(54) **Title:** A VIDEOCONFERENCING TERMINAL AND METHOD OF OPERATING THE SAME

Figure 5

(57) **Abstract:** A method of videoconferencing comprises displaying an image of a remote user on a display and capturing an image
of a local user at a user positron in front of the display. The at least one camera is located at a camera position behind the display.
The method comprises modifying an image to be displayed based on the camera position of the at least one camera with respect to the
display and based on the user position of the local user with respect to the display.

[Continued on next page]

WO 2020/153890 A1

A VIDEOCONFERENCING TERMINAL AND METHOD OF OPERATING THE SAME

The present invention relates to a videoconferencing terminal and method of
operating the same.

Today in the business environment there is an increasing demand not to travel
and face to face meetings are being replaced with alternatives such as
videoconferencing.   However, one problem with videoconferencing is that
making eye contact with the remote user may not be possible if the camera for
the videoconferencing terminal is located adjacent to display screen. In this
way, the local user looks at the remote user on the display screen, but the local
user will not be looking directly at the camera. This can mean the eye contact
is not maintained and this can be distracting to the users and reduce the efficacy
of videoconferencing as a viable alternative to face to face meetings.

One known arrangement is discussed in US201 2/0257004 which discloses
mounting the camera behind a transparent display screen on a mechanism for
moving the position of the camera. The camera is then moved with respect to
the position of the local user to maintain eye contact with the camera. A
problem with this arrangement is that additional mechanical components are
required to enable moving the position of the camera. This means the
videoconferencing terminal is usually dedicated to a specific room because
setup is time consuming and complex. Furthermore, movement of the camera
during a videoconference call may distract the local user if the mechanism is
actuated and makes a sound.

Another known solution is discussed in US2009/0278913 which discloses
moving the displayed image of the remote user's face until it is aligned with the
axis of the camera behind the screen. A problem with this arrangement is that
the local user may be looking at the displayed image but still not make direct
eye contact with the camera and the remote user due to parallax error.

Embodiments of the present invention aim to address the aforementioned problems.

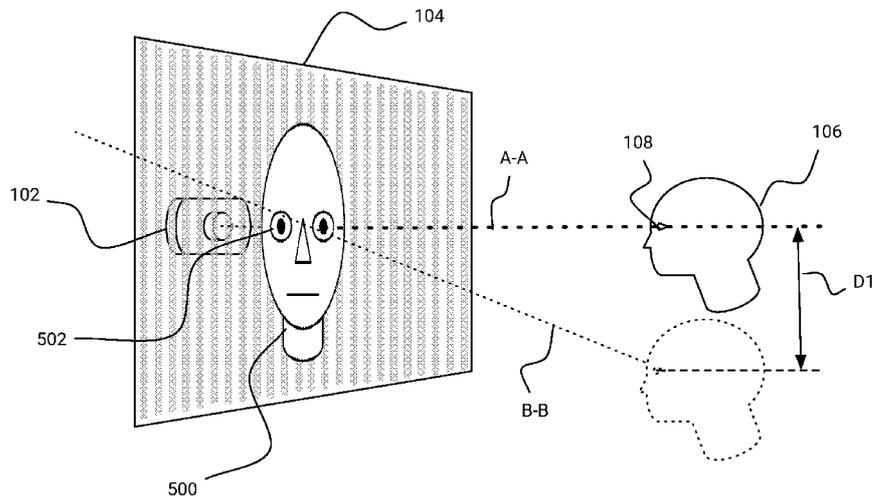According to a first aspect of the present invention there is a method of videoconferencing comprising: displaying an image of a remote user on a display; capturing an image of a local user at a user position in front of the display, with at least one camera being located at a camera position behind the display; and modifying an image to be displayed based on the camera position of the at least one camera with respect to the display and based on the user position of the local user with respect to the display.

Optionally the method comprises determining a position of the eyes of the local user with respect to the display.

Optionally the method comprises determining an axis of the at least one camera based on the position of the eyes of the local user.

Optionally the method comprises determining a position of the eyes of the remote user with respect to the display.

Optionally the method comprises determining an offset between the axis of the camera and the eyes of the remote user in a displayed image.

Optionally the modifying comprises translating the image to be displayed such that the displayed eyes of the remote user intersect with the axis.

Optionally the method comprises determining one or more pixel artifacts captured by the at least one camera from the display.

Optionally method comprises compensating the captured camera image to remove the determined one or more pixel artifacts.

Optionally the method comprises determining one or more occlusion artifacts from one or more display elements.

Optionally the method comprises compensating the captured camera image to remove the one or more occlusion artifacts.

5   Optionally the occluding display elements are out of focus.

Optionally the user position of the user and / or the camera position of the at least one camera is moveable with respect to the display.

10   Optionally the at least one camera is one or more of the following: an RGB camera or an infrared camera.

Optionally the display is transmissive to electromagnetic radiation.

15   Optionally one or more of the steps is carried out during calibration and / or one or more of the steps is carried out during operation.

According to a second aspect of the present invention there is a videoconferencing terminal comprising: a display for displaying an image of a
20   remote user; at least one camera for capturing an image of a local user at a user position in front of the display, the camera being located at a camera position behind the display; and a controller configured to modify an image to be displayed based on the camera position of the at least one camera with respect to the display and based on the user position of the local user with
25   respect to the display.

According to a third aspect of the present invention there is a method of videoconferencing comprising: displaying an image of a remote user on a display; capturing an image of a local user at a user position in front of the
30   display, with at least one camera being located at a camera position behind the display; and modifying an image to be displayed based on the camera position of the at least one camera with respect to the display.

According to a fourth aspect of the present invention there is a videoconferencing terminal comprising: a display for displaying an image of a remote user; at least one camera for capturing an image of a local user at a user position in front of the display, the camera being located at a camera

5   position behind the display; and a controller configured to modify an image to be displayed based on the camera position of the at least one camera with respect to the display.

Various other aspects and further embodiments are also described in the

10   following detailed description and in the attached claims with reference to the accompanying drawings, in which:

Figure 1a shows a schematic perspective view of a videoconferencing terminal;

Figure 1b shows a schematic side view of a videoconferencing terminal;

15   Figure 2 shows a schematic cross-sectional side view of a videoconferencing terminal;

Figure 3 shows a schematic perspective view of a videoconferencing terminal;

Figure 4a, 4b, and 4c show a schematic view of a captured image by a videoconferencing terminal;

20   Figure 5 shows a schematic perspective view of a videoconferencing terminal;

Figure 6 shows a schematic perspective view of a videoconferencing terminal;

Figure 7 shows a schematic view of a videoconferencing terminal;

Figure 8 shows a flow diagram of the operation of a videoconferencing terminal; and

25   Figure 9 shows a flow diagram of the operation of a videoconferencing terminal.

Figure 1a shows a perspective view of a schematic perspective view of a videoconferencing terminal 100. The videoconferencing terminal 100 comprises at least one camera 102 positioned behind a display 104. The

30   display 104 is configured to display an image 500 of a remote user to a local user 106 who is positioned in front of the display 104.

The local user 106 is positioned in close proximity to the videoconferencing terminal 100 and the camera 102 is configured to capture on or more images,

and or videos of the local user 106. For example, the local user 106 is in the same room as the videoconferencing terminal 100. In contrast, the remote user is not in close proximity to the videoconferencing terminal 100 or the local user 106 and the video stream and / or images of the local user 106 are transmitted

5   to a videoconferencing terminal (not shown) associated with the remote user. In the embodiments described with reference to the Figures there are two users a local user 106 and a remote user. In other embodiments (not shown), there may be any number of local users 106 and remote users on the videoconference call.

10

The process of receiving and transmitting video and image data between videoconferencing terminals 100 is carried out with respect to known techniques and will not be discussed in any further detail.

15   In some embodiments, the remote user has an identical videoconferencing terminal 100 to the videoconferencing terminal 100 of the local user 106. However, this is not necessary and only one of the users participating in the videoconference can have the videoconferencing terminal 100 according to the embodiments described in reference to the Figures. In a preferred

20   embodiment, all users participating in the videoconference have a videoconferencing terminal 100 according to the embodiments.

Figure 1b shows a schematic side view of a videoconferencing terminal 100. The camera 102 comprises an axis A-A which is in some embodiments

25   arranged substantially perpendicular to the plane of the surface of the display 104. Figure 1b shows that the axis A-A is in alignment with the eyes 108 of the local user 106. In this way, axis A-A is an "eye-contact" axis. In this arrangement, the local user 106 is looking directly along the axis of the camera 102. This means that the camera 102 will capture an image or a video of the

30   local user 106 looking directly at the camera 102. This means the remote user will receive an image of the local user 106 with the eyes 108 of the local user in the correct direction to simulate a face to face meeting. In some alternative embodiments, the camera 102 is moveable with respect to the display 104 and

the axis of the camera 102 can be positioned at an angle with respect to the plane of the display 104.

