



- (51) **International Patent Classification:**
H04N 21/2343 (2011.01) *H04N 21/44* (2011.01)
H04N 21/4223 (2011.01) *H04N 21/6587* (2011.01)
- (21) **International Application Number:** PCT/IB2016/000262
- (22) **International Filing Date:** 26 January 2016 (26.01.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
201510094975.X 3 March 2015 (03.03.2015) CN
201510207760.4 27 April 2015 (27.04.2015) CN
- (71) **Applicant:** ALCATEL LUCENT [FR/FR]; 148/152 route de la Reine, 92100 Boulogne-Billancourt (FR).
- (72) **Inventor:** CHEN, Yu; 388# Ningqiao Road, Pudong Jin-qiao, Shanghai 201206 (CN).
- (74) **Agent:** BERTHIER, Karine; Alcatel-Lucent International, 148/152 route de la Reine, 92100 Boulogne-Billancourt (FR).
- (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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, 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, ST, 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,

[Continued on next page]

(54) **Title:** METHOD AND APPARATUS FOR TRANSMITTING A VIDEO

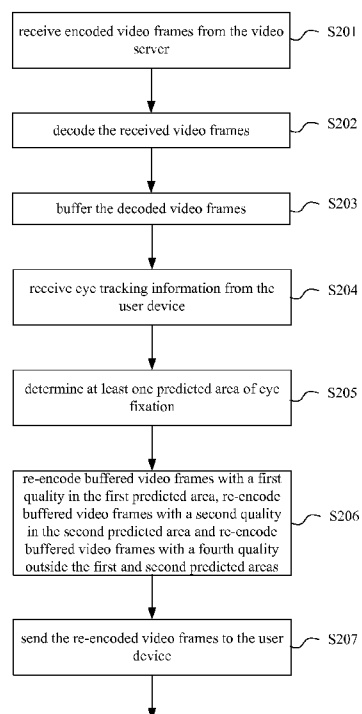


Figure 6

(57) **Abstract:** The invention provides a method, for use in a communication device, of transmitting a video, the method comprising the steps of: receiving encoded video frames from a video server; decoding the received video frames; buffering the decoded video frames; receiving eye tracking information from a user device; determining a first predicted area of eye fixation based on the eye tracking information; re-encoding buffered video frames with a first quality in the first predicted area and re-encoding buffered video frames with a fourth quality outside the predicted area, wherein the first quality is better than the fourth quality; and sending the re-encoded video frames to the user device.





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))

METHOD AND APPARATUS FOR TRANSMITTING A VIDEO

Field of the Invention

The present disclosure relates generally to communication systems, and more particularly, to video transmissions based on eye tracking techniques.

Background of the Invention

The study of human visual system has a long history. One of the most important findings is the structure of the retina. The retina covers the back surface of the eye ball connecting the ciliary body, which hosts the huge number of photo-receptors, rod and cone cells. Cone cells are smaller than rod cells but more important for color vision. Most of the cone cells are located in the macula area which is near the blind spot for blood vessels and nerves into the brain. There are around 90 million rod cells, compared to 4.5 million cone cells. Rods are responsible for the night vision or scotopic vision and rod cells cannot differentiate colors. In bright light, the color-sensitive cones are predominant, so we see a colorful world. In bright daylight environment, rods may saturate and then only cones work. In the indoor environment, both rods and cones contribute to the vision. Next, it takes some time for the cells to adapt to the environments, for rods it takes about 30 minutes fully adapted to the dim light, slower than cones. Hence, when talking about video, only cone cells are relevant.

The density of the photo-receptors in the retina varies greatly. For rods, the peak appears in 20 degree around the center and decreases to the edge. In contrast, most of the cones are located in the center, a very small area called fovea, around 1.5mm wide and this is the core of the human vision, but it is rods free. The distribution of rods and cones are shown in Figure 1. The human color vision is limited in this several degree area. Hence, the human only have good eyesight in the gaze point. Few people notice this because the eye ball moves and our brain can piece the small patches of good vision together.

One may consider if a system only transmits where the eye gazes at by eye tracking technique, huge gains may be achievable. However, the eye moves at a very high speed, about 400 degree per second. This requires the system performing with extremely low latency. For example, when the eye moves from one corner of a display to the other across 20 degree, it only takes 50ms. Hence, the eye tracking information reporting and video

transmission switching should be finished within 50ms. This poses a big challenge to the network, which may only be supported by 5G. 5G is the network expected to be available by 2020 and featured by high capacity and extremely low transmission delay. New techniques might be used to support these including large scale antenna array, new frame structure, new scheduling mechanism and etc. Moreover, fundamental changes to the network architecture are also required to support end to end eye tracking based video transmission.

Object and Summary of the Invention

There are some studies to link the human visual perception and video transmission. In “Robert-Inacio, F. ; Scaramuzzino, R. ; Stainer, Q. ; Kussener-Combier, E., Biologically inspired image sampling for electronic eye, Biomedical Circuits and Systems Conference (BioCAS), 2010, pages: 246-249”, an image sampling scheme is proposed for electronic eye. The sampling is made based on a hexagon pavement where the area of each hexagon increases with the distance to the focus. The human gazing behavior is studied in “Laura Muir Iain, Iain Richardson, Steven Leaper, Gaze Tracking and Its Application to Video Coding for Sign Language, Picture Coding Symposium 2003, pages 32-325” to discover which part of a picture is likely to be gazed by human. Mohsen M. proposed a real-time eye tracking based video coding system in “Mohsen M. Farid, Fatih Kurugollu, Fionn D. Murtaghk, Adaptive wavelet eye-gaze-based video compression, Proc. SPIE 4877, Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision, 255 (March 17, 2003)”. In this system, the video frame is sub-blocked and encoded according to the eye tracker feedback. This scheme is realized between computers in laboratory, where the latency might not be the constraint. At the moment, the model of human visual perception is only studied in “Robert-Inacio, F. ; Scaramuzzino, R. ; Stainer, Q. ; Kussener-Combier, E., Biologically inspired image sampling for electronic eye, Biomedical Circuits and Systems Conference (BioCAS), 2010, pages: 246-249” but without the consideration of eye movement behavior. The next aspect is latency, which is the main challenge of video transmission in a mobile network and is not studied yet.

