images with ratings greater than the threshold will be tagged 78 with their ratings and inserted into the preprocessing output buffer 46.

With reference to Figure 4, further details are provided about the prioritization step. The thread 80 is initialised by obtaining 82 the camera ID. The camera ID is then used to lookup 84 the camera location weight. This weight is one of the two factors that determine the final priority value for the images.

A loop calculates the average image rating of N first images in the preprocessing output buffer, as defined by the following steps in the illustrated diagram:

20 86 Sum=0, Loop variable j=1
 88 Look for image j in preprocessing output buffer
 90 If image j is not available, then wait 92;
 If image j is available, then retrieve Image Rating 94

- 96 add image rating to Sum
- 98 Increment j by 1
- 100 If $j < N$ then return to the step of checking if image j is available 88
- If $j > N$ then calculate average 102
- 5 104 calculate priority = average + camera location weight
- 106 fetch first image
- 108 tag image with priority,
- then insert image into prioritization output buffer 48
- 10 This average image rating and the camera location weight are used to determine the final priority value for the current image. In order to appreciate these two factors that determine the final image priority value, camera location weight and image rating average are considered.
- 15 With regard to camera location weight, in a case where some cameras (or image producing devices) are located at secure/sensitive sites, it is appropriate to assign such cameras a bigger weight. This is because images coming from these cameras should have higher priority and be processed first.
- 20 With regard to image rating average, in the case where images in a video stream have been identified to have high priority, it's likely that subsequent images will have important information or activity in them. Therefore, through averaging, the subsequent images will have higher priority values.

Thus the current image is tagged with the priority value and inserted into the prioritization output buffer 48. This process is repeated as long as there are enough images in the preprocessing output buffer 46 to calculate the Image rating average.

5

With reference to Figure 5, further details are provided about the buffer aggregator 28 which is initialised 110 as a separate thread.

A transport buffer (memory buffer) is allocated 112, and is tasked with moving high
10 priority images from CPU to GPU. The buffer aggregator picks images with the highest priority and inserts them into the transport buffer. This is accomplished by comparing the priority values of the first images of all N prioritization output buffers, and repeated until the transport buffer is full, as defined by the following steps in the illustrated diagram:

```
15      114   Loop variable i = 1
          116   Priority array[i] = 0
          118   If prioritization output buffer i does not contain M or more images,
                  then 120 increment i by 1;
                  If prioritization output buffer i contains M or more images, then 122
20      retrieve priority value of first image in prioritisation output buffer i
          124   store in priority array [i]
          126   If i < number of cameras then 120 increment i by 1
```

If $i >$ number of cameras then 128 find prioritization output buffer
with largest priority value

130 fetch M images from the buffer

132 insert into transport buffer

5 134 if transport buffer is full, then transfer 136 to GPU, otherwise return to
start 114 of loop

With regard to Figure 6 the buffer aggregator activity is illustrated wherein $N=M=2$
and each image has two values: R which stands for image rating and P for priority
10 value.

The transport buffer is always filled with M images at a time, taken from the buffer
holding the first image with highest priority value. M images are required because
some video analytics require at least M images for processing. For example, for
15 motion detection at least two images are needed.

The streams from four cameras (CAM1-4) in this example each contain a number of
images wherein each image is tagged with a priority value (P value). The cameras
may be given a prioritization weighting (W-value). The Buffer Aggregator will look
20 at the priority value of the first image in each stream to determine which is first to be
moved to the transport buffer. In this example, the stream from camera 1 goes first
because the first image has a P value of 5.5 which is highest among its peers. Second
is Camera 2 as the first image of the stream has a P-value of 5.0.

Notably, since the images are moved to the transport buffer in pairs ($M=2$ in this example), IMG3 and IMG4 from camera 1 will be the third pair selected, because the P value (4.5) of IMG3 from camera 1, which at that time is the first image in stream
5 1, is the highest among its peers.

Now the buffer has six images and is therefore full and ready to be transported to the GPU. This entire selection process then repeats.

10 With regard to Figure 7 the segmentation step is illustrated wherein the concatenated images are split into the original individual images for GPU consumption.

The border of individual images is found 138 and the concatenated images are split 140. The kernel is run 142 to process the images, and then the processed images are
15 concatenated 144 and transferred 146 back to the CPU.

With regard to Figure 8 the GPU load estimation step is illustrated. In this step, GPU load is estimated based on the GPU execution time.

20 The analytics and GPU results are transferred 148 from the GPU to the CPU and the GPU execution time per image is measured 150.