Whilst Figures 1a and 1b show one camera 102, in some embodiments there can be a plurality of cameras 102 for capturing and image or a video of a plurality of local users 106 or for capturing an image of a video of a large room. The embodiments described hereinafter are only described in reference to using one camera, but some embodiments use a plurality of cameras 102 are used instead. The camera 102 as shown in Figure 1 is static and positioned in the centre of the display 104. However, in some embodiments, the camera 102 is moveable with respect to the display 104.

The display 104 in some embodiments is a transparent OLED display 104. The display 104 is substantially planar and can be any suitable size for the videoconferencing call. In other embodiments any other suitable transparent display can be used. For example, infrared cameras (not shown) can be used and the infrared cameras can see the local user 106 through the display 104. In this way, the display 104 is transmissive to electromagnetic radiation which can be in the visible spectrum, near visible, infrared or ultraviolet or any other suitable frequency of electromagnetic radiation.

Turning to Figure 7, the videoconferencing terminal 100 will be described in further detail. Figure 7 shows a schematic view of a videoconferencing terminal 100 according to some embodiments.

As previously mentioned, the videoconferencing terminal 100 comprises a camera 102 and a display 104. The videoconferencing terminal 100 selectively controls the activation of the camera 102 and the display 104. As shown in Figure 7, the camera 102 and the display 104 are controlled by a camera controller 702 and a display controller 704 respectively.

The videoconferencing terminal 100 comprises a videoconferencing controller 700. The videoconferencing controller 700, the camera controller 702 and the

display controller 704 may be configured as separate units, or they may be incorporated in a single unit.

The videoconferencing controller 700 comprises a plurality of modules for
5   processing the videos and images received from a remotely from an interface 706 and videos and images captured locally. The interface 706 and the method of transmitted and receiving videoconferencing data is known and will not be discussed any further. In some embodiments, the videoconferencing controller 700 comprises a face detection module 710 for detecting facial features and an
10   image processing module 712 for modifying an image to be displayed on the display 104. The face detection module 710 and the image processing module 712 will be discussed in further detail below.

One or all of the videoconferencing controller 700, the camera controller 702
15   and the display controller 704 may be at least partially implemented by software executed by a processing unit 714. The face detection modules 710 and the image processing modules 712 may be configured as separate units, or they may be incorporated in a single unit. One or both of the modules 710, 712 may be at least partially implemented by software executed by the processing unit
20   714.

The processing unit 714 may be implemented by special-purpose software (or firmware) run on one or more general-purpose or special-purpose computing devices. In this context, it is to be understood that each "element" or "means"
25   of such a computing device refers to a conceptual equivalent of a method step; there is not always a one-to-one correspondence between elements/means and particular pieces of hardware or software routines. One piece of hardware sometimes comprises different means/elements. For example, a processing unit 714 may serve as one element/means when executing one instruction but
30   serve as another element/means when executing another instruction. In addition, one element/means may be implemented by one instruction in some cases, but by a plurality of instructions in some other cases. Naturally, it is conceivable that one or more elements (means) are implemented entirely by analogue hardware components.

The processing unit 714 may include one or more processing units, e.g. a CPU ("Central Processing Unit"), a DSP ("Digital Signal Processor"), an ASIC ("Application-Specific Integrated Circuit"), discrete analogue and/or digital components, or some other programmable logical device, such as an FPGA ("Field Programmable Gate Array"). The processing unit 714 may further include a system memory and a system bus that couples various system components including the system memory to the processing unit. The system bus may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures. The system memory may include computer storage media in the form of volatile and/or non-volatile memory such as read only memory (ROM), random access memory (RAM) and flash memory. The special-purpose software and associated control parameter values may be stored in the system memory, or on other removable/non-removable volatile/non-volatile computer storage media which is included in or accessible to the computing device, such as magnetic media, optical media, flash memory cards, digital tape, solid state RAM, solid state ROM, etc. The processing unit 714 may include one or more communication interfaces, such as a serial interface, a parallel interface, a USB interface, a wireless interface, a network adapter, etc, as well as one or more data acquisition devices, such as an A/D converter. The special-purpose software may be provided to the processing unit 714 on any suitable computer-readable medium, including a record medium, and a read-only memory.

Figures 1a and 1b show the videoconferencing terminal 100 which is operating optimally and the remote user and the local user 106 can make eye contact. However, calibration of the videoconferencing terminal 100 and dynamic modification of the displayed image 500 may be required in order for the local user 106 to experience a good connected feel during a video conference call.

Calibration of the videoconferencing terminal 100 will now be discussed in reference to Figures 2, 3, 4a, 4b, 4c and 9. Figure 2 shows a schematic cross-sectional side view of a videoconferencing terminal. Figure 3 shows a schematic perspective view of a videoconferencing terminal. Figure 4a, 4b, and

4c show a schematic view of a processing sequence for a captured camera image 400 on the videoconferencing terminal 100. Figure 9 shows a flow diagram of the operation of a videoconferencing terminal.

5   During operation of the camera 102 and the display 104 the videoconferencing controller 700 can optionally interleave operation of the camera 102 and the display 104. In this way, the camera 102 and the display 104 sequentially operate so that the camera 102 captures an image of the local user 106 when the display 104 is off. Likewise, the camera 102 is not capturing an image when
10   the display 106 is displaying an image. For example, the camera 102 can be turned off or the shutter is closed when not capturing an image of the local user 106. This means that the camera 102 takes an image when the display 104 is dark. As mentioned previously, in some embodiments the display is an OLED display. The OLED display has a low persistence, and this reduces pixel
15   artifacts 300 which are received and captured by the camera 102 originating from the display 104.

However, the camera 102 may still receive light from pixel artifacts 300 from the display 104. This can be a function of the display image 500 being
20   displayed on the display 104 as well as the properties of the display 104 itself. Turning to Figure 2, the display 104 will be described in further detail. The display 104 comprises an LED matrix 200 of selectively operable pixels 202. For the purposes of clarity, only one pixel 202 has been labelled in Figure 2. The LED matrix 200 can comprise any number of pixels 202 to achieve the
25   required resolution for the videoconferencing call. An optically transmissive cover 204 such as a glass sheet, a transparent film or another clear medium is placed over the LED matrix 200. In some circumstances, one or more light rays B can be reflected back from the optically transmissive cover 204 towards the camera 102.

30

In some embodiments, the videoconferencing controller 700 is configured to determine one or more pixel artifacts 300 captured by the at least one camera 102 from the display 104 as shown in 900 of Figure 9. Once the pixel artifacts 300 have been determined, the videoconferencing controller 700 is configured

to compensate the captured camera image 400 to remove the mapped one or more pixel artifacts 300.   Figure 3 shows a perspective schematic representation of the video conferencing terminal 100.   The display 104 is shown with exemplary pixel artifacts 300 and occlusion artifacts 302 on the display 104. Figure 4a shows the captured camera image 400 including a local user captured image 406 of the local user 106 together with the pixel artifacts 300 and / or occlusion artifacts 302. Whilst the pixel artifacts 300 and occlusion artifacts 302 are represented by a series of vertical lines, the pixel artifacts 300 and occlusion artifacts 302 can be any distribution across the display 104.

In some embodiments, in order to compensate for the pixel artifacts 300 from the display 104 in the captured camera image 400, the contribution from each pixel 202 of the display 104 in the captured camera image 400 is determined as shown in step 900. Optionally, this is achieved with per-pixel information of the LED matrix 200 which maps the pixel output to the contribution as a pixel artifact map 402 in the captured camera image 400.

The pixel output is a function of the digital RGB (red green blue) colour output of the display image 500 and properties of the display 104. The videoconferencing controller 700 uses information relating to displayed image 500 and the display 104 properties and determines each display pixel's contribution in the captured camera image 400.   In this way, the videoconferencing controller 700 determines a pixel artefact map 402 as shown in Figure 4b.

The videoconferencing controller 700 then subtracts the contribution of all display pixels 202 in the pixel artifact map 402 to obtain a compensated camera image 404 as shown in Figure 4c and step 902 of Figure 9.   The videoconferencing controller 700 then determines the compensated camera image 404 as it would have looked without any light contribution of pixel artifacts 300 from the pixels 202. The compensated camera image 404 comprises the local user captured image 406 as well.