One potential solution for a mobile network like 5G could be the eye tracker feeds back the eye gaze information to the video server and the server encodes the video accordingly based on the gaze information. However, this suffers too long delay. The

general network architecture is depicted in Figure 2, where the video server distributes the video to the base stations. Based on such architecture, the total delay is evaluated in Table I.

The end to end delay is summarized in Table I, showing totally 106ms. The response delay from saccadic movement to fixation is around 30ms, and a maximum delay should be less than 50ms. Technique advance may shorten these delay components. For example, by using 100Hz eye tracker, the measurement delay could be shortened to 10ms. However, transmission delay is still not acceptable. There is need to optimize the main delay component, the transmission from the base station to the video server.

Table I Delay analysis

Delay contributor	Delay (ms)
Eye tracker measurement delay	20ms for 50Hz
Eye tracking info report delay to BS	8ms for two HARQ RTT
Eye tracking info report to video server	20ms
Video encoding	10ms
Encoded video to BS	20ms
Primary BS to secondary BS	10ms
Video transmission from BS to MS	8ms for two HARQ RTT
BS to BS signaling	10ms
Total	106ms

Based on above concerns, the purpose of this invention is to provide an eye tracking based video transmission system which could reduce system delay and save resources.

In one aspect of the invention, there is provided a method, for use in a communication device, of transmitting a video, the method comprising the steps of: receiving encoded video frames from a video server; decoding the received video frames; buffering the decoded video frames; receiving eye tracking information from a user device; determining a first predicted area of eye fixation based on the eye tracking information; re-encoding buffered video frames with a first quality in the first predicted area and re-encoding buffered video frames with a fourth quality outside the predicted area, wherein the first quality is better than the fourth quality; sending the re-encoded video frames to the user device.

In an example, the method may further comprise the steps of: determining a second predicted area of eye fixation based on the buffered video frames and the eye tracking information; re-encoding buffered video frames with a second quality in the second predicted area, wherein the first quality is better than the second quality and the second

quality is better than the fourth quality.

In an example, the method may further comprise the steps of: determining a predicted area of eye saccadic route based on the buffered video frames and the eye tracking information; re-encoding buffered video frames with a third quality in the predicted area of eye saccadic route, wherein the first quality is better than the third quality and the third quality is better than the fourth quality.

In another aspect of the invention, there is provided a method, for use in a communication device, of transmitting a video, the method comprising the steps of:

- receiving encoded video frames from a video server;
- 10 - decoding the received video frames;
- buffering the decoded video frames;
- receiving eye tracking information from a user device;
- determining an eye status based on the eye tracking information;
- if the eye status is in a fixation status, then re-encoding the buffered video frames
- 15 using a resolution y :

$$y = \begin{cases} \max \left(\frac{1 - e^{-t/33.3}}{a_1 (\max(g(x), s))^2 + a_2 (\max(g(x), s)) + a_3}, \hat{y} \right), x = \arg(g(x) \leq \hat{x} + s) \\ \frac{1 - e^{-t/33.3}}{a_1 (g(x) - \hat{x})^2 + a_2 (g(x) - \hat{x}) + a_3}, x = \arg(g(x) > \hat{x} + s) \end{cases}$$

wherein $t \in [t_1, +\infty)$, x is the position of a point on a screen of the user device,

$g(x)$ is a distance from the point to a center of a gaze point, t_1 is a system delay,

$\hat{y} = \frac{1 - e^{-t_1/33.3}}{a_1 s^2 + a_2 s + a_3}$, s is a diameter of the gaze point, $\hat{x}$ is derived from an equation of

20 $\hat{y} = \frac{1}{a_1 \hat{x}^2 + a_2 \hat{x} + a_3}$, and $\arg$ is a function to calculate a suitable x according to an input formula;

- if the eye status is in a saccadic status, then re-encoding the buffered video frames using a resolution y :

$$y = \max \left(\frac{1 - e^{-t/33.3}}{a_1 \max(f(x), s)^2 + a_2 \max(f(x), s) + a_3}, k_i \tilde{y}_i \right)$$

wherein $t = \frac{\Delta x}{v}$, Δx is an eye tracker resolution, v is a velocity of eye movement, x is the position of a point on a screen of the user device, $f(x)$ is a minimum distance from the point to an estimated moving trajectory of eye, $\tilde{y}_i = \frac{1 - e^{-t_1/33.3}}{a_1(\max(g(x), \hat{x}))^2 + a_2 \max(g(x), \hat{x}) + a_3}$, $g(x)$ is a distance from the point to a center

5 of a predicted fixation area i , $k_i \leq 1$ is a parameter to control the resolution of the predicted area i ;

- sending the re-encoded video frames to the user device.

Brief Description of the Drawings

10 The invention is explained in further detail, and by way of example, with reference to the accompanying drawings wherein:

Figure 1 shows a schematic view of rod and cone cell distribution;

Figure 2 shows a schematic view of network architecture;

Figure 3 shows a schematic view of equally readable chart;

15 Figure 4 shows a schematic view of visual acuity;

Figure 5 shows a schematic view of an eye tracking based video transmission system according to one embodiment of the invention;

Figure 6 shows a flow chart of a method of transmitting a video according to one embodiment of the invention;

20 Figure 7 shows a schematic view of video information adapted to the eye.

Throughout the above drawings, like reference numerals will be understood to refer to like, similar or corresponding features or functions.

Detailed Description

25 First, the human visual system modeling is described below.

The human visual acuity is related to the density of cone cells, and this is also the baseline in most studies. However, there are some other factors to affect the visual acuity, e.g. the ganglion cells. Multiple photoreceptor cells connect to one ganglion cell, and usually there are more photoreceptor cells in the connection for retina peripheral area.

Hence, the most accurate model of the human acuity is still by experiment. In “Anstis SM. A chart demonstrating variations in acuity with retinal position (Letter). Vision Res. 1974;14:589–592” (“Anstis” for short), it gives the acuity threshold model and an interesting equally readable chart. As shown in Figure. 3, when the eye is fixated in the center, all the letters should be equally readable even though the letters in the outer ring are much bigger than those in the center. This reveals on the other hand the visual acuity in the center retina is much better than the peripheral area.