If it is found 152 that the execution timer is greater than the upper threshold, then the GPU-driven variable threshold is decreased 156 by 1 level. If not, and it is found 154 that the execution timer is less than the lower threshold, then the GPU-driven variable threshold is increased 158 by 1 level

5

As such the estimated load adjusts the GPU driven variable threshold which is used in the preprocessing image step for filtering less significant images. Thus if the GPU is busy, the threshold is tightened to allow only highly significant images to pass through the preprocessing image step, which is useful because the GPU is already
10 busy and has limited free resources to process the incoming data.

It will be appreciated by persons skilled in the art that the present invention may also include further additional modifications made to the device which does not affect the overall functioning of the device.

15

CLAIMS

1. An image processing system (2) comprising:
 - a plurality of imaging sources (4, 4') for capturing images and/or video;
 - 5 a central processing unit (38) for processing the images and/or video;
 - at least one graphics processing unit (40) for further processing the images and/or video received from the central processing unit;
 - characterised in that the central processing unit (38) includes a prioritization unit for assigning one or more priority values to the images and/or video based on the location (26, 26') of the imaging sources and/or the
10 rating (24, 24') of the images and/or video, and selectively transferring (136) the images and/or video with priority values above a predetermined threshold to the at least one graphics processing unit (40).
- 15 2. An image processing system according to claim 1 wherein the images and/or video with priority values above a predetermined threshold are concatenated (28) in a buffer prior to being transferred to a graphics processing unit.
3. An image processing system according to claim 2 wherein the images and/or
20 video with the highest priority values, from those captured at any instant from the imaging sources, are concatenated first.

4. An image processing system according to claim 2 wherein the graphics processing unit segments (36) the concatenated images and/or video into a non-concatenated form so that the images and/or video can be processed thereby.

5

5. An image processing system according to claim 2 wherein the images and/or video from the imaging sources are provided as multiple streams for prioritization, but are concatenated as a single stream for transfer to a graphics processing unit.

10

6. An image processing system according to claim 2 wherein the concatenated images and/or video are only transferred to the graphics processing unit when the buffer is full.

- 15 7. An image processing system according to claim 1 wherein the central processing unit includes preprocessing means for filtering out images and/or video with a rating below a predesignated threshold.

- 20 8. An image processing system according to claim 7 wherein the rating of the images and/or video is determined by video analytics.

9. An image processing system according to any preceding claim wherein the load of the at least one graphics processing unit is estimated and provided as feedback to the preprocessing means to adjust the predetermined threshold as appropriate.

5

10. An image processing system according to claim 9 wherein the load is determined by measuring the GPU execution time per image.

11. A method of processing images comprising the steps of:

10 capturing images and/or video from a plurality of imaging sources (4, 4');

processing the images and/or video via a central processing unit (38);

further processing the images and/or video received from the central processing unit via at least one graphics processing unit (40);

15 characterised in that one or more priority values are assigned to the images and/or video based on the location (26, 26') of the imaging sources and/or the rating (24, 24') of the images and/or video, and images and/or video with priority values above a predetermined threshold are selectively transferred (136) to the at least one graphics processing unit.

20

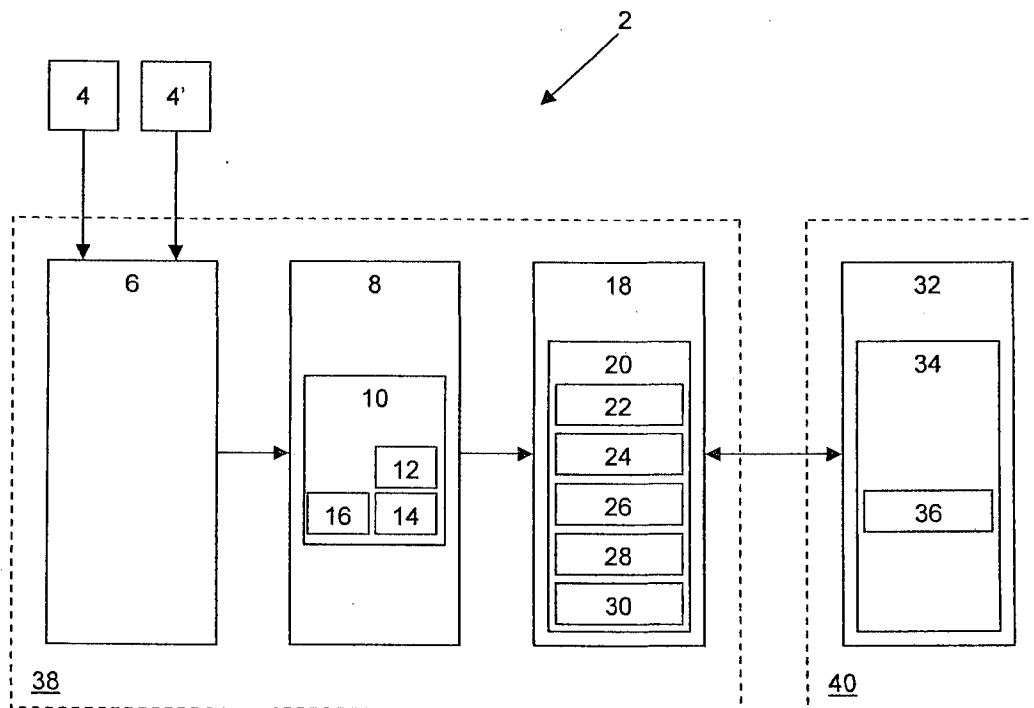
12. A method of processing images according to claim 11 wherein the images and/or video with priority values above a predetermined threshold are