The videoconferencing controller 700 receives information relating to the digital RGB colours of the display image 500 sent to the display 104. This means that the information relating to the digital RGB colours are directly available to the videoconferencing controller 700 for carrying out the compensation algorithm as shown in Figure 9.

In some embodiments, the videoconferencing controller 700 optionally determines the display 104 properties can be determined in a calibration step. In the calibration step the videoconferencing controller 700 selectively controls the LED matrix 200 to light up each pixel 202 individually, at different illumination levels, to learn the mapping from digital RGB colour output to contribution in the captured camera image 400.

After the display pixel artifacts 300 have been removed, in some circumstances the captured camera image 400 may still have occlusion artifacts 302 in the captured camera image 400 from elements of the display 104. The occlusion artifacts 302 arise from one or more elements of the display 104 in front of the camera 102 which blocks light from the local user 106. The occlusion artifacts 302 can be described as having an occlusion factor between 0.0 and 1.0 wherein 0.0 indicates total occlusion and 1.0 indicates no occlusion.

In some embodiments, the videoconferencing controller 700 determines the occlusion factors of the occlusion artifacts 302 in a calibration step, when the camera 102 is directed at a uniform (e.g., all white) and evenly illuminated target. This means that the camera image pixel levels are uniform if no occlusions artifacts 302 are present.

Figure 4b also represents the determined occlusion artifact map 408 of occlusion artifacts 302 on the display occluded image after the calibration step. As mentioned above, in the calibration step the camera 102 is looking at a smooth white surface. The videoconferencing controller 700 determines the maximum pixel level of a particular pixel 202 in the LED matrix 200. For each other pixel in the LED matrix 200, the videoconferencing controller 700 divides

its pixel value by the maximum pixel value to get the occlusion factor for each particular pixel 200.

In this way, the videoconferencing controller 700 sets, a notional "correct" level
to be the one of the maximum pixels. The videoconferencing controller 700 implicitly assumes that the maximum pixel is unoccluded. If this is not the case, the effect is a uniformly darker image, but this is not an effect that is apparent to the local user 106, and not experienced as a significant artifact. Accordingly, the videoconferencing controller 700 determines on or more occlusion artifacts
302 as shown in step 904 of Figure 9.

In a similar way, it may be the case that the target and illumination properties during calibration are such that the ideal, unoccluded, image is not uniform, but has slight variations. Typically, such variations are of low spatial frequency, and
will cause low frequency artifacts in the compensated results that are either not noticeable at all to the user or not experienced as significant artifacts to the local user 106.

The videoconferencing controller 700 assumes that occlusions are not severe
enough to completely occlude parts of a camera pixel (not shown) (e.g. occlusion factor 0.0), but only occlude parts of the incoming light, for each camera pixel. In some embodiments, at least some of the occluding display elements are out-of-focus. In some embodiments, the optics of the camera 102 are designed to keep occluding display elements are out-of-focus.

The videoconferencing controller 700 then multiples the "correct", "unoccluded",
pixel value is multiplied by. 0.0 gives total occlusion and 1.0 no occlusion. In this way by having information relating to the occlusion factor for each pixel 202, the videoconferencing controller 700 can determine the compensated camera
image 404 according to step 906 in Figure 9 by dividing each pixel value by its occlusion factor, obtaining an unoccluded and compensated camera image 404 as shown in Figure 4c.

Optionally the steps 900, 902 relating to the compensation of the pixel artifacts 300 and steps 904, 906 relating to the compensation of the occlusion artifacts 302 can be carried out in a different order than as show in Figure 9. Furthermore, optionally one, some or all of the steps 900, 902 relating to the

5   compensation of the pixel artifacts 300 and steps 904, 906 relating to the compensation of the occlusion artifacts 302 can be omitted. For example, compensation for pixel artifacts 300 can be omitted. Likewise, additionally or alternatively, compensation for occlusion artifacts 302 can be omitted.

10   Steps 900, 902, 904, 906 are dependent on the position of the camera 102 with respect to the display 104. Accordingly, the compensation of the pixel artifacts 300 and compensation for occlusion artifacts 302 is based on the relative position of the camera 102 with respect to the display 104. This means that if the camera 102 moves with respect to the display 104, one or more of the steps

15   as shown in Figure 9 are repeated to recalibrate the video conferencing terminal 100. In this way, videoconferencing controller 700 modifies an image based on the camera position of the at least one camera 102 with respect to the display.

Another embodiment will now be described in reference to Figures 5, 6 and 8.
20   Figure 5 and 6 show a schematic perspective view of a videoconferencing terminal 100 and Figure 8 shows a flow diagram of the operation of a videoconferencing terminal. Optionally, the method steps discussed with respect to Figure 9 can be used together with the method steps in Figure 8, but this is not necessary.

25

Turning to Figure 5, again the axis A-A of the camera 102 is in alignment with the eyes 108 of the local user 106. In Figure 5 the eyes 108 of the local user 106 are aligned with eyes 502 of the displayed image 500 of the remote user. Accordingly, the local user 106 and the remote user are able to make direct eye

30   contact.

As can be seen from Figure 5, if the local user 106 moves with respect to the display 104, the local user 106 is no longer aligned with the axis A-A of the camera 102. Figure 5 shows one possible new position of the local user 106

represented by a dotted outline. In the new position, the local user's 106 line of sight B-B is still focused on the eyes 502 of the displayed image 500 of the remote user. However, the local user 106 is no longer looking directly at the camera 102 due the parallax error introduced by the local user 106 also moving

5   with respect to the camera 102. This means that the captured camera image 400 of the local user 106 will not be looking directly at the camera 102.

However, Figure 6 shows the local user 106 in the new position shown in Figure 5. Here the position of the local user 106 is offset by a distance D1 from the

10 axis A-A of the camera 102. This means that the eyes 108 of the local user 106 have moved from the axis A-A by a distance D1. Specifically, as shown in Figure 6, the local user 106 is lower than the axis A-A. However, in other embodiments the local user 106 can be offset from the axis A-A of the camera 102 in any direction. For example, the local user 106 may have moved

15 sideways with respect to the axis A-A or may be standing and the eyes 108 of the local user are above the axis A-A.

The videoconferencing controller 700 sends the image 500 of the remote user to be displayed to the face detection module 710. The face detection module

20 710 determines the position of the eyes 502 of the displayed image 500 of the remote user as shown in step 800 in Figure 8. The face detection module 710 uses feature detection on an image 500 of the remote user to detect where the eyes 502 of the displayed image 500 of the remote user. The face detection module 710 then sends position information of the eyes 502 of the displayed

25 image 500 of the remote user to the videoconferencing controller 700.

Then the videoconferencing controller 700 determines the position of the camera 102 with respect to the display 104. If the camera 102 is fixed with respect to the display 104, the videoconferencing controller 700 can store the

30 position of the camera 102 and the axis of the camera 102 in memory. Alternatively, the videoconferencing controller 700 can determine the relative position of the camera 102 with respect to the display 104 based on movement information of the camera 102. For example, the videoconferencing controller 700 determines the position of the camera 102 from servo information on a

mechanism for moving the camera 102. Alternatively, the videoconferencing controller 700 determines the position of the camera 102 based on reference points in the captured camera image 400. For example, a reference point could be a QR code fixed to a wall behind the local user 106. In this way, the

5   videoconferencing controller 700 determines the position and orientation of the camera 102 and the axis A-A of the camera 102 as shown in step 802 of Figure 8.

Then the videoconferencing controller 700 sends a captured camera image 400

10   of the local user 106 to the face detection module 710. The face detection module 710 determines the position of the eyes 108 of the local user in the image 400 as shown in step 804 in Figure 8. The face detection module 710 uses feature detection on the image 400 of the local user 106 to detect where the eyes 108 are in the image 400. This is similar to the step 800 in Figure 8

15   for determining the position of the eyes 502 of the displayed image 500 of the remote user.

The videoconferencing controller 700 then determines a position of the eyes 108 of the local user 106 with respect to the display 104. Based on the

20   determined position of the camera 102, the videoconferencing controller 700 determines an offset D1 between the position of the eyes 108 of the local user 106 and an axis A-A of the at least one camera 102. In this way, the videoconferencing controller 700 determines how much the local user 106 has moved from the axis A-A of the camera 102. This means that the

25   videoconferencing controller 700 determines, a new axis A'-A' of the camera 102 based on a light ray from the new position of the local user 106 and the position of the camera 102. Accordingly, A'-A' is the new eye contact axis.

