



(51) International Patent Classification:

H04N 7/15 (2006.01) *H04N 7/18* (2006.01)
H04N 7/14 (2006.01)

(21) International Application Number:

PCT/US2009/038310

(22) International Filing Date:

26 March 2009 (26.03.2009)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

12/055,861 26 March 2008 (26.03.2008) US

(71) Applicant (for all designated States except US): **CISCO TECHNOLOGY, INC.** [US/US]; 170 West Tasman Drive, San Jose, CA 95134-1706 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MACDONALD, Alan, J.** [US/US]; 127 Paoli Pike, Malvern, PA 19355 (US). **MAUCHLY, J., William** [US/US]; 199 Cassatt Road, Berwyn, PA 19312 (US). **SOWA, David, W.** [US/US]; 119 Timber Spring Lane, Exton, PA 19341 (US). **FRIEL, Joseph, T.** [US/US]; 310 Ardmore Avenue, Ardmore, PA 19003 (US).

(74) Agent: **FRAME, Thomas, J.**; 6119 McCommas Blvd., Dallas, TX 75214 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: VIRTUAL ROUND-TABLE VIDEOCONFERENCE

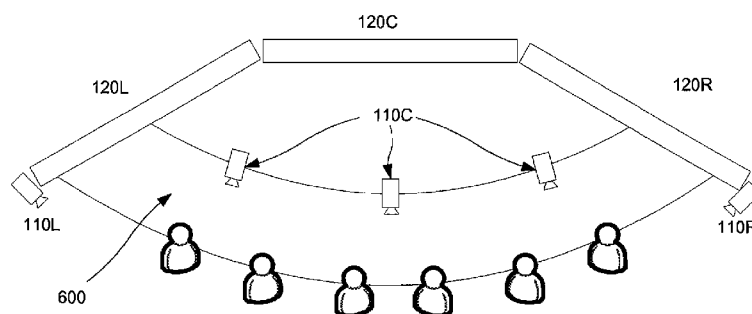


FIG. 1A

(57) Abstract: A system and method for creating a virtual round table videoconference is described. An embodiment of the system comprises a plurality of displays arranged in an configuration with a table to create a virtual round table. Cameras are arranged around the plurality of displays such that when a participant looks at a display with a image of a remote participant, the camera associated with the display captures an image of the participant's gaze, making eye contact with the camera. The image is displayed at the remote participant's endpoint creating the effect of eye contact between the participants. In another embodiment, audio speakers are arranged to provide directional sound that the video source for a display and the audio source for the associated speaker are from the same endpoint.



VIRTUAL ROUND-TABLE VIDEOCONFERENCE**FIELD OF THE INVENTION**

[0001] The present invention relates to teleconferencing. More particularly, the present invention relates to multipoint videoconferencing. Specifically, embodiments according to the present invention relate to a system and method for maintaining eye contact thus creating a more realistic environment during multipoint videoconferences.

BACKGROUND

[0002] Multipoint videoconferencing is a natural extension of point-to-point video conferencing. Multipoint videoconferencing usually includes a multipoint video bridge combining the video signals from multiple videoconference endpoints to provide a single output video signal which can be displayed to and shared by all the participants. When there are a large number of participants in the videoconference, multipoint systems have difficulty maintaining an accurate perspective of the entire videoconference. Ideally, a participant should be able to view all other participants at the other endpoints. However, because of limited display space and a potential for a large number of participants, it is not always possible to display the video images of all participants in an image size that is useful to the viewers.

[0003] To account for this problem, designers have relied on many different methods. One prior art method is to limit the number of participants displayed at any one endpoint such that each image is large enough to be beneficial to the participants

viewing them. As a participant speaks, her image is displayed at the other endpoints, replacing an existing image of a different participant. While this method has the advantage of displaying video of participants in an image size that is useful to other participants, it creates other problems. Because participants are not able to see all other participants at one time, a speaker must frequently address someone she cannot see. A speaker would often ask the person she is addressing to speak as a way of “tricking” the view switching algorithm, which may be based on audio activity, to switch the video image to the addressee.

[0004] Another prior art method used to deal with a large number of participants is to display the image of all participants. This “Hollywood Squares” approach, while giving participants the opportunity to see everyone in a videoconference, has its own problems. As the number of participants increases, the size of the individual images decreases making it more difficult for a participant to figure out who among the sea of faces is actually speaking.

[0005] While the current methods provide for some level of perspective in a multipoint videoconference, they do not create the perception that all participants are in the same room and leave speakers and audience members searching their displays for the right image.

[0006] Therefore, what is desired is a system and method that overcomes challenges found in the art, including a method for creating a more realistic videoconference environment allowing participants to all see each other at the same time and make eye contact with other participants without creating a “Hollywood Squares” effect.