concatenated (28) in a buffer prior to being transferred to a graphics processing unit.

13. A method of processing images according to claim 12 wherein the
5 concatenated images and/or video are segmented into a non-concatenated form
so that they can be processed by the graphics processing unit.

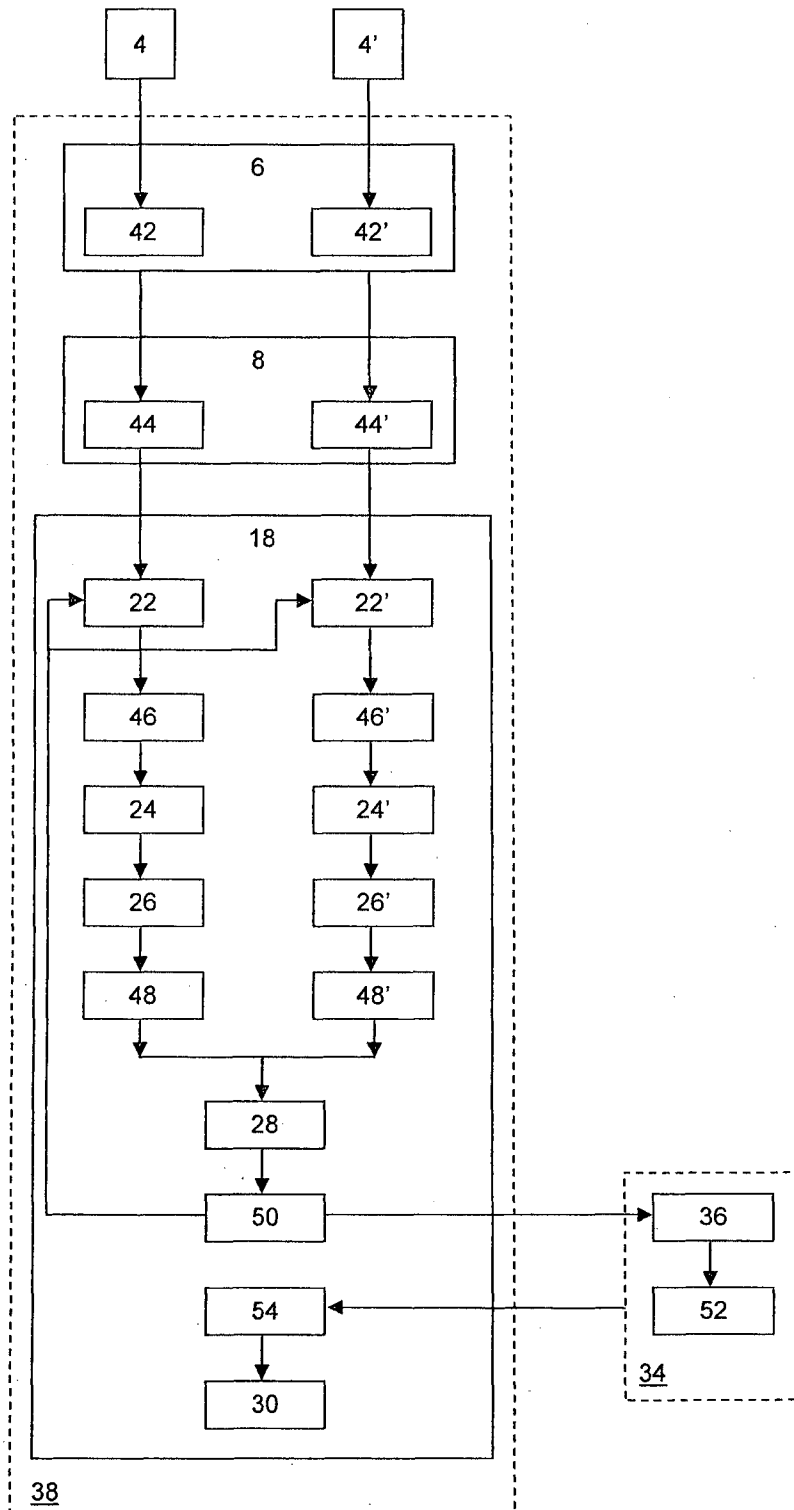
1/6

Figure 1

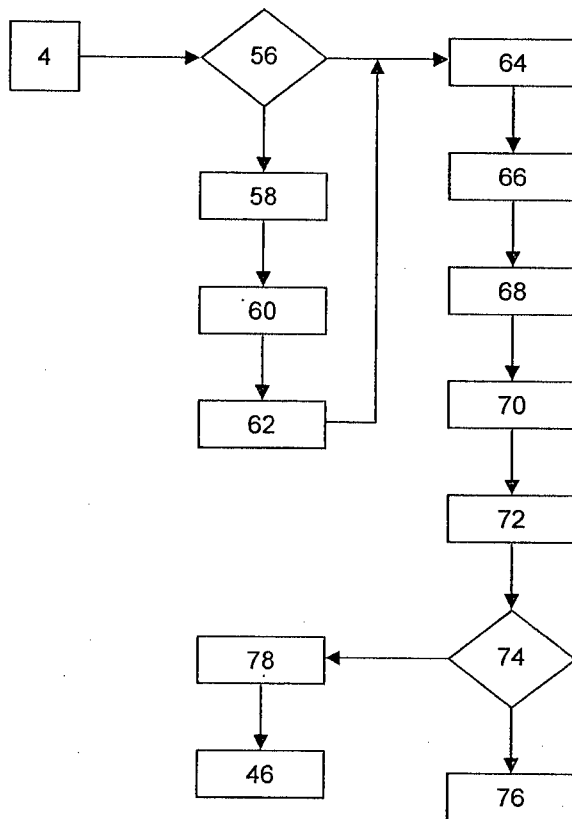


2/6

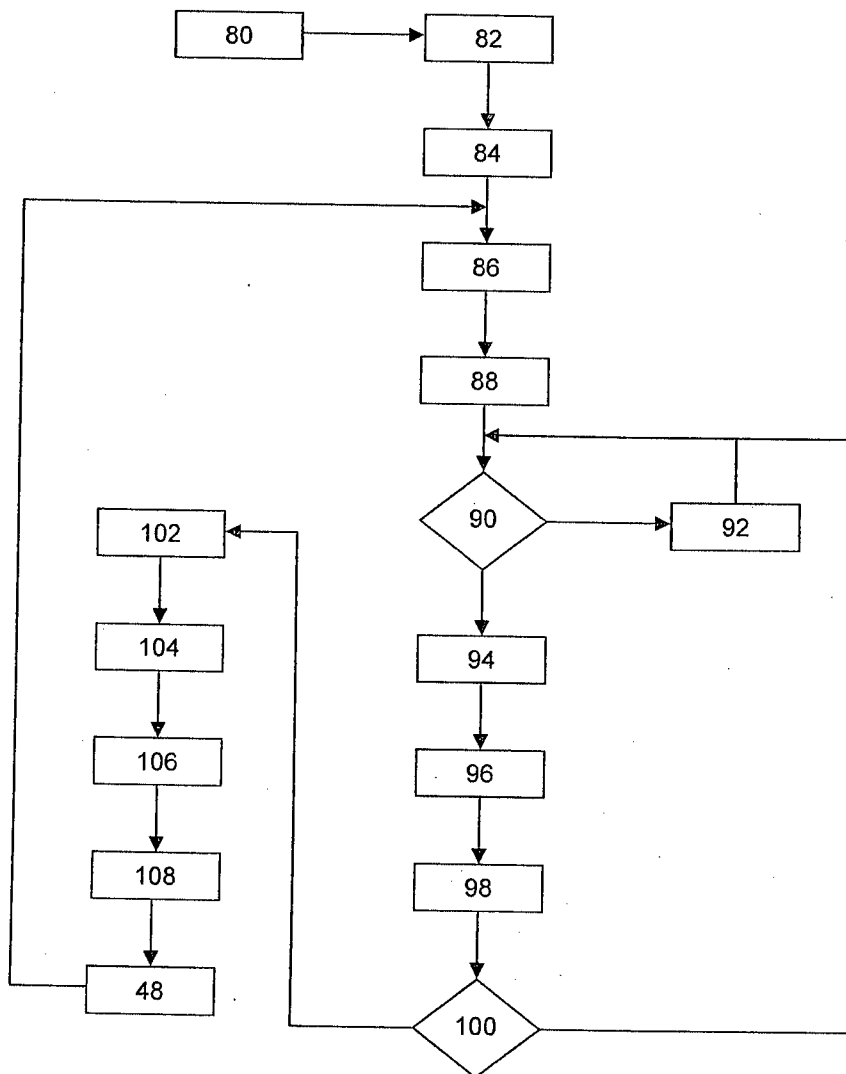
Figure 2



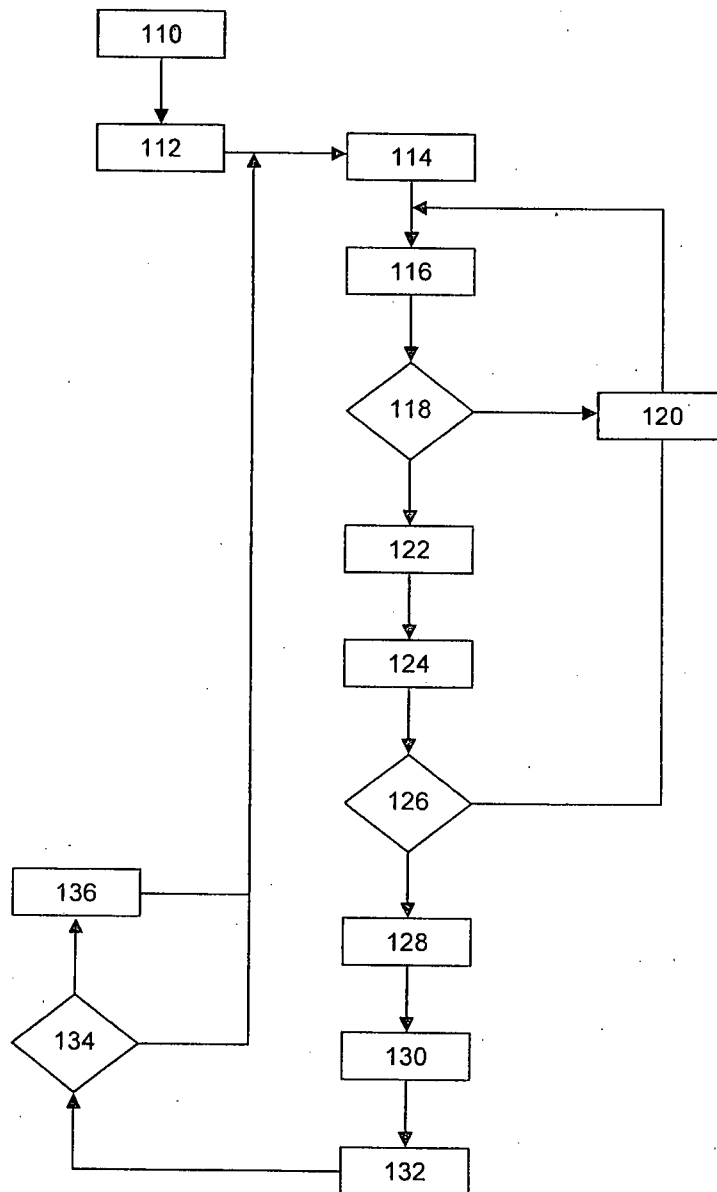
3/6

Figure 3

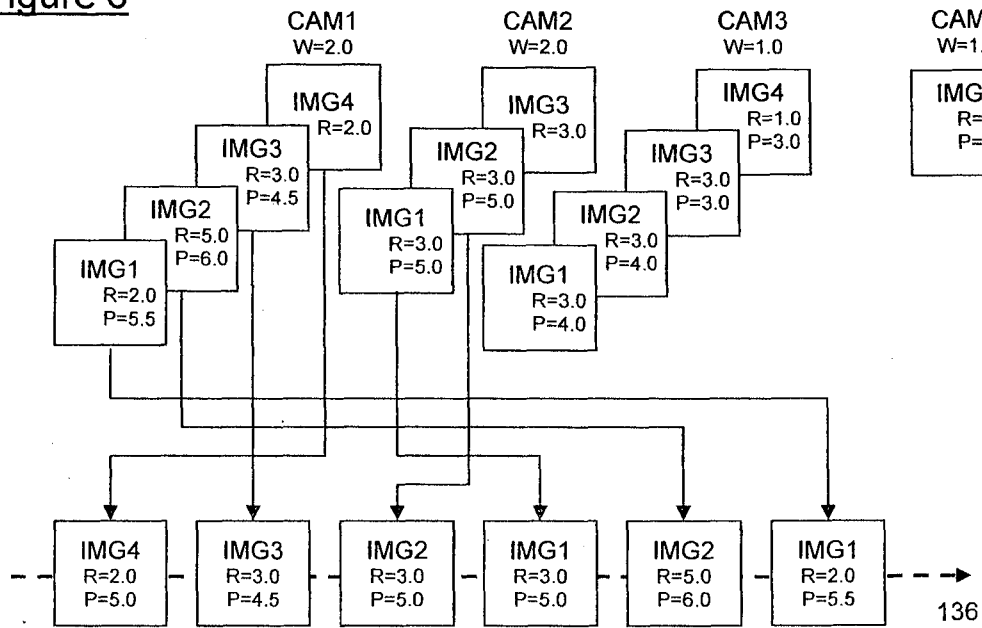