The videoconferencing controller 700 determines a position of the eyes 502 of

30   the displayed image 500 of the remote user with respect to the display 104. That is, the videoconferencing controller 700 determines where the image 500 would be positioned on the display 104 with no modification to the image 500.

The videoconferencing controller 700 then determines whether the position of the eyes 502 of the displayed image 500 of the remote user is offset D2 from the new axis A'-A based on the new position of the local user 106. If the videoconferencing controller 700 determines that the displayed image 500 is

5   offset greater than a predetermined threshold, the videoconferencing controller 700 sends an instruction to the image processing module 712 to modify the image 500 as show in step 806 in Figure 8. In Figure 6, the eyes 502 of the displayed image 500 of the remote user are translated downwards by a distance of D2 to intersect the new axis A'-A'.

10

In some embodiments, the videoconferencing controller 700 instructs the image processing module 712 to modify the image 500 when the new position of the local user 106 requires the local user 106 to adjust their line of sight through an arc having an angle greater than 10 degrees. In some embodiments, the image

15   processing module 712 to modifies the image 500 when the local user 106 adjusts their line of sight through an arc having an angle greater than 10 degrees in a horizontal and / or a vertical directions from the axis A-A. In this way, if the local user 106 is required to move their head or the eyes 108 of the local user to maintain eye contact with the eyes 502 of the displayed image 500

20   of the remote user, the videoconferencing controller 700 modifies the image 500 and returns modified image 600. This means that there is no parallax error that prevents direct eye contact between the local user 106 and the remote user because the videoconferencing controller 700 modifies an image based on the position of the camera 102 and the local user 106 with respect to the displayed

25   image 500.

In some embodiments, the videoconferencing controller 700 sends an instruction that a co-ordinate corresponding to the centre of the eyes 502 of the displayed image 500 of the remote user is translated to a new position. The

30   image processing module 712 returns a modified image 600 to the videoconferencing controller 700. The modified image 600 of the remote user is shown in Figure 6.

In this way, the eyes 502 of the displayed image 500 of the remote user are moved to intersect with the new axis A'-A'. In this way, the image processing module 712 modifies the image 500 such that the eyes 502 of the displayed image 500 of the remote user intersect with the new axis A'-A'. In the new

5   position, the local user's 106 line of sight B-B is focused on the eyes 502 of the displayed image 500 of the remote user and aligned with the new axis A'-A'. In some embodiments, the image processing module 712 modifies the image 500 by translating, scaling, or transforming or any other suitable image modification to move the position of the eyes 502 of the displayed image 500 of the remote

10   user.

In this way, videoconferencing controller 700 modifies an image based on the camera position of the at least one camera 102 with respect to the display 104 and on the user position of the local user 106 with respect to the display 104.

15

As mentioned above, in some embodiments, there is only one video conferencing terminal 100 with a videoconferencing controller 700 and the image processing module 712 as discussed with reference to the previous embodiments. In these embodiments, the videoconferencing controller 700

20   performs the image processing as discussed with reference to embodiments as shown in the Figures e.g. Figures 8 and 9 for both the local video conferencing terminal 100 and the remote video conferencing terminal. This means that the advantages of the invention can be achieved for both sides of the video conference with only one video conferencing terminal 100, e.g. the local video

25   conferencing terminal 100, according to the present invention.

When the local video conferencing terminal 100 is modifying the image for both the local and the remote video conferencing terminals 100, the videoconferencing controller 700 performs the methods described with

30   references to the Figures for both local and the remote video conferencing terminals. The local videoconferencing controller 700 then sends instructions for modifying the displayed image to the remote video conferencing terminal. For example, translation coordinates for modifying the displayed image on the

remote video conferencing terminal are sent by the local video conferencing controller 700 to the remote video conferencing terminal 100.

In another embodiment two or more embodiments are combined. Features of one embodiment can be combined with features of other embodiments.

Embodiments of the present invention have been discussed with particular reference to the examples illustrated. However it will be appreciated that variations and modifications may be made to the examples described within the scope of the invention.

Claims

1.   A method of videoconferencing comprising:

displaying an image of a remote user on a display;

capturing an image of a local user at a user position in front of the display, with at least one camera being located at a camera position behind the display; and

modifying an image to be displayed based on the camera position of the at least one camera with respect to the display and based on the user position of the local user with respect to the display.

2.   A method according to claim 1 wherein the method comprises determining a position of the eyes of the local user with respect to the display.

3.   A method according to claims 1 or 2 wherein the method comprises determining an axis of the at least one camera based on the position of the eyes of the local user.

4.   A method according to any of the preceding claims wherein the method comprises determining a position of the eyes of the remote user with respect to the display.

5.   A method according to claim 4 wherein the method comprises determining an offset between the axis of the camera and the eyes of the remote user in a displayed image.

6.   A method according to claim 5 wherein the modifying comprises translating the image to be displayed such that the displayed eyes of the remote user intersect with the axis.

7.   A method according to any of the preceding claims wherein the method comprises determining one or more pixel artifacts captured by the at least one camera from the display.

8.   A method according to claim 7 wherein the method comprises compensating the captured camera image to remove the determined one or more pixel artifacts.

5   9.   A method according to any of the preceding claims wherein the method comprises determining one or more occlusion artifacts from one or more display elements.

10.   A method according to claim 9 wherein the method comprises
10   compensating the captured camera image to remove the one or more occlusion artifacts.

11.   A method according to claims 9 or 10 wherein the occluding display elements are out of focus.
15

12.   A method according to any of the preceding claims wherein the user position of the user and / or the camera position of the at least one camera is moveable with respect to the display.

20   13.   A method according to any of the preceding claims wherein the at least one camera is one or more of the following: an RGB camera or an infrared camera.

14.   A method according to any of the preceding claims wherein the display
25   is transmissive to electromagnetic radiation.

15.   A method according to any of the preceding claims wherein one or more of the steps is carried out during calibration and / or one or more of the steps is carried out during operation.
30

16.   A videoconferencing terminal comprising:
        a display for displaying an image of a remote user;

20

at least one camera for capturing an image of a local user at a user position in front of the display, the camera being located at a camera position behind the display; and

5   a controller configured to modify an image to be displayed based on the camera position of the at least one camera with respect to the display and based on the user position of the local user with respect to the display.

17.   A method of videoconferencing comprising:

displaying an image of a remote user on a display;

10   capturing an image of a local user at a user position in front of the display, with at least one camera being located at a camera position behind the display; and

modifying an image to be displayed based on the camera position of the at least one camera with respect to the display.

15

18   A method according to claim 17 wherein the method comprises determining one or more pixel artifacts captured by the at least one camera from the display.

20   19.   A method according to claim 18 wherein the method comprises compensating the captured camera image to remove the determined one or more pixel artifacts.

20.   A method according to any of claims 17 to 19 wherein the method
25   comprises determining one or more occlusion artifacts from one or more display elements.

21.   A method according to claim 20 wherein the method comprises compensating the captured camera image to remove the one or more occlusion
30   artifacts.

22.   A videoconferencing terminal comprising:

a display for displaying an image of a remote user;

at least one camera for capturing an image of a local user at a user position in front of the display, the camera being located at a camera position behind the display; and