The recognition threshold model is given in “Anstis”, as:

$$y = 0.046x - 0.031 \quad (1)$$

where the unit is degree and x represents the eccentricity to the fovea center. It is indicated in “Anstis” that the small negative intercept might be probably caused by experimental error, so we use general coefficients to replace the numbers. Next, the area recognition threshold should vary to the square of the eccentricity. The visual acuity is defined as the inverse of the threshold:

$$\tilde{y} = \frac{1}{a_1 x^2 + a_2 x + a_3} \quad (2)$$

When we watch a video, a picture or anything, the eye repeats the two major types of movement it performs, saccadic and fixating. When the eye starts to gaze at a new element in a picture, it needs some time to prepare and also to accumulate sufficient light to stimulate the neural. This has special meaning to video because video frame changes and the eye needs time to recognize it and if there is no change in the picture, there is no information. Hence, the eye recognizes the picture needs two conditions, time and size.

In the study of congenital nystagmus, the exponential relationship between visual acuity and the time is found in “Mario Cesarelli, Paolo Bifulco, Luciano Loffredo, Marcello Bracale, Relationship between visual acuity and eye position variability during foveations in congenital nystagmus, Documenta Ophthalmologica, July 2000, Volume 101, Issue 1, pp 59-72” (“Mario Cesarelli” for short). Hence, take this into account, the visual acuity model is:

$$\tilde{y} = \frac{1 - e^{-t/33.3}}{a_1 x^2 + a_2 x + a_3} \quad (3)$$

where t is the foveation time and 33.3 is the coefficient defined in “Mario Cesarelli” and in unit of millisecond.

Let $a_1=0.046$, $a_2=a_3=0$, Retina eccentricity from 2 to 15 degree for a duration from 0 to 100ms, the visual acuity is illustrated in Figure 4 based on equation (3). It can be seen that the visual acuity increases with time distinctly for the first tens of milliseconds. However, the main factor to affect the acuity is retina eccentricity. The visual acuity drops quickly to the floor beyond 8 degree.

The eye movement behavior can be modeled by a piecewise function corresponding to fixating and saccadic movement. The visual acuity is very low during saccadic movement as the eye turns at a speed as high as 400 degree per second. Hence, only the fixate movement needs to be considered. It is straightforward to model the movement by a Markov transition model, which could be as simple as two states. The saccadic movement is quite stereotype which can be modeled by three steps, initial preparation, fast open loop movement and final adjustment, where the second steps depends on the distance between the eye and the target, so the duration of saccadic movement duration is:

$$D(r) = \delta_1 + S(r) + \delta_2 \quad (4)$$

where r is the screen size of the display, δ_1 is preparation latency, δ_2 is the final adjustment delay, $S(r)$ is the second stage delay. Usually the total delay varies between 20ms to 200ms. The sense to model the saccadic movement is the visual acuity is low in this stage and this is useful to optimize the video transmission.

The fixation duration can be modeled by lognormal or exponential distribution as proposed in “Arthur Lugtigheid, Distributions of fixation durations and visual acquisition rates, Lugtigheid, A.J.P., 2007”. The duration is usually on the hundreds of millions level though depending on the content of the video. This means the eye might not “watch” for about 1/3 time, corresponding to 30% transmission resource saving in principle.

Then, respective embodiments of the present invention are described below.

Referring to Figure 5, the eye tracking based video transmission system comprises a video server 101, three communication devices 102a, 102b and 102c, and a user device 103. The communication device may be a base station or an eNode B, for example. The user device may be a cell phone or a tablet, for example.

In the following, a method of transmitting a video based on eye tracking techniques according to one embodiment of the invention will be described using a primary cell (Pcell) as an example of the communication device 102a, secondary cells (Scells) as an example of the communication devices 102b and 102c. Referring to Figure 6, in step S201, the Pcell 102a receives encoded video frames from the video server 101. Then, in step S202, the

Pcell 102a decodes the received video frames. For example, the video frames may be encoded by a low decoding complexity encoder in the video server 101 such that it would be easier for the Pcell 102a to do transcoding. Next, in step S203, the Pcell 102a buffers the decoded video frames.

5 Also, in step S204, the Pcell 102a receives eye tracking information from the user device 103. The eye tracking information may comprise eye gaze position and/or eye movement direction, for example. As the eye movement is quite stereotype, following saccadic and fixating pace, the fixation area can be predicted when the saccadic movement just starts. Saccadic movement has an interesting characteristic, ballistic. The fixation area
10 of interest is usually predicable, e.g., moving items, human, outstanding color objects, etc. Thus, based on the buffered video frames and the eye tracking information, in step S205, the Pcell 102a determines at least one predicted area of eye fixation. For example, the Pcell 102a may determine two predicted area of eye fixation, i.e., the first predicted area of eye fixation and the second predicted area of eye fixation.

15 For the two predicted areas of eye fixation, in step S206, the Pcell 102a re-encodes buffered video frames with a first quality in the first predicted area, re-encodes buffered video frames with a second quality in the second predicted area and re-encodes buffered video frames with a fourth quality outside the first and second predicted areas. The first quality and the second quality are better than the fourth quality. The first quality may be
20 the same as the second quality, or be better than the second quality if the first predicted area is closer to the eye. The quality may comprise resolution, for example.

 Moreover, the Pcell 102a may further determine a predicted area of eye saccadic route based on the buffered video frames and the eye tracking information. For the predicted area of eye saccadic route, the Pcell 102a re-encodes buffered video frames with a third quality
25 in the predicted area of eye saccadic route. The first quality and second quality are better than the third quality and the third quality is better than the fourth quality.

 Then, in step S207, the Pcell 102a sends the re-encoded video frames to the user device 103.