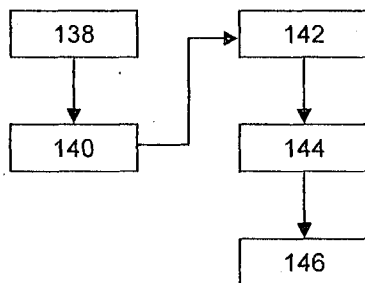
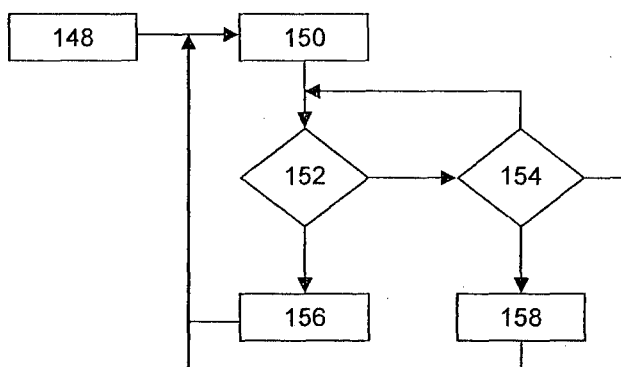
4/6

Figure 4

5/6

Figure 5

6/6

Figure 6Figure 7Figure 8

INTERNATIONAL SEARCH REPORT

International application No
PCT/MY2013/000164

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/18 G09G5/00 G08B13/196
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)
H04N G09G G08B

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)

EP0-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	WO 2011/109935 A1 (GLOBAL ADVANCED VISION LTD [CN]; JIN JESSE SHENG [AU]; AU SZE LOK [CN]) 15 September 2011 (2011-09-15) paragraph [0034] - paragraph [0055] -----	1-13
A	US 2012/119879 A1 (ESTES ANDREW D [US] ET AL) 17 May 2012 (2012-05-17) paragraph [0069] - paragraph [0071]; figure 9 paragraph [0077] - paragraph [0078] -----	1,11
A	US 2007/001512 A1 (SATO MASAYUKI [JP] ET AL) 4 January 2007 (2007-01-04) abstract; figure 1 -----	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

13 December 2013

Date of mailing of the international search report

02/01/2014

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

Beaudoin, Olivier

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/MY2013/000164

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2011109935 A1	15-09-2011	CN 102860003 A	02-01-2013
		US 2013011014 A1	10-01-2013
		WO 2011109935 A1	15-09-2011

US 2012119879 A1	17-05-2012	US 2012119879 A1	17-05-2012
		US 2012212611 A1	23-08-2012

US 2007001512 A1	04-01-2007	DE 102006028981 A1	11-01-2007
		JP 4615380 B2	19-01-2011
		JP 2007013497 A	18-01-2007
		US 2007001512 A1	04-01-2007