a controller configured to modify an image to be displayed based
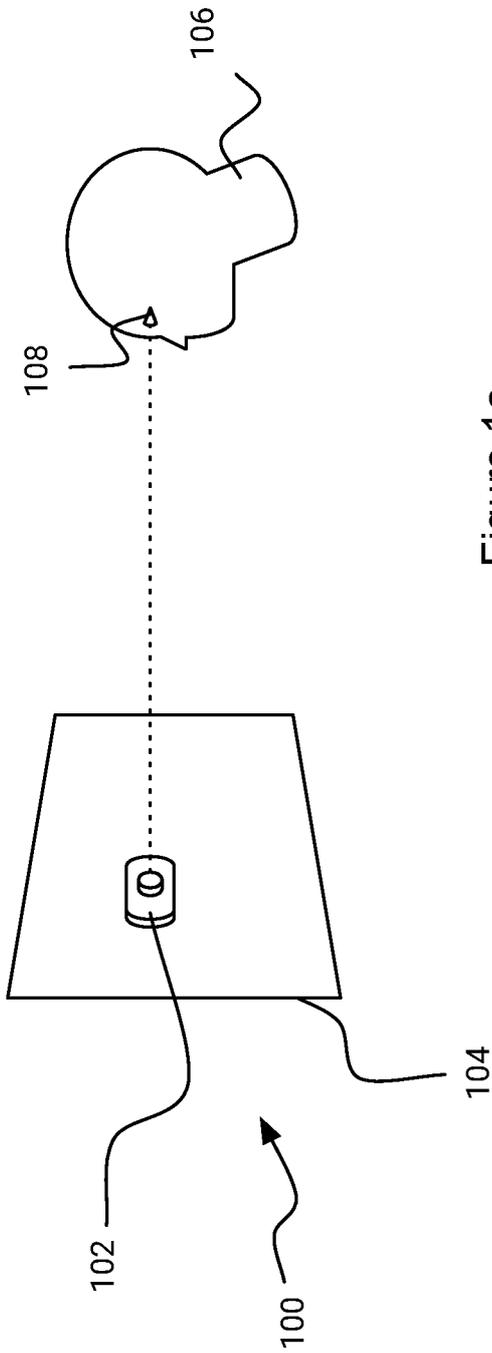5 on the camera position of the at least one camera with respect to the display.
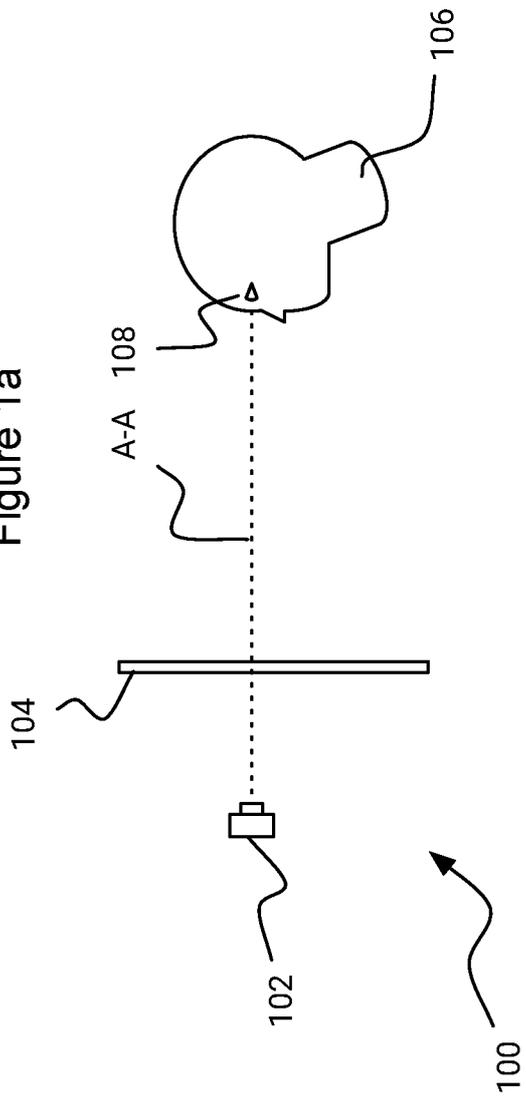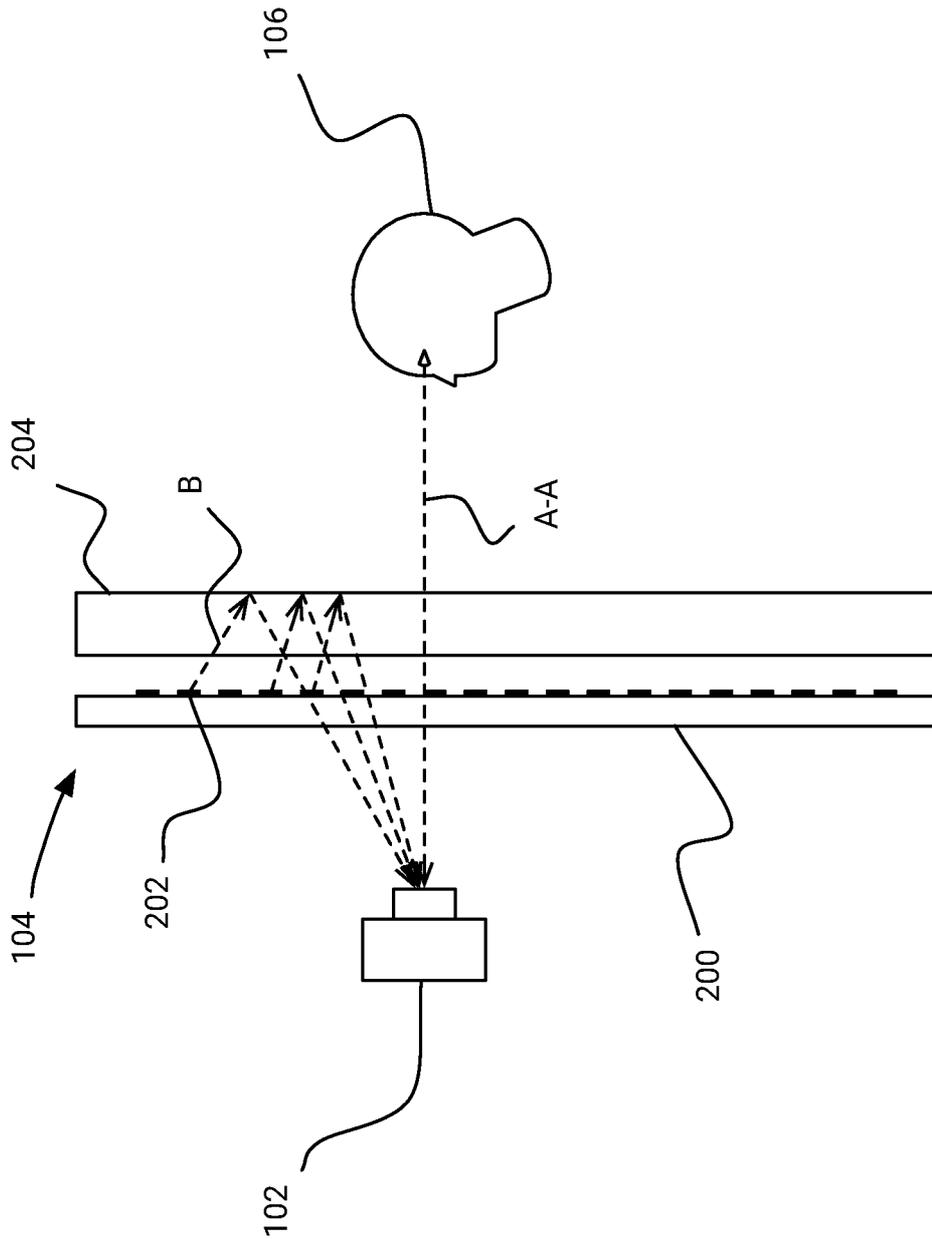
Figure 1a