SUMMARY OF THE INVENTION

[0007] In order to provide a realistic conference environment in a multipoint videoconference, it is desirable to have a system that can display video images of all participants at the same time. In one exemplary embodiment, a system is setup such that the meeting has the apparent geometry of a large round table, where each endpoint of the videoconference is a portion of the table.

[0008] In one exemplary embodiment of a method for practicing an aspect of the invention, a method for maintaining eye contact between participants of a multipoint conference is described. The method comprises creating a configuration of displays in an arch shape. Cameras are associated with each display so that when a participant at an endpoint looks at a particular display, the associated camera captures an image of the participant looking into and making eye contact with the associated camera. This image is displayed at other remote endpoint, giving the participants at the remote endpoint the visual effect of making eye contact with participants at other endpoints.

[0009] Additional advantages will be set forth in part in the description which follows or may be learned by practice. The advantages will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are examples and explanatory only and are not restrictive, as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The accompanying drawings, not drawn to scale, which are incorporated in and constitute a part of this specification, illustrate embodiments and together with the description, serve to explain the principles of the methods and systems:

FIG. 1A illustrates an embodiment according to the present invention;

FIG. 1B illustrates a three dimensional embodiment according to the present invention;

FIG. 2 illustrates an embodiment according to the present invention showing video scaling and compositing and the generated macro view;

FIG. 3 illustrates an embodiment according to the present invention showing the relationship of the displays and cameras between the plurality of endpoints;

FIG. 4 illustrates an embodiment according to the present invention showing video connections for one endpoint when there are more endpoints than the number of displays available at each endpoint;

FIG. 5 illustrates an embodiment according to the present invention showing a video image displayed at an endpoint when there are more endpoints in a videoconference than the number of displays at the endpoint; and

FIG. 6 illustrates an exemplary flow chart describing the steps to implement the method used by the videoconference system, according to one embodiment.

DETAILED DESCRIPTION

[0011] Before the present methods and systems are disclosed and described, it is to be understood that the methods and systems are not limited to specific methods, specific components, specific systems or to particular compositions, as such may,

of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

[0012] As used in the specification and the appended claims, the singular forms “a”, “an” and “the” include plural referents unless the context clearly dictates otherwise. Ranges may be expressed herein as from “about” one particular value, and/or to “about” another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent “about,” it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint.

[0013] “Optional” or “optionally” means that the subsequently described event or circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.

[0014] “Exemplary” means “an example of” and is not intended to convey a meaning of an ideal or preferred embodiment.

[0015] The present methods and systems may be understood more readily by reference to the following detailed description of embodiments and the examples included therein and to the figures and their previous and following description.

[0016] Embodiments according to the invention can be understood in the context of a multipoint videoconferencing system where a plurality of endpoints are displayed as to create a virtual round table.

[0017] In accordance with the embodiments according to the present invention, each endpoint in the multipoint videoconference has a plurality of displays configured with an arch shaped table such that the table and displays form a virtual round table. Cameras are located near of the plurality of display configuration so that when a participant turns to view the images on a given display, the participant makes eye contact with the associated camera. For example, when a participant turns to the left to address the audience displayed in the left screen, the left side camera captures the participant making eye contact. This image is relayed to the remote endpoint sourcing the left side display where it is displayed on the rights side screen of the remote endpoint. Through this configuration, the two endpoints appear to be side by side to the participants at both endpoints.

[0018] FIG. 1A illustrates a simplified non-limiting example of components according to one embodiment of the invention. A plurality of screens 120L, 120C, 120R are placed at predetermined angles to each other facing an arc shaped table 600. Wide angle cameras 110L, 110R are placed at the left and right sides of the plurality of screens to capture side views of the participants at the endpoint. Three center cameras 110C are used to capture close-up images of the participants and scaled and combined to create a composite image of all participants at an endpoint.

[0019] Note, while the illustration depicts the wide angle cameras 110R, 110L to be located to the left and right ends of the display configuration, this is not required to

practice the invention. The wide angle cameras 110R, 110L may be placed in a plurality of locations as long as they are aligned such that when participants look to the left or right screens 120L, 120R, the wide angle cameras 110L, 110R capture the image of the participant making eye contact with the cameras 110L, 110R. Additionally, while the illustrations and description refer to a system with three displays, this is not required to practice the invention. The plurality of displays can consist of any number of displays depending on the needs of the user.