 For the multi-cell transmission, e.g., COMP, the Pcell 102a would send video content
30 to Scells 102b and 102c. In one example, the Pcell 102a may send the decoded video frames to the Scells 102b and 102c respectively. In another example, the Pcell 102a may directly send encoded video frames to the Scells 102b and 102c after receiving them from the video server 101. The Pcell 102a and the Scells 102b, 102c use a video control protocol

to ensure the video frames are re-encoded in the same way that can be combined at the user device 103. The video control protocol may define the video encoder and decoder types and its version. The video control protocol may also define the encoder parameters, e.g. quantization configuration and parameters in equation (3). The video control protocol may also include the timing information for the video frames to be re-encoded. Moreover, for each transmission, the Pcell 102a would send the eye tracking information to the Scells 102b and 102c. Then each cell could do the same video re-encoding based on the information.

Additionally, as for the complexity, the video content is distributed to the related cells and decoded and buffered for some time to absorb the delay variation, so the eye tracking based video encoder only needs to perform encoding process, not transcoding (decoding and then encoding). As the decoded video content will be buffered for some time, for example 1 second, this is very useful to smooth the decoding and encoding computation demand.

Moreover, a shorten subframe structure is used where the granularity is one slot, i.e. 0.5ms. This reduces the latency of one transmission plus one retransmission from 16ms to 8ms. Suppose the video re-encoding could be decreased to 5ms and inter base station signaling to 2ms, the total delay of the system will be 25ms. Further possible latency reduction includes shortening the HARQ retransmission cycle, reducing the re-encoding delay and eye tracker processing delay.

In another embodiment, after receiving the eye tracking information from the user device 103, the Pcell 102a determines eye status based on the eye tracking information.

If the eye status is in a fixation status, then the Pcell 102a re-encodes the buffered video frames using resolution y :

$$y = \begin{cases} \max \left(\frac{1 - e^{-t/33.3}}{a_1(\max(g(x), s))^2 + a_2(\max(g(x), s)) + a_3}, \hat{y} \right), & x = \arg(g(x) \leq \hat{x} + s) \\ \frac{1 - e^{-t/33.3}}{a_1(g(x) - \hat{x})^2 + a_2(g(x) - \hat{x}) + a_3}, & x = \arg(g(x) > \hat{x} + s) \end{cases} \quad (4)$$

wherein $t \in [t_1, +\infty)$, x is the position of a point on the screen of the user device

103, $g(x)$ is the distance from this point to the center of the gaze point, t_1 is system delay,

$$\hat{y} = \frac{1 - e^{-t_1/33.3}}{a_1 s^2 + a_2 s + a_3}, \quad s \text{ is a diameter of the gaze point, } \hat{x} \text{ is derived from the equation of}$$

$$\hat{y} = \frac{1}{a_1 \hat{x}^2 + a_2 \hat{x} + a_3}, \text{ and arg is a function to calculate the suitable } x \text{ according to an input formula.}$$

- 5 If the eye status is in a saccadic status, then the Pcell 102a re-encodes the buffered video frames using resolution y :

$$y = \max\left(\frac{1 - e^{-t/33.3}}{a_1 \max(f(x), s)^2 + a_2 \max(f(x), s) + a_3}, k_i \tilde{y}_i\right) \quad (5)$$

wherein $t = \frac{\Delta x}{v}$, Δx is an eye tracker resolution, v is a velocity of eye movement,

- 10 x is the position of a point on the screen of the user device 103, $f(x)$ is a minimum distance from this point to an estimated moving trajectory of eye.

$$\tilde{y}_i = \frac{1 - e^{-t_i/33.3}}{a_1 (\max(g(x), \hat{x}))^2 + a_2 \max(g(x), \hat{x}) + a_3}, \quad g(x) \text{ is the distance from this point to the}$$

center of the predicted fixation area i . In addition, $k_i \leq 1$ is the parameter to control the resolution of each predicted area i . In an example, for the first predicted fixation area, k_1 can be set to 1. The first predicted area is the one nearest to the previous gaze point. When the eye passes the first predicted fixation area, the second one may be upgraded to the first one and so on. The passed predicted fixation area is then deleted. There may be zero, one or multiple predicted fixation area.

Then, the Pcell 102a sends the re-encoded video frames to the user device 103.

- 20 The end to end delay depends on multiple factors. Hence, the adaptive delay compensation could be proposed. The high resolution area size is set based on equation (3). A threshold can be set if the system delay exceeds it, the system will switch to none eye tracking mode. Next, a similar slow start transmission can be used to absorb the delay variation. It is assumed the improper video encoding configuration in delay variation may affect the user experience, a target can be set e.g. the configuration should work in 99% cases, and then based on the end to end delay statistics, the base station can have the optimal configuration.

Below, simulation of the gains based on the proposed model is described.

- 30 Based on the proposed model, i.e. equation (3), the gains of different terminals are tested. First, a maximum end to end delay of 25ms is assumed. The distance between the phone and the eye is assumed to be 60cm based on the measurement of the author and his

phone. The eye movement is modeled as saccadic->fixation->saccadic... Each time a random position on the screen is selected and the eye makes saccadic movement from the current position to the next position. At the new position, the fixation is modeled by an ex-Gaussian process based on the paper of “Adrian Staub, Ashley Benatar, Individual differences in fixation duration distributions in reading, Psychonomic Bulletin & Review, December 2013, Volume 20, Issue 6, pp 1304-1311”.

The end-to end delay is considered in the modeling to ensure when the eyes starts the saccadic movement, the viewer will not notice a change of the video quality. This enlarges the high resolution circle. The effect is shown in Figure 7. The 25ms end to end delay corresponds to about 12 degree, which exhibits a two-step shaped resolution distribution figure.

The simulation results are summarized in the below table, where one can see the proposed model can save 55.5% to 80.6% resources for different terminals. Bigger screen gives higher performance gain.

Device name	Screen size	Saved resources
Iphone 6 plus	5.5'	55.5%
Ipad mini	7'	63.8%
Amazon Kindle fire HDx	9.7'	72.6%
Microsoft Surface 3	12'	80.6%

In one or more exemplary designs, the functions of the present application may be implemented using hardware, software, firmware, or any combinations thereof. In the case of implementation with software, the functions may be stored on a computer readable medium as one or more instructions or codes, or transmitted as one or more instructions or codes on the computer readable medium. The computer readable medium comprises a computer storage medium and a communication medium. The communication medium includes any medium that facilitates transmission of the computer program from one place to another. The storage medium may be any available medium accessible to a general or specific computer. The computer-readable medium may include, for example, but not limited to, RAM, ROM, EEPROM, CD-ROM or other optical disc storage devices, magnetic disk storage devices, or other magnetic storage devices, or any other medium that carries or stores desired program code means in a manner of instructions or data structures accessible by a general or specific computer or a general or specific processor.