Figure 1b

Figure 2
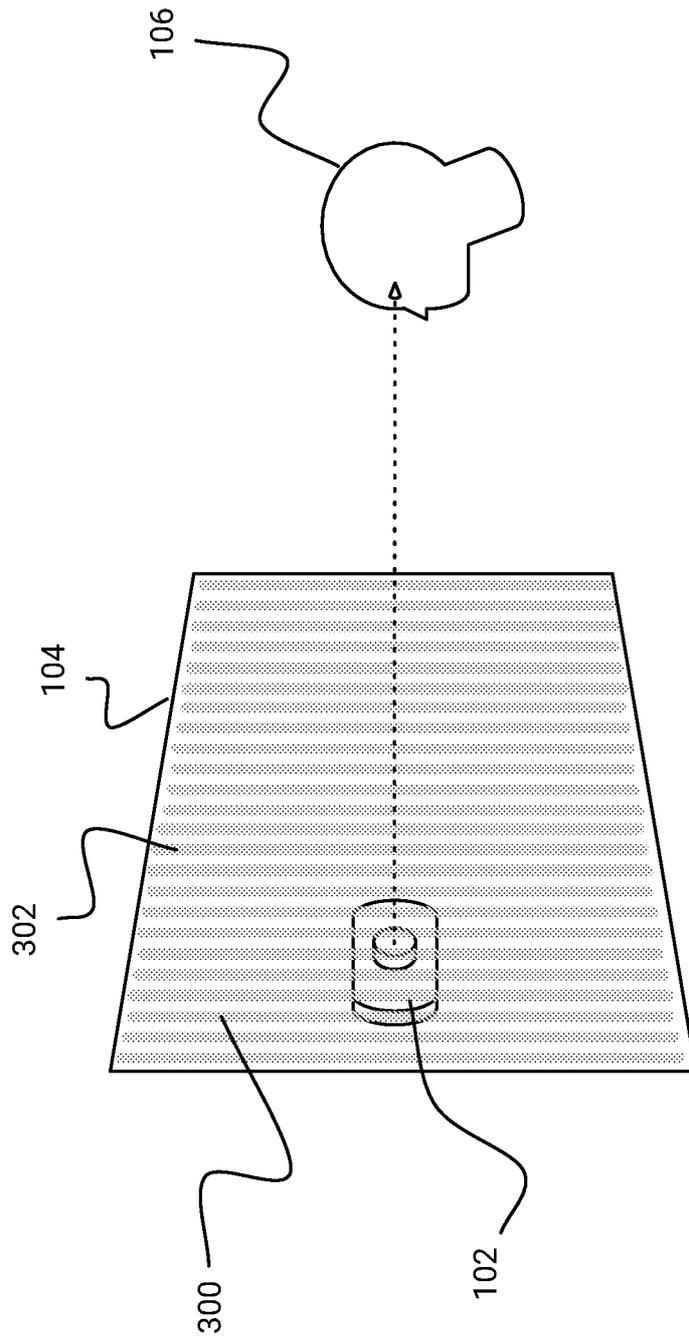
Figure 3

300
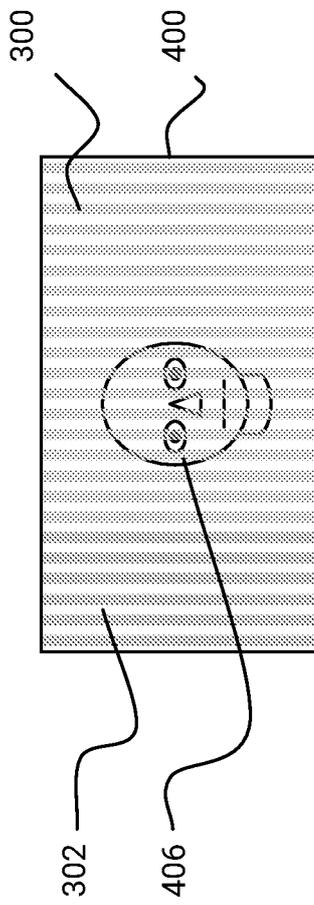
400

302

406

Figure 4a

402

408

302

300

Figure 4b

404

406
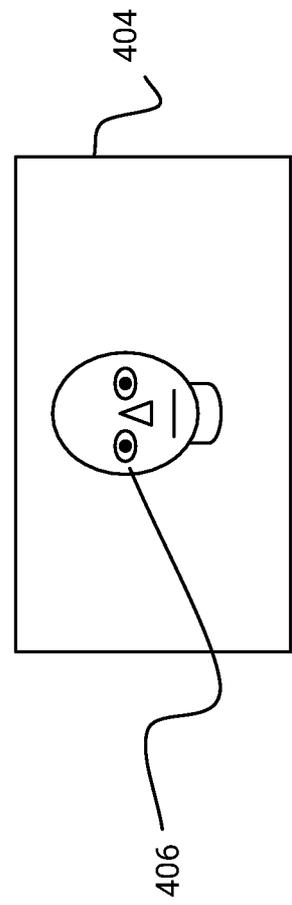
Figure 4c

Figure 5

Figure 6

Figure 7

Determine position of remote user's eyes in image — 800

Determine axis of camera — 802

Determine position of local user's eyes — 804

Modify the displayed image to align remote user's eyes and local user's eyes with axis — 806