[0020] FIG. 1B illustrates the exemplary system of FIG. 1A, including the possible images displayed on the display screens. In accordance with one embodiment according to the present invention, the image displayed on the left screen 120L is associated with the endpoint on the participants 131, 132, 133 virtual left. From the perspective of the participants in the image displayed in the left screen 120L, the participants 131, 132, 133 at the current endpoint appear to the virtual right. The image displayed on the center screen 120C is a scaled composite image of the participants from a third endpoint. The image displayed on the right screen 120R is associated with the endpoint on the participants 131, 132, 133 virtual right. From the perspective of the participants in the image displayed in the right screen 120R, the participants 131, 132, 133 at the current endpoint appear to the virtual left.

[0021] FIG. 2 illustrates a simplified non-limiting example of a center composite image, comprised of a plurality of close-up images. The center close up cameras 110C-1, 110C-2, 110C-3 are each configured to capture an image 110C-1i, 110C-2i, 110C-3i of one or more participants. Through the usage of scaling and

compositing methods known to practitioners of the art, in step 620 the three images 110C-1i, 110C-2i, 110C-3i are incorporated into a single image 640.

[0022] FIG. 3 illustrates a simplified non-limiting example of an embodiment according to the present invention with four endpoints. The center screens 120C, 220C, 320C, 420C display composite images from the opposite endpoints close up cameras 310C, 410C, 110C, 210C. Left side screens 120L, 220L, 230L, 420L use the wide angle right side cameras 210R, 310R, 410R, 110R of the endpoints 200, 300, 400, 100 on their virtual left. Right side screens 120R, 220R, 320R, 420R use the wide angle left side cameras 410L, 110L, 210L, 310L of the endpoints 400, 100, 200, 300 on their virtual right.

[0023] Through creating a virtual round table where each endpoint comprises a portion of the table, participants at remote locations have a more natural environment in which to conduct a videoconference. In accordance with the embodiments according to the present invention, participants at one endpoint 100 looks straight ahead at the center screen 120C to talk to one endpoint 300, looks to the right side screen 120R to talk to a second endpoint 400, and looks to the left side screen 120L to talk to the third endpoint 200. Cameras associated with the various screens, 120L, 120R, 120C capture the image of the participants looking directly into the camera, creating the appearance of eye contact with participants at other endpoints.

[0024] FIG. 4 illustrates a simplified non-limiting embodiment according to the invention of a multipoint videoconference where there are more endpoints than display screens. Note, this illustration shows one of a plurality of possible

examples and can be applied to situations where there are more than five endpoints. In accordance with the embodiments according to the present invention, the number of displays at each endpoint may be equal to the total number of endpoints in a plurality of endpoints. For instance, in a five endpoint videoconference, an additional center display screen may be added to the system described in FIG. 3 with the fifth endpoint being virtually inserted into the virtual roundtable between the plurality of existing endpoints 100, 200, 300, 400 in FIG. 3. However, there is a possibility that there may be more endpoints than display screens in a videoconference. In an embodiment of the invention, when there is one endpoint more than display screens available, the images of endpoints 300, 400 not on the virtual left or right of a given endpoint 100 is displayed on the center screen 120C. FIG. 5 illustrates this configuration with the top image 320 being from one endpoint 300 and the lower image 420 being from the second endpoint 400.

[0025] In another embodiment of the invention, when there are more endpoints than the number of display screens 120L, 120R, 120C available at each endpoint, a general solution is to employ endpoint switching. In this embodiment of the invention, the endpoints in the videoconference are all connected in a virtual ring. The left screen 120L for the endpoint 100 is connected to the right side camera of the remote endpoint 200 to the virtual left. The right screen 120R for the endpoint 100 is connected to the left side camera of the remote endpoint 500 to the virtual right. The left and right screen's image sources remain constant during the videoconference. The center screen 120C switches between the endpoints not

being displayed on the left or right screen 120L, 120R, using the composite image generated from the close-in center cameras of the endpoints 300, 400. A switch 520 controls which video image is currently being displayed on the center screen 120C. A plurality of algorithms known in the prior art may be used to control the switching of the two center images. For example, the center screen image may be based on the endpoint 300, 400 with a participant that has most recently spoken, or the image may be based on manual selection at the plurality of endpoints 100, 200, 300, 400, 500. For practical reasons, conferences rarely include large numbers of endpoints. The solution provided above preserves the effective eye-contact for the two fixed left and right side screens 120L, 120R.

[0026] Note, the non-limiting examples described above uses a three display system.

However, the invention is not limited to three displays. For example, in an embodiment, there could include more than one center display. Additionally, in another embodiment, there could be one display where the left, right and center view images are scaled and composited into one image.