Furthermore, any connection may also be considered as a computer-readable medium. For example, if software is transmitted from a website, server or other remote source using a co-axial cable, an optical cable, a twisted pair wire, a digital subscriber line (DSL), or radio technologies such as infrared, radio or microwave, then the co-axial cable, optical cable, twisted pair wire, digital subscriber line (DSL), or radio technologies such as infrared, radio or microwave are also covered by the definition of medium.

The various illustrative logical blocks, modules, and circuits described in connection with the disclosure herein may be implemented or performed with a general-purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general-purpose processor may be a microprocessor, but in the alternative, the processor may be any normal processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

Those of skill would further appreciate that the various illustrative logical blocks, modules, circuits, and algorithm steps described in connection with the disclosure herein may be implemented as electronic hardware, computer software, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various illustrative components, blocks, modules, circuits, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. Skilled artisans may implement the described functionality in varying ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the present disclosure.

The above depiction of the present disclosure is to enable any of those skilled in the art to implement or use the present invention. For those skilled in the art, various modifications of the present disclosure are obvious, and the general principle defined herein may also be applied to other transformations without departing from the spirit and protection scope of the present invention. Thus, the present invention is not limited to the

examples and designs as described herein, but should be consistent with the broadest scope of the principle and novel characteristics of the present disclosure.

What is claimed is:

1. A method, for use in a communication device, of transmitting a video, the method comprising the steps of:

- 5 - receiving encoded video frames from a video server;
 - decoding the received video frames;
 - buffering the decoded video frames;
 - receiving eye tracking information from a user device;
 - determining a first predicted area of eye fixation based on the eye tracking
10 information;
 - re-encoding buffered video frames with a first quality in the first predicted area
and re-encoding buffered video frames with a fourth quality outside the predicted area,
wherein the first quality is better than the fourth quality;
 - sending the re-encoded video frames to the user device.

15

2. The method of claim 1, further comprising the steps of:

- determining a second predicted area of eye fixation based on the buffered video
frames and the eye tracking information;
 - re-encoding buffered video frames with a second quality in the second predicted
20 area, wherein the first quality is better than the second quality and the second quality is
better than the fourth quality.

3. The method of claim 1, further comprising the steps of:

- determining a predicted area of eye saccadic route based on the buffered video
25 frames and the eye tracking information;
 - re-encoding buffered video frames with a third quality in the predicted area of eye
saccadic route, wherein the first quality is better than the third quality and the third quality
is better than the fourth quality.

30 4. The method of claim 1, further comprising the steps of:

- sending the encoded video frames to one or more other communication devices; or
sending the decoded video frames to the one or more other communication devices;
 - sending the eye tracking information to the one or more other communication

devices.

5. The method of claim 1, wherein the eye tracking information comprises eye gaze position and/or eye movement direction.

5

6. The method of claim 1, wherein the communication device is a base station or an eNode B.

7. A method, for use in a communication device, of transmitting a video, the method comprising the steps of:

10

- receiving encoded video frames from a video server;
- decoding the received video frames;
- buffering the decoded video frames;
- receiving eye tracking information from a user device;
- 15 - determining an eye status based on the eye tracking information;
- if the eye status is in a fixation status, then re-encoding the buffered video frames

using a resolution $\mathcal{Y}$:

$$\mathcal{Y} = \begin{cases} \max \left(\frac{1 - e^{-t/33.3}}{a_1 (\max(g(x), s))^2 + a_2 (\max(g(x), s)) + a_3}, \hat{y} \right), x = \arg(g(x) \leq \hat{x} + s) \\ \frac{1 - e^{-t/33.3}}{a_1 (g(x) - \hat{x})^2 + a_2 (g(x) - \hat{x}) + a_3}, x = \arg(g(x) > \hat{x} + s) \end{cases}$$

wherein $t \in [t_1, +\infty)$, x is the position of a point on a screen of the user device,

20 $g(x)$ is a distance from the point to a center of a gaze point, t_1 is a system delay,

$\hat{y} = \frac{1 - e^{-t_1/33.3}}{a_1 s^2 + a_2 s + a_3}$, s is a diameter of the gaze point, $\hat{x}$ is derived from an equation of

$\hat{y} = \frac{1}{a_1 \hat{x}^2 + a_2 \hat{x} + a_3}$, and $\arg$ is a function to calculate a suitable x according to an input formula;

- if the eye status is in a saccadic status, then re-encoding the buffered video frames

25 using a resolution $\mathcal{Y}$:

$$y = \max\left(\frac{1 - e^{-t/33.3}}{a_1 \max(f(x), s)^2 + a_2 \max(f(x), s) + a_3}, k_i \tilde{y}_i\right)$$

wherein $t = \frac{\Delta x}{v}$, Δx is an eye tracker resolution, v is a velocity of eye movement, x is the position of a point on a screen of the user device, $f(x)$ is a minimum distance from the point to an estimated moving trajectory of eye,

$$\tilde{y}_i = \frac{1 - e^{-t_i/33.3}}{a_1 (\max(g(x), \hat{x}))^2 + a_2 \max(g(x), \hat{x}) + a_3}, \quad g(x) \text{ is a distance from the point to a}$$

center of a predicted fixation area i , $k_i \leq 1$ is a parameter to control the resolution of the predicted area i ;

- sending the re-encoded video frames to the user device.

8. The method of claim 7, wherein the eye tracking information comprises eye gaze position and/or eye movement direction.

9. An apparatus, for use in a communication device, for transmitting a video, the apparatus comprising:

a receiver configured to receive encoded video frames from a video server and eye tracking information from a user device;

a decoder configured to decode the received video frames;

a buffer configured to buffer the decoded video frames;

a determining unit configured to determine a first predicted area of eye fixation based on the eye tracking information;

an encoder configured to re-encode buffered video frames with a first quality in the first predicted area and to re-encode buffered video frames with a fourth quality outside the predicted area, wherein the first quality is better than the fourth quality;

a transmitter configured to send the re-encoded video frames to the user device.