Figure 8

Determining one or more pixel artifacts

900

↓

Compensating the displayed image

902

↓

Determining one or more occlusion artifacts

904

↓

Compensating the displayed image

906
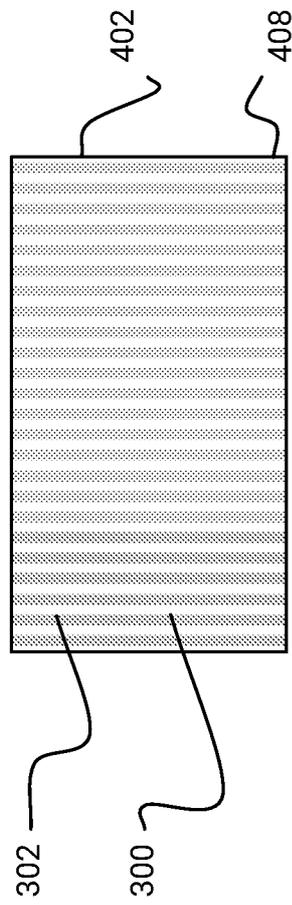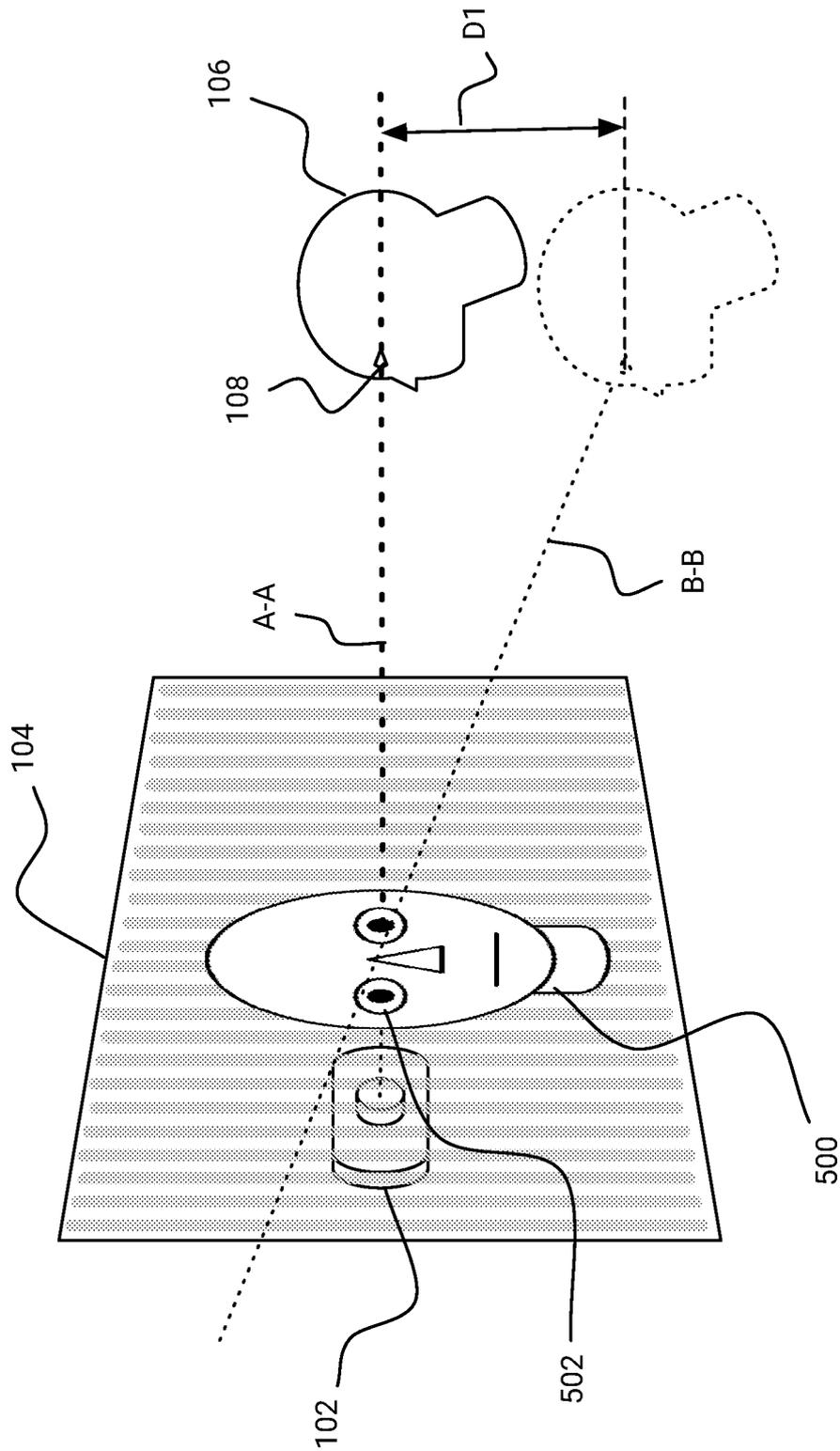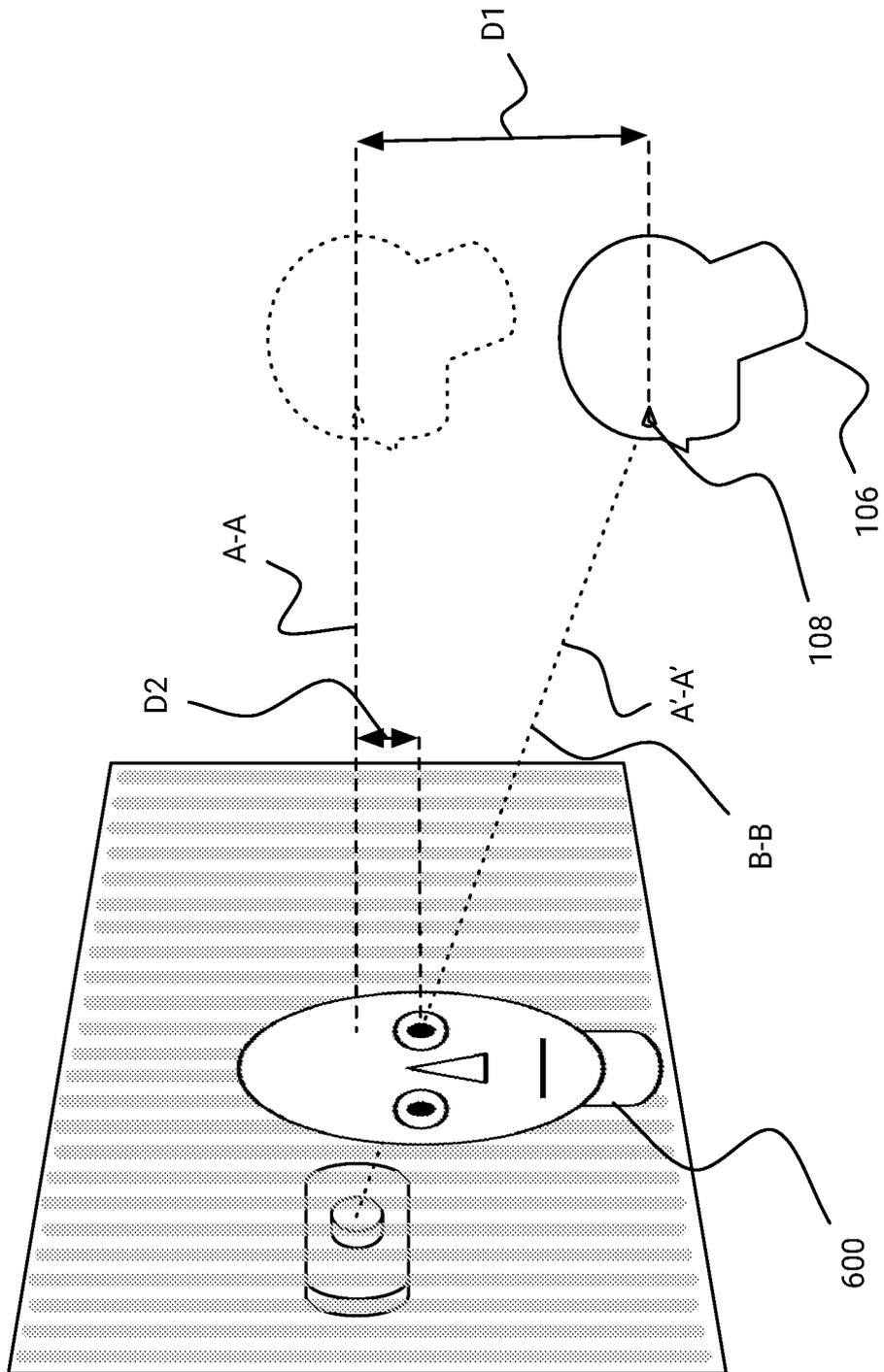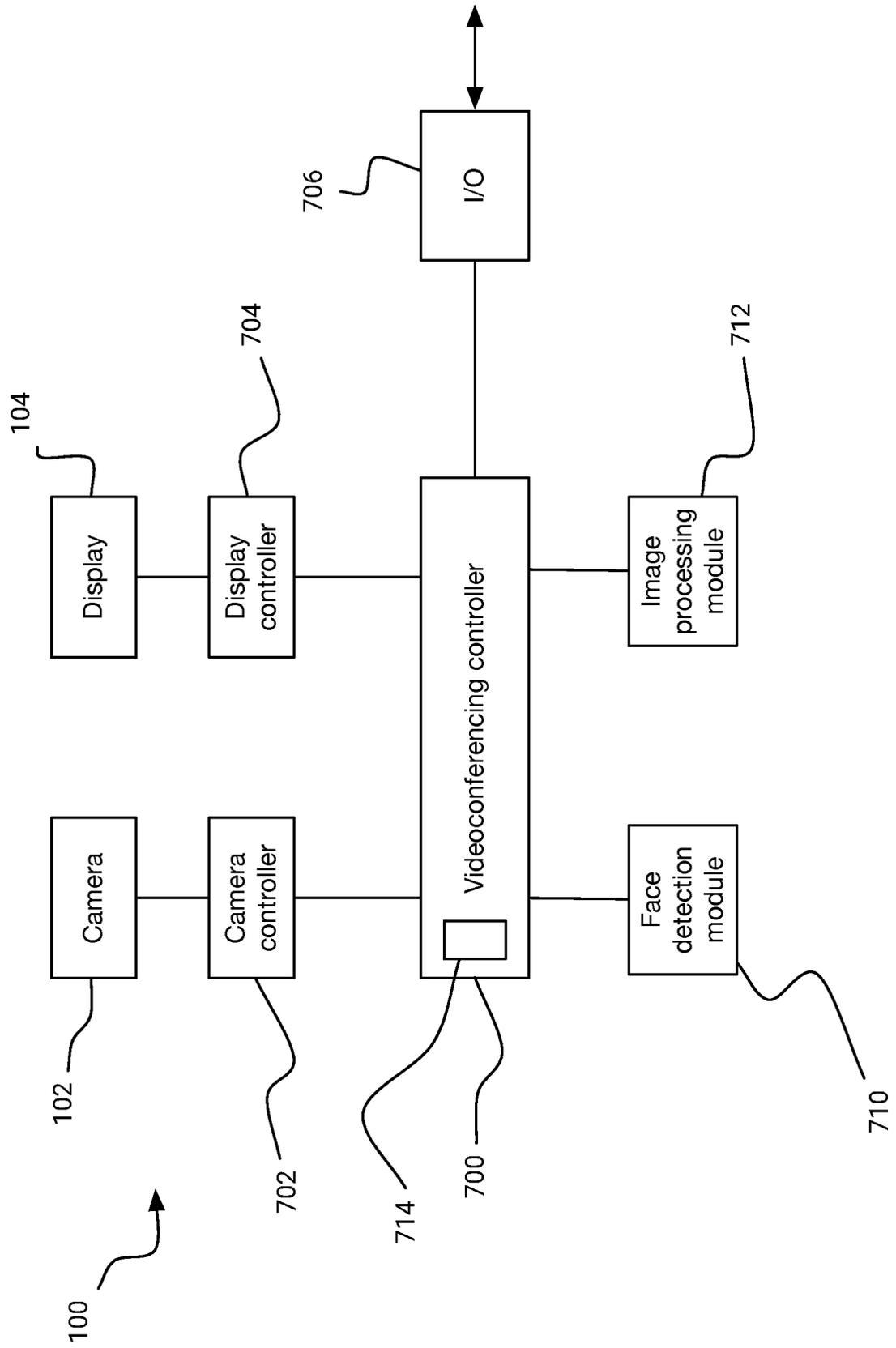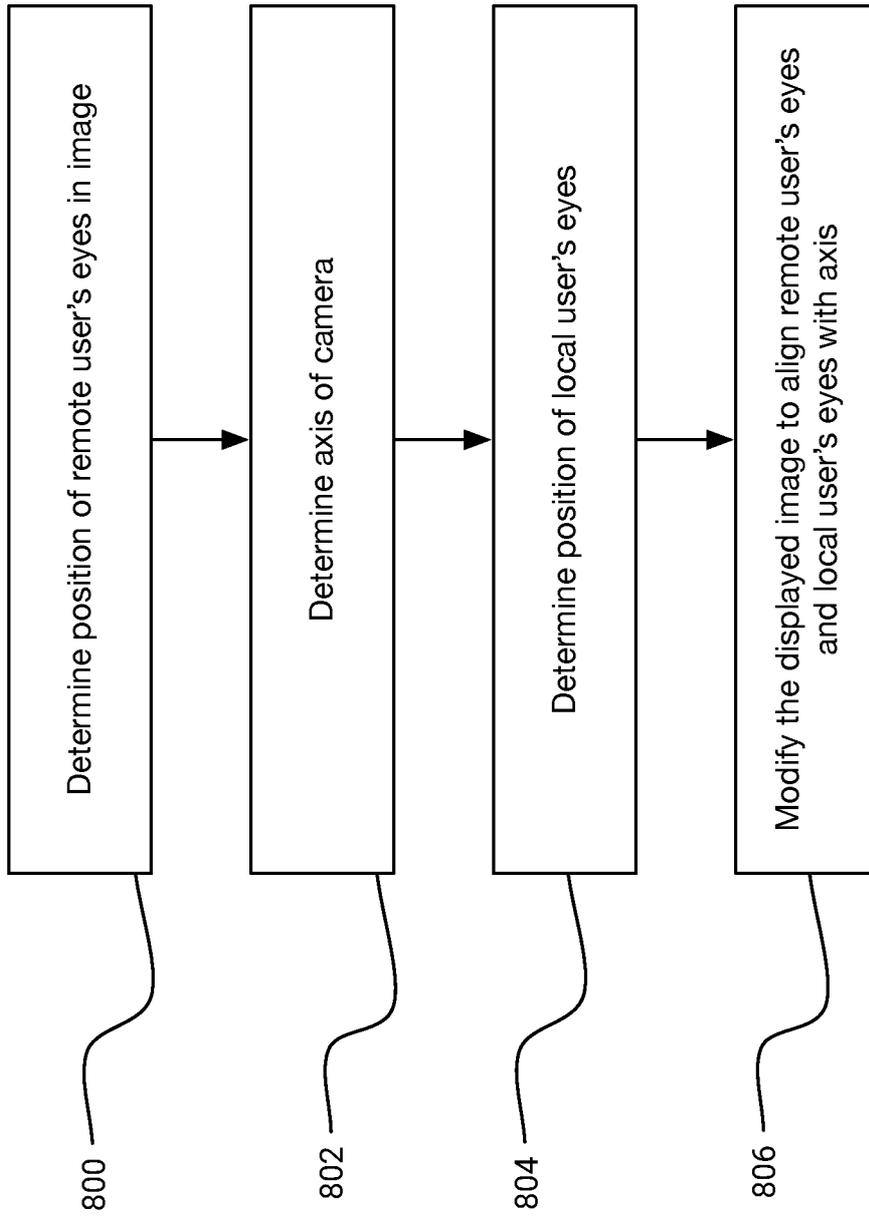
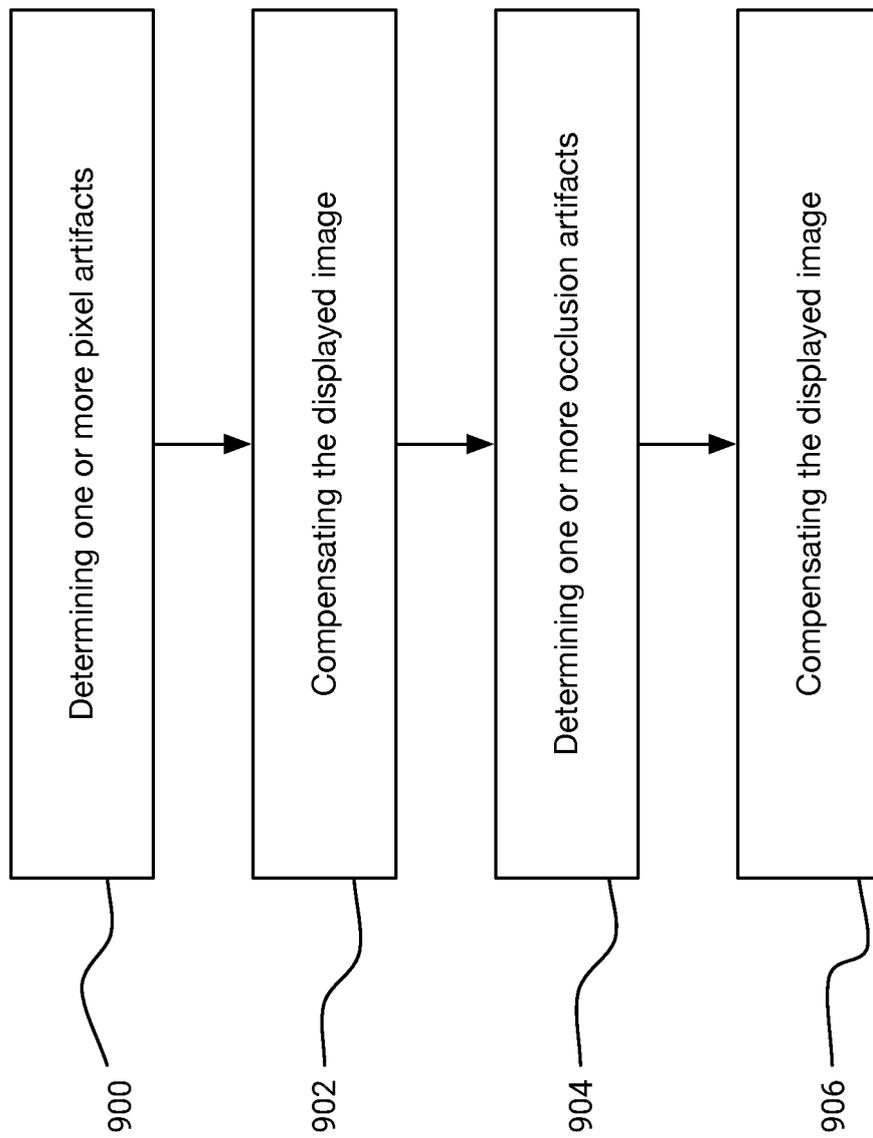Figure 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SE2020/050043