[0027] FIG. 6 illustrates an exemplary flow chart describing the steps needed to implement the method used by the videoconference system, according to one embodiment. In step 610, to create left, right and center views, a plurality of displays are setup in an arch configuration such that participants at an endpoint turns to the left, right or center when looking at displays showing the images of participants at other endpoints. To capture the image of participants as they look at left, right or center displays, in step 620, cameras are setup such that when a participant looks at the left, right or center display, one of the plurality of cameras

captures the participant's gaze, making eye contact with the camera. In step 630, a center view for center displays can be created by one camera or by scaling and compositing close-up images from a plurality of cameras. In step 640, when a participant looks left, right or center to view other participants at a remote endpoint, the camera associated with the left, right or center view captures the image of the participant looking directly at the camera. This image is operably connected to the same endpoint being viewed on the left, right or center display. Thus, when a participant looks to the left display, a left side camera captures the image of the participant looking into the camera. This image is then displayed at the endpoint sourcing the left display creating the effect of the participants making eye contact with each other.

[0028] In an embodiment of the invention, to further enhance the virtual effects of the invention, audio speakers may be associated with the plurality of display screens. Referring to FIG. 3, when a participant 230 at an endpoint 200 speaks, the participant's 230 voice is relayed to an audio speaker associated with the display screen 120L currently showing the participant's 230 image. In associating the audio with the video image, a participant's 130 attention, at a remote endpoint 100, will be naturally drawn to the video image associated with the source of the audio.

[0029] In accordance with embodiments of the invention, the wide angle and composite images of participants may be used in addition to a plurality of other viewing options. For example, when a participant at one endpoint speaks for a sufficiently long duration, the display showing the wide angle or composite image of the participant's endpoint may switch to a close-up image of the participant

from one of the center cameras. The close up images may also be used as part of manual selection by a participant.

[0030] While the methods and systems have been described in connection with preferred embodiments and specific examples, it is not intended that the scope be limited to the particular embodiments set forth, as the embodiments herein are intended in all respects to be illustrative rather than restrictive.

[0031] Unless otherwise expressly stated, it is in no way intended that any method set forth herein be construed as requiring that its steps be performed in a specific order. Accordingly, where a method claim does not actually recite an order to be followed by its steps or it is not otherwise specifically stated in the claims or descriptions that the steps are to be limited to a specific order, it is no way intended that an order be inferred, in any respect. This holds for any possible non-express basis for interpretation, including: matters of logic with respect to arrangement of steps or operational flow; plain meaning derived from grammatical organization or punctuation; the number or type of embodiments described in the specification.

[0032] It will be apparent to those skilled in the art that various modifications and variations can be made without departing from the scope or spirit. Other embodiments will be apparent to those skilled in the art from consideration of the specification and practice disclosed herein. It is intended that the specification and examples be considered as examples only, with a true scope and spirit being indicated by the following claims.

CLAIMS

What is claimed is:

1. A method for creating a virtual illusion of meeting around a continuous table for a multipoint videoconference comprising:

showing, on a display system, a center perspective image of participants in a plurality of remote endpoints;

showing, on a display system, a left side perspective image of a plurality of participants at a first remote endpoint; and

showing, on the display system, a right side perspective image of a plurality of participants at a second remote endpoint.

2. The method of claim 1, wherein showing on the display system, the center perspective image of participants at the plurality of remote endpoints comprises using a scaled and composited image of participants at the plurality of remote endpoints.

3. The method of claim 2, wherein using the scaled and composited image of participants at the plurality of remote endpoints comprises scaling the image such that participants shown in the center perspective image appear visually farther away than those at the right or left side of the display system.

4. The method of claim 1, wherein showing, on the display system, the left side perspective image of the plurality of participants at the first remote endpoint comprises

displaying on a right side of the display system the left side perspective image of the participants at the first remote endpoint.

5. The method of claim 1, wherein showing, on the display system, the right side perspective image of the plurality of participants at the second remote endpoint comprises displaying on a left side of the display system the right side perspective image of the participants at the second remote endpoint.

6. The method of claim 4, wherein displaying on the right side of the display system the left side perspective image of the participants at the first remote endpoint comprises displaying the left side perspective image such that participants in the left side perspective image nearer to the center portion of the display system appear visually farther away than those at the right side of the display system.

7. The method of claim 5, wherein displaying on the left side of the display system the right side perspective image of the participants at the second remote endpoint comprises displaying the right side perspective image such that participants in the right side perspective image nearer to the center portion of the display system appear visually farther away than those at the left side of the display system

8. The method of claim 1, wherein the display system comprises a plurality of displays configured in an arc pattern such that the images on the display system in combination with a table creates the illusion of the continuous table.