25

10. The apparatus of claim 9, wherein the determining unit is further configured to determine a second predicted area of eye fixation based on the buffered video frames and the eye tracking information; and the encoder is further configured to re-encode buffered video frames with a second quality in the second predicted area, wherein the first quality is

better than the second quality and the second quality is better than the fourth quality.

11. The apparatus of claim 9, wherein the determining unit is further configured to determine a predicted area of eye saccadic route based on the buffered video frames and the eye tracking information; and the encoder is further configured to re-encode buffered video frames with a third quality in the predicted area of eye saccadic route, wherein the first quality is better than the third quality and the third quality is better than the fourth quality.

12. The apparatus of claim 9, wherein the transmitter is further configured to send the encoded video frames to one or more other communication devices, or send the decoded video frames to the one or more other communication devices; and to send the eye tracking information to the one or more other communication devices.

13. The apparatus of claim 9, wherein the eye tracking information comprises eye gaze position and/or eye movement direction.

14. The apparatus of claim 9, wherein the communication device is a base station or an eNode B.

15. An apparatus, for use in a communication device, for transmitting a video, the apparatus comprising:

a receiver configured to receive encoded video frames from a video server and eye tracking information from a user device;

a decoder configured to decode the received video frames;

a buffer configured to buffer the decoded video frames;

a determining unit configured to determine an eye status based on the eye tracking information;

an encoder configured to re-encode the buffered video frames using a resolution $\mathcal{V}$, if

the eye status is in a fixation status:

$$y = \begin{cases} \max \left(\frac{1 - e^{-t/33.3}}{a_1 (\max(g(x), s))^2 + a_2 (\max(g(x), s)) + a_3}, \hat{y} \right), x = \arg(g(x) \leq \hat{x} + s) \\ \frac{1 - e^{-t/33.3}}{a_1 (g(x) - \hat{x})^2 + a_2 (g(x) - \hat{x}) + a_3}, x = \arg(g(x) > \hat{x} + s) \end{cases}$$

wherein $t \in [t_1, +\infty)$, x is the position of a point on a screen of the user device,

$g(x)$ is a distance from the point to a center of a gaze point, t_1 is a system delay,

$\hat{y} = \frac{1 - e^{-t_1/33.3}}{a_1 s^2 + a_2 s + a_3}$, s is a diameter of the gaze point, $\hat{x}$ is derived from an equation of

5 $\hat{y} = \frac{1}{a_1 \hat{x}^2 + a_2 \hat{x} + a_3}$, and $\arg$ is a function to calculate a suitable x according to an input formula;

and to re-encode the buffered video frames using a resolution y , if the eye status is in a saccadic status:

$$y = \max \left(\frac{1 - e^{-t/33.3}}{a_1 \max(f(x), s)^2 + a_2 \max(f(x), s) + a_3}, k_i \tilde{y}_i \right)$$

10 wherein $t = \frac{\Delta x}{v}$, Δx is an eye tracker resolution, v is a velocity of eye movement, x is the position of a point on a screen of the user device, $f(x)$ is a minimum distance from the point to an estimated moving trajectory of eye,

$\tilde{y}_i = \frac{1 - e^{-t_1/33.3}}{a_1 (\max(g(x), \hat{x}))^2 + a_2 \max(g(x), \hat{x}) + a_3}$, $g(x)$ is a distance from the point to a center

15 of a predicted fixation area i , $k_i \leq 1$ is a parameter to control the resolution of the predicted area i ;

a transmitter configured to send the re-encoded video frames to the user device.