A. CLASSIFICATION OF SUBJECT MATTER

IPC: see extra sheet

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: H04N

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

SE, DK, FI, NO classes as above

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

EPO-Internal, PAJ, WPI data, COMPENDEX, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 20130093838 A1 (TAN KAR-HAN ET AL), 18 April 2013 (2013-04-18); abstract; paragraphs [0024], [0015], [0036]; figures 1-2,5-8 --	1-22
X	Kar-Han Tan, Robinson I N, Culbertson B, Apostolopoulos J, 'ConnectBoard: Enable Genuine Eye Contact and Accurate Gaze in Remote Collaboration', In: IEEE Transaction on Multimedia, 2011, June, vol: 13, no: 3, ISSN: 1520-9210; whole document --	1-22
X	US 20110102538 A1 (TAN KAR-HAN), 5 May 2011 (2011-05-05); whole document --	1, 16, 17, 22

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

*	Special categories of cited documents:	"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
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"D"	document cited by the applicant in the international application	"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
"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)	"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
"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	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24-02-2020	24-02-2020

Name and mailing address of the ISA/SE Patent- och registreringsverket Box 5055 S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Gordana Ninkovic Telephone No. +46 8 782 28 00

Form PCT/ISA/210 (second sheet) (July 2019)

C (Continuation).	DOCUMENTS CONSIDERED TO BE RELEVANT	
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2016192688 A (DAIWA HOUSE IND ET AL), 10 November 2016 (2016-11-10); whole document --	1, 2, 16, 17, 22
A	WO 2007047685 A2 (I2IC CORP ET AL), 26 April 2007 (2007-04-26); abstract; page 1, line 1 - line 6; page 2, line 26 - line 31; page 9, line 18 - page 10, line 28 -- --------	1-22

Continuation of: second sheet
International Patent Classification (IPC)
H04N 7/15 (2006.01)

US	20130093838	A 1	18/04/201 3	US		8908008	B2	09/12/2014
				W O		201 2008972	A 1	19/01/201 2
US	201 10 102538	A 1	05/05/201 1	US		8384760	B2	26/02/201 3
JP	20161 92688	A	10/1 1/201 6	J P		6461 679	B2	30/01/201 9
				W O		201 6 1591 66	A 1	06/10/201 6
W O	2007047685	A 2	26/04/2007	J P		5 1561 52	B2	06/03/201 3
				J P		2009524947	A	02/07/2009
				US		8022977	B2	20/09/201 1
				US		201 30088630	A 1	11/04/201 3
				US		201 10292255	A 1	01/12/201 1
				US		20070120879	A 1	31/05/2007
				US		83451 29	B2	01/01/201 3