9. The method of claim 1, wherein the display system comprises one display configured to display the center, left and right perspectives such that the images on the display system in combination with the table creates the illusion of the continuous table.
10. A method for providing directional sound for audio signals associated with each remote endpoint to draw the attention of participants to a new speaker when the participants are not looking at an image associated with the new speaker.
11. The method of claim 10, wherein providing directional sound for audio signals associated with each remote endpoint to draw the attention of participants to a new speaker when the participants are not looking at the image associated with the new speaker comprises:
- locating a plurality of audio speakers near a display system; and
 - associating the source for the audio speaker with the image being displayed on the display system.
12. A system for creating a virtual illusion of meeting around a continuous table for a multipoint videoconference comprising:
- a network; and
 - a plurality of endpoints operably connected to the network wherein an endpoint is configured to:

display, on a display system, a center perspective image of participants in a plurality of remote endpoints;

display, on the display system, a left side perspective image of a plurality of participants at a first endpoint; and

display, on the display system, a right side perspective image of a plurality of participants at a second endpoint.

13. The system of claim 12, wherein the endpoint is configured to display the center perspective image of participants in the plurality of endpoints comprises the endpoint being further configured to display the center perspective image generally in the center of a display system.

14. The system claim of 12, wherein the endpoint is configured to display the center perspective image of participants in the plurality of endpoints comprises the plurality of endpoints being configured to use scaled and composited images of participants at the plurality of remote endpoints such that the participants shown in the center perspective image appear visually farther way than those at the right or left side of the display system.

15. The system of claim 12, wherein the endpoint is configured to display the left side perspective image of the plurality of participants at the first endpoint comprises the endpoint being further configured to display on a right side of the display system, the left side perspective image of the participants at the first endpoint.

16. The system of claim 12, wherein the endpoint is configured to display the right side perspective image of the plurality of participants at the second endpoint comprises the endpoint being further configured to display on a left side of the display system, the right side perspective image of the participants at the second endpoint.

17. The system of claim of 15, wherein the endpoint is configured to display on the right side of the display system, the left side perspective image of the participants at the first endpoint comprises the endpoint being further configured to display the left side perspective image such that participants in the left side perspective image nearer to a center portion of the display system appear visually farther away than those at the right side of the display system.

18. The system of claim 16, wherein the endpoint is configured to display on the left side of the display system, the right side perspective image of the participants at the second endpoint comprises the endpoint being further configured to display the right side perspective image such that participants in the right side perspective image nearer to the center portion of the display system appear visually farther away than those at the left side of the display system.

19. The system of claim 12, wherein the display system comprises a plurality of displays configured in an arc pattern such that the images on the display system in combination with a table creates the illusion of the continuous table.

20. The system of claim 12, wherein the display system comprises one display configured to display the center, left and right perspectives such that the images on the display system in combination with the table creates the illusion of the continuous table.