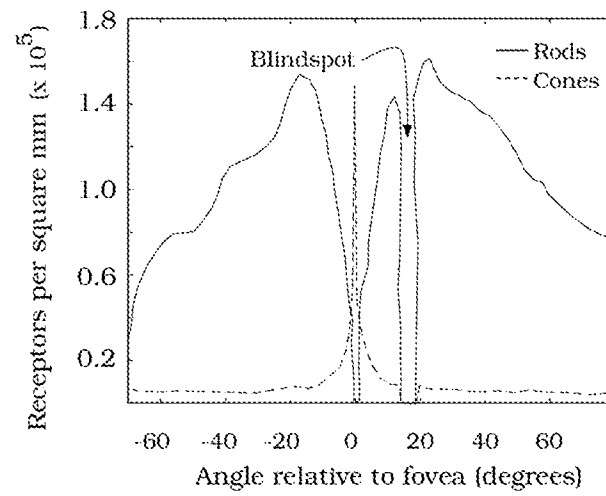


Figure 1

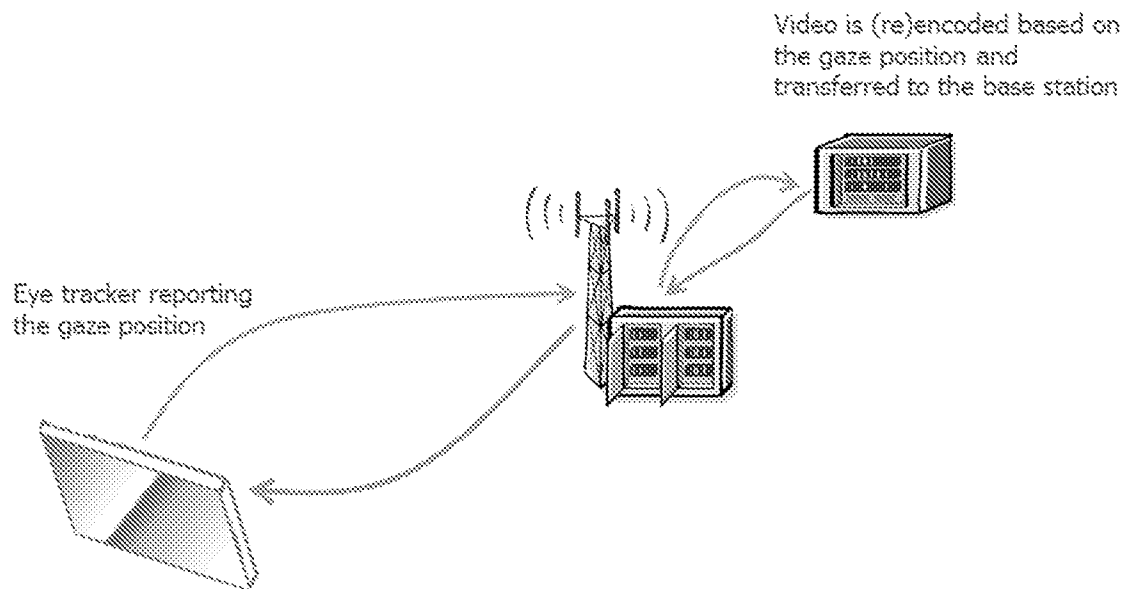


Figure 2

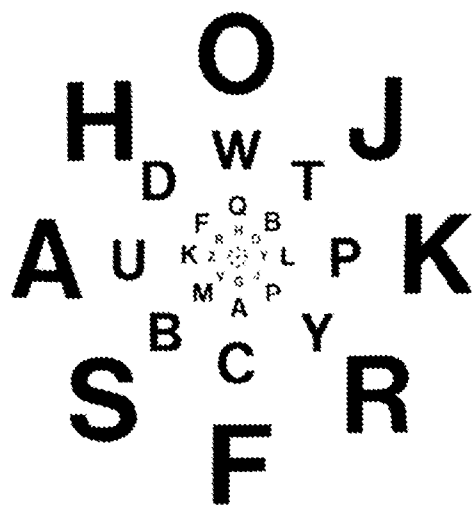


Figure 3

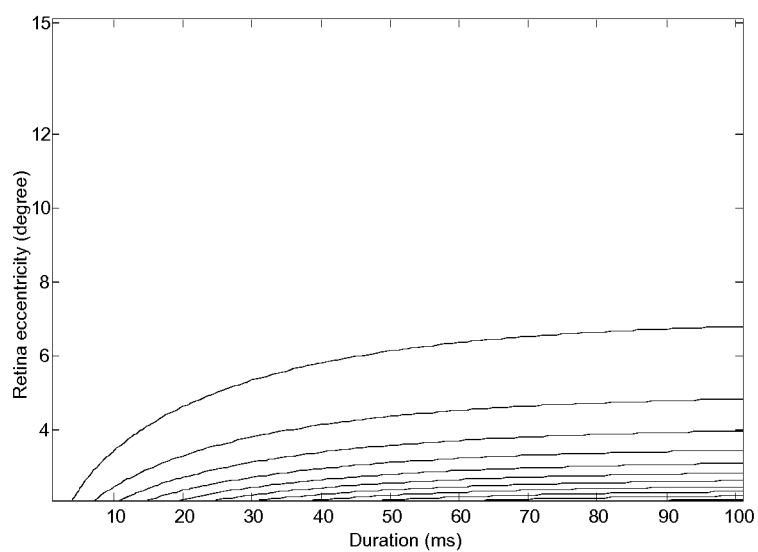


Figure 4

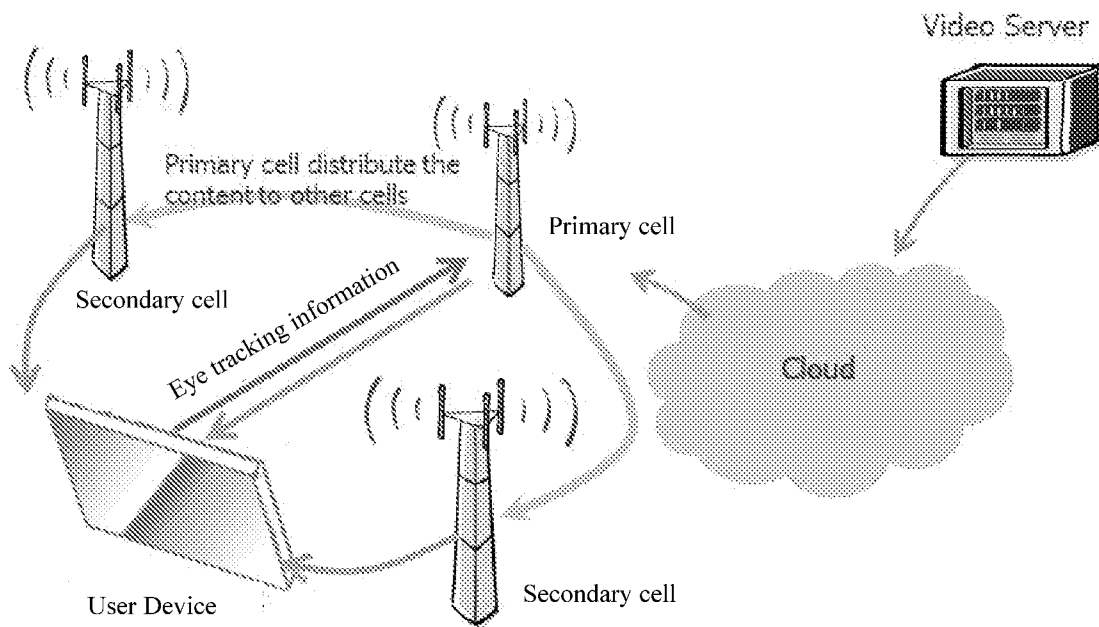


Figure 5

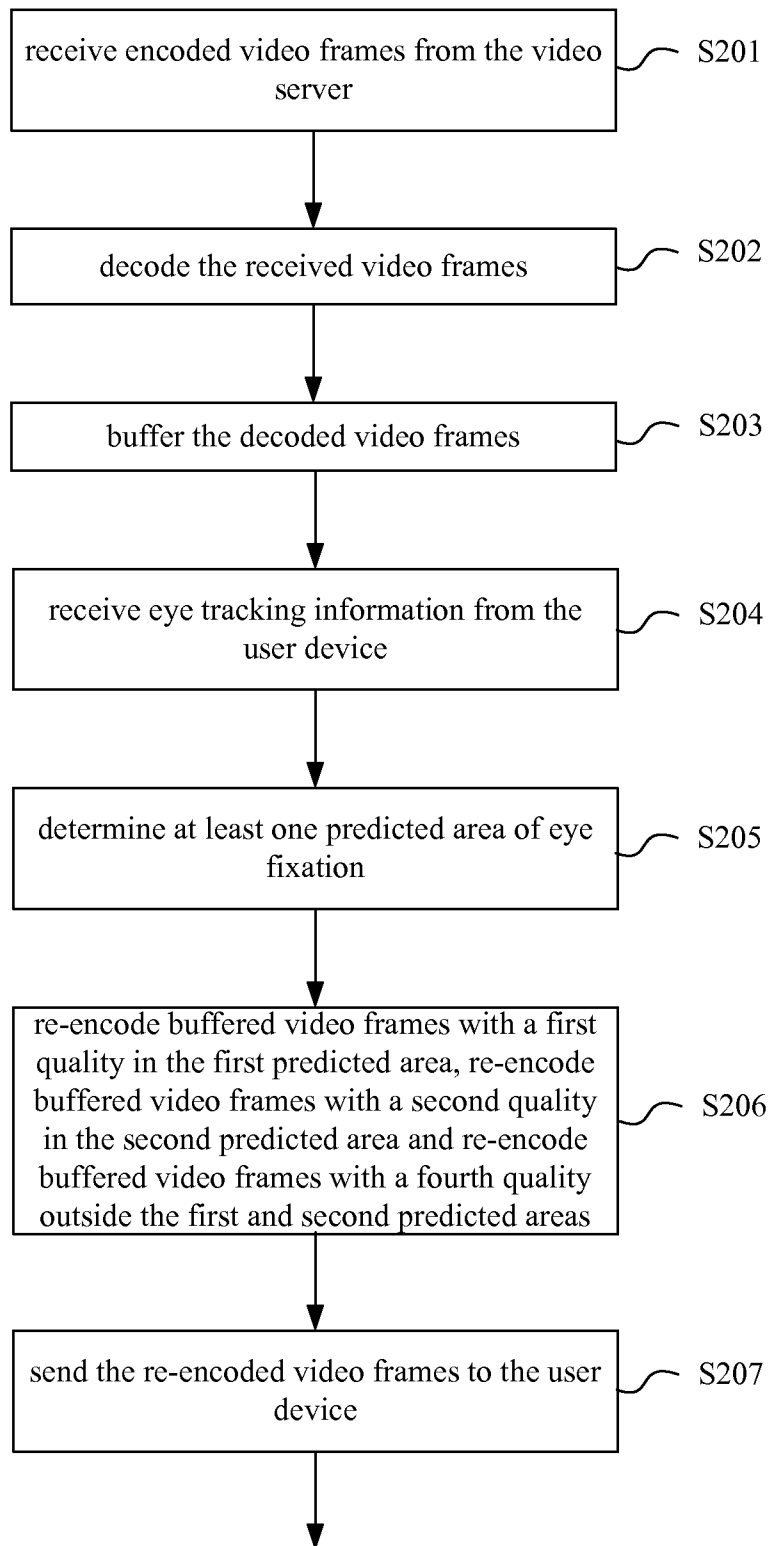


Figure 6

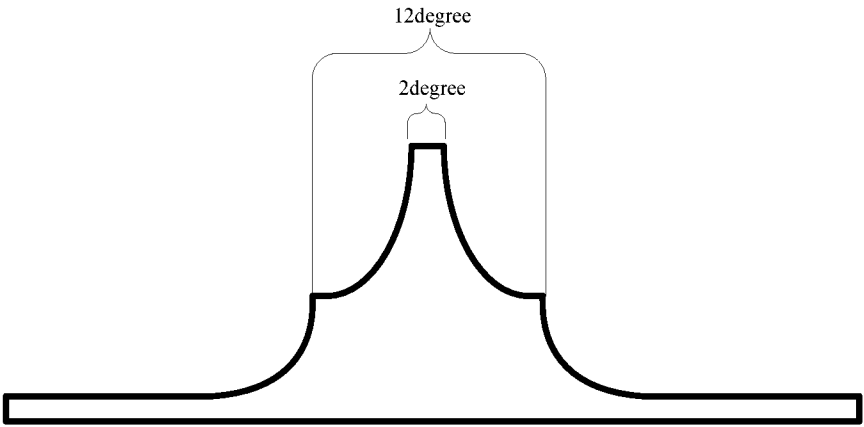


Figure 7

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2016/000262

A. CLASSIFICATION OF SUBJECT MATTER

INV. H04N21/2343 H04N21/4223 H04N21/44 H04N21/6587
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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012/146891 A1 (KALINLI OZLEM [US]) 14 June 2012 (2012-06-14)	1,2,4-6, 9,10, 12-14
Y	abstract; figure 2	3,11
A	paragraphs [0026], [0027], [0031], [0040], [0041]	7,8,15
Y	----- US 6 028 608 A (JENKINS BARRY [US]) 22 February 2000 (2000-02-22) abstract column 52, line 20 - column 53, line 13	3,11
A	----- US 4 479 784 A (MALLINSON RICHARD B [US] ET AL) 30 October 1984 (1984-10-30) abstract; figures 8,9 column 3, line 67 - column 4, line 53 column 8, lines 3-9 ----- -/-	1-15



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 May 2016

Date of mailing of the international search report

23/05/2016

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

Güvener, Cem

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2016/000262

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PARKHURST D ET AL: "Variable-Resolution Displays: A theoretical, practical and behavioral evaluation", HUMAN FACTORS, HUMAN FACTORS AND ERGONOMICS SOCIETY, SANTA MONICA, CA, US, vol. 44, no. 4, 1 January 2002 (2002-01-01), page 611629, XP002332177, ISSN: 0018-7208, DOI: 10.1518/0018720024497015 abstract -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2016/000262

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
US 2012146891	A1	14-06-2012	CN 103249352 A	14-08-2013
			CN 103559006 A	05-02-2014
			EP 2648604 A1	16-10-2013
			JP 2014508311 A	03-04-2014
			US 2012146891 A1	14-06-2012
			WO 2012078207 A1	14-06-2012

US 6028608	A	22-02-2000	NONE	

US 4479784	A	30-10-1984	NONE	