1/5

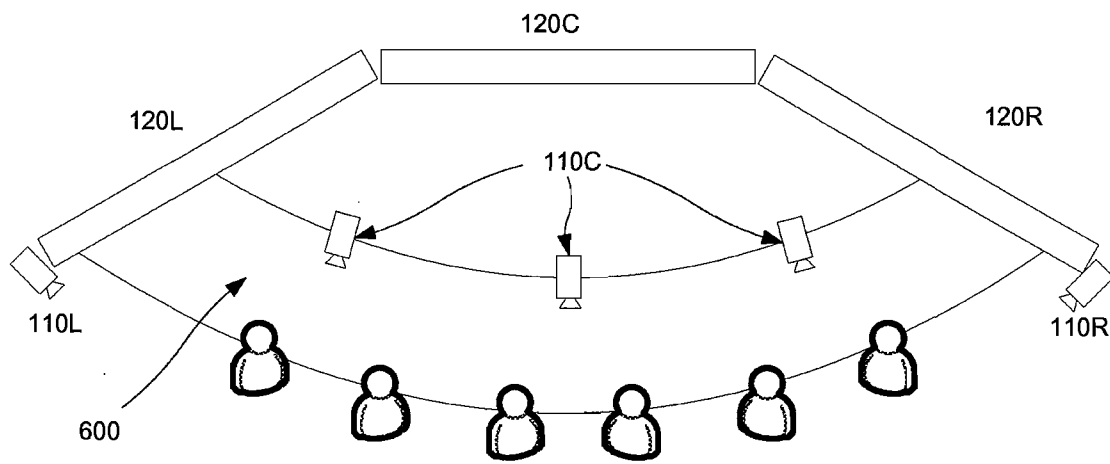


FIG. 1A

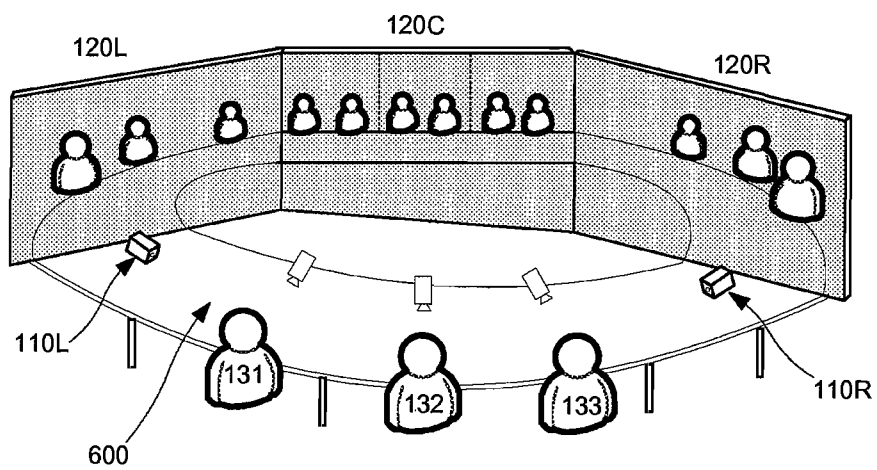
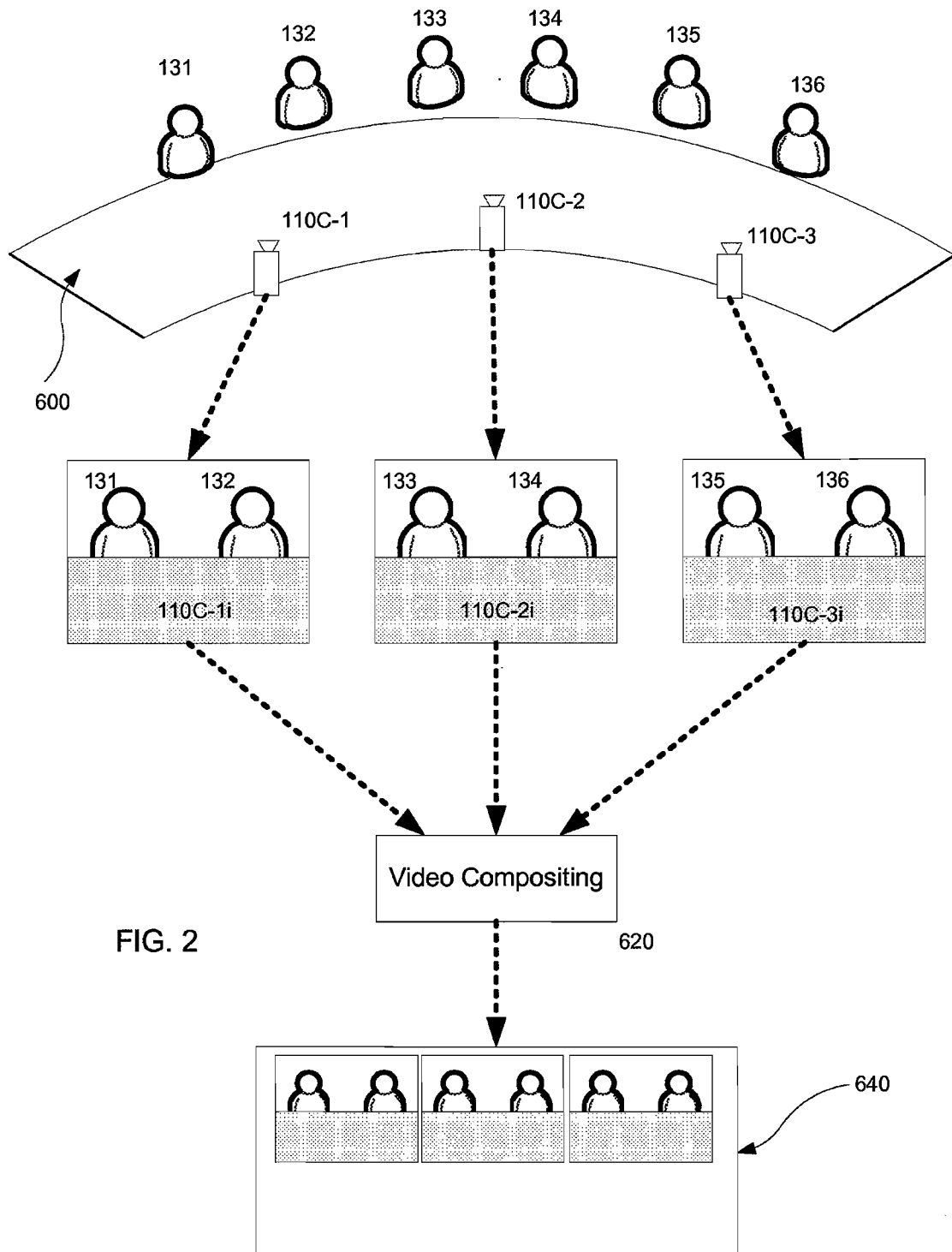


FIG. 1B

2/5



3/5

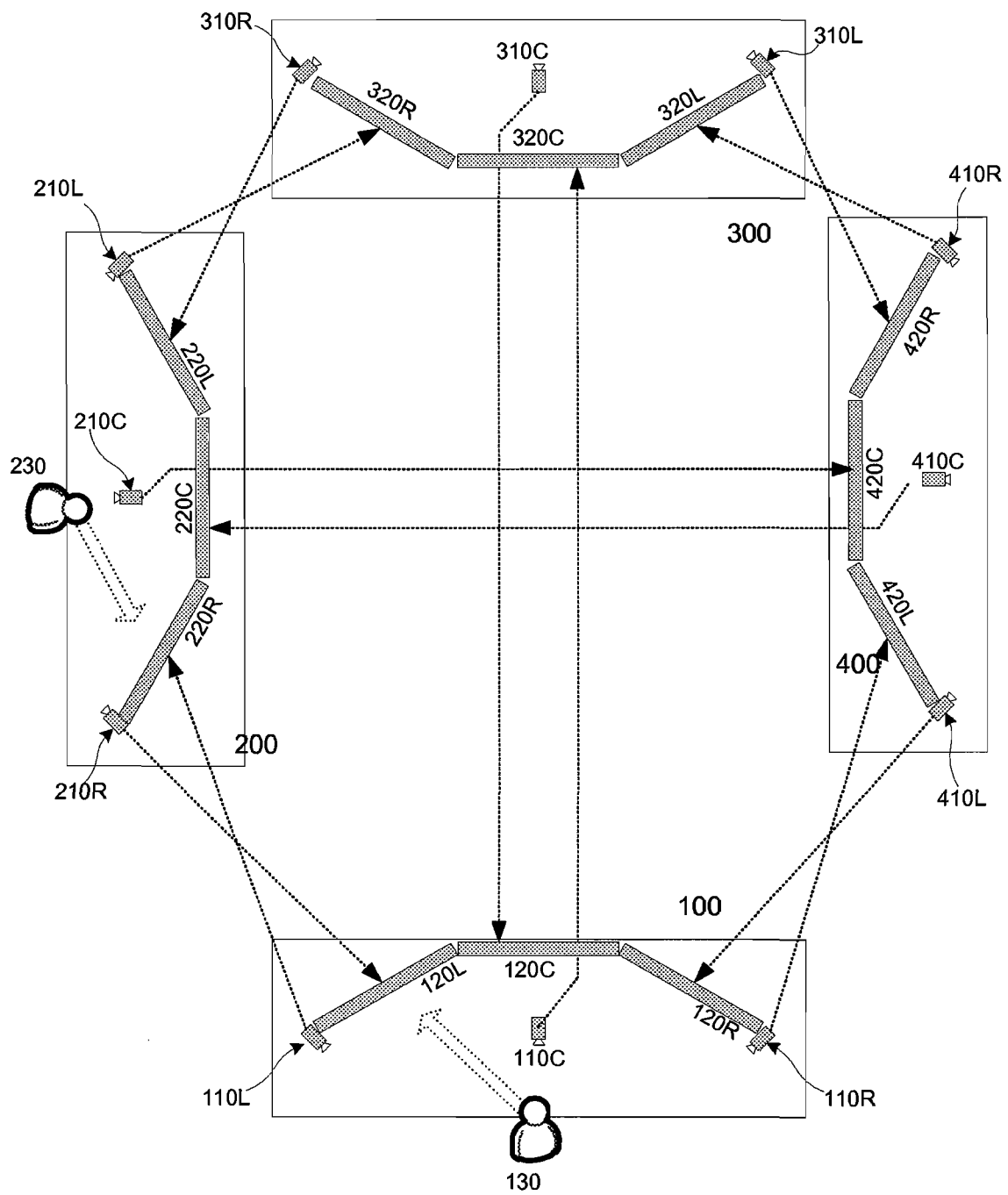
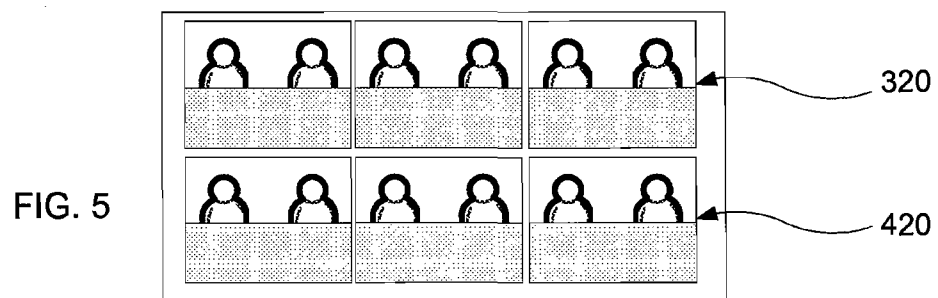
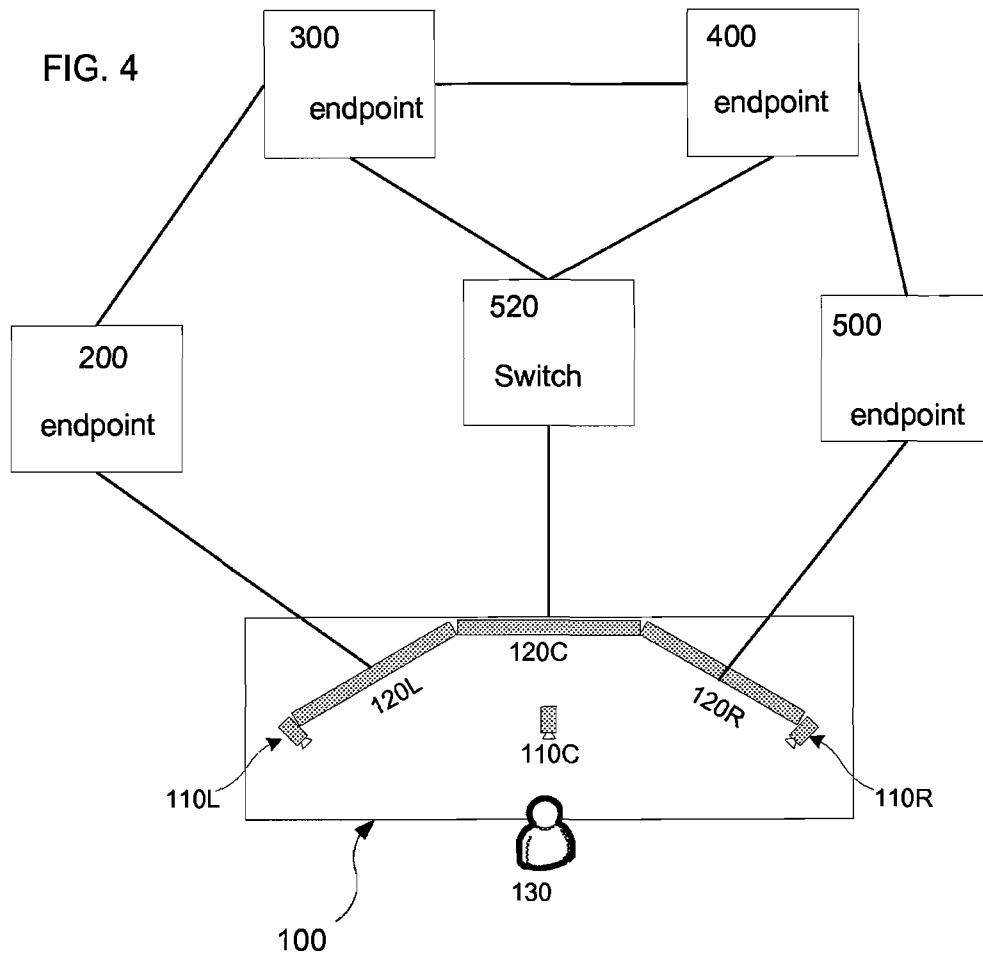


FIG. 3

4/5



5/5

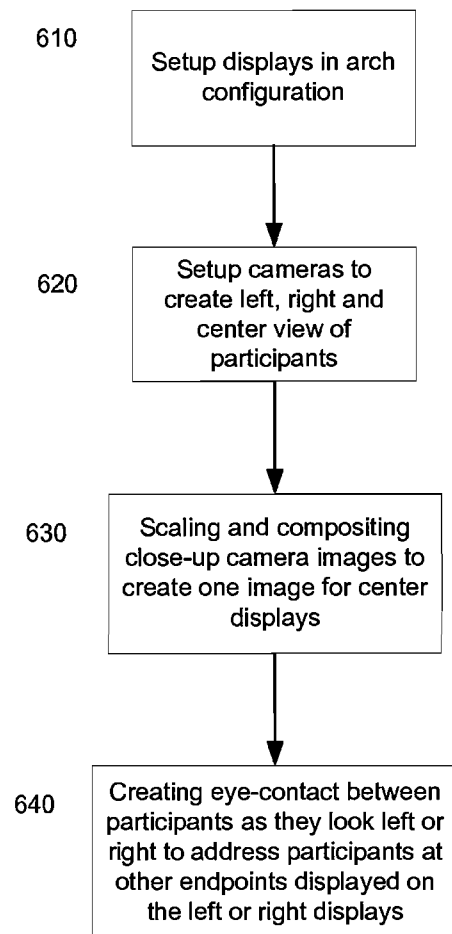


FIG. 6