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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 January 2009 (22.01.2009)

PCT

(10) International Publication Number
WO 2009/011592 A1

(51) International Patent Classification:
H04N 7/15 (2006.01)

(21) International Application Number:
PCT/NO2008/000249

(22) International Filing Date: 30 June 2008 (30.06.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
20073621 13 July 2007 (13.07.2007) NO
60/949,718 13 July 2007 (13.07.2007) US

(71) Applicant *(for all designated States except US)*: **TANDBERG TELECOM AS** [NO/NO]; Philip Pedersens vei 22, N-1366 Lysaker (NO).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ENSTAD, Gisle** [NO/NO]; Haugjordet 17, N-1336 Sandvika (NO). **KORNELIUSSEN, Jan, Tore** [NO/NO]; Schweigaards gt. 92, N-0656 Oslo (NO). **HUSØY, Per, Ove** [NO/NO]; Linjeveien 51D, N-1087 Oslo (NO).

(74) Agents: **ONSAGERS AS** et al.; P.O. Box 6963 St. Olavs plass, N-0130 Oslo (NO).

(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, 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, 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).

Published:
— *with international search report*

(54) **Title:** METHOD AND SYSTEM FOR AUTOMATIC CAMERA CONTROL

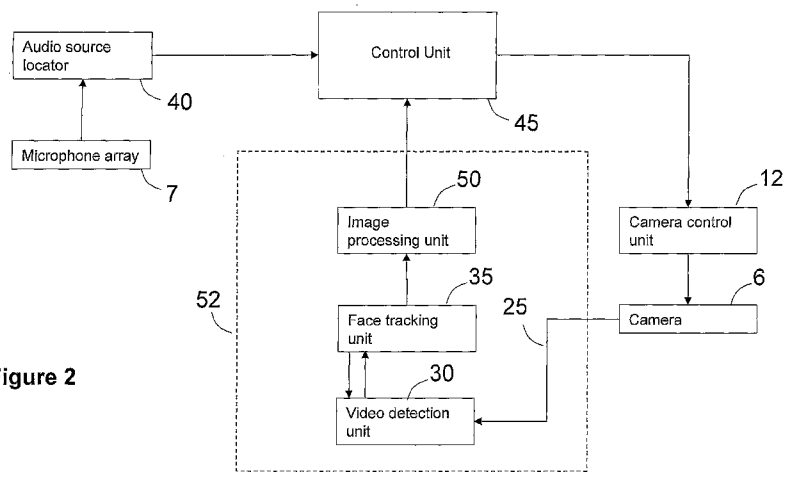

Figure 2

(57) **Abstract:** The present invention uses video detection techniques to detect participants and their respective locations in video frames captured by the camera, and based on the location and sizes of said detected participants, automatically determine and use the optimal camera orientation and zoom for capturing the best view of all the participants.

WO 2009/011592 A1

METHOD AND SYSTEM FOR AUTOMATIC CAMERA CONTROL

Field of the invention

The present invention relates to video conferencing and automatic adjustment of camera orientation and zoom.

Background of the invention

In most high end video conferencing systems, high quality cameras with pan-, tilt-, and zoom capabilities are used to frame a view of the meeting room and the participants in the conference. The cameras typically have a wide field-of-view (FOV), and high mechanical zooming capability. This allows for both good overview of a meeting room, and the possibility of capturing close-up images of participants. The video stream from the camera is compressed and sent to one or more receiving sites in the video conference. All sites in the conference receive live video and audio from the other sites in the conference, thus enabling real time communication with both visual and acoustic information.

Video conferences vary a great deal when it comes to purpose, the number of participants, layout of conference rooms, etc. Each meeting configuration typically requires an individual adjustment of the camera in order to present an optimal view. Adjustments to the camera may be required both before and during the video conference. E.g. in a video conference room seating up to 16 persons, it is natural that the video camera is preset to frame all of the 16 available seat locations. However, if only 2 or 3 participants are present, the wide field of view camera setting will give the receiving end a very poor visual representation.

Adjustments to the camera are typically done via a remote control, either by manually controlling the camera pan, tilt and zoom, or by choosing between a set of pre-defined

camera positions. These pre-defined positions are manually programmed. Often, before or during a video conference, the users do not want to be preoccupied with the manual control of the camera, or the less experienced user may not even be aware of the possibility (or how) to change the cameras field of view. Hence, the camera is often left sub-optimally adjusted in a video conference, resulting in a degraded video experience.

Therefore, in order to ensure a good camera orientation for every situation in a video conferencing room, an automatic field of view adjustment system is desirable.

Some video conferencing systems with Camera Tracking capabilities exist. However, the purpose of these systems is to automatically focus the camera on an active speaker. These systems are typically based on speaker localization by audio signal processing with a microphone array, and/or in combination with image processing.

Some digital video cameras (for instance web-cams) use video analysis to detect, center on and follow the face of one person within a limited range of digital pan, tilt and zoom. However, these systems are only suitable for one person, require that the camera is initially correctly positioned and have a very limited digital working range.

Hence, none of the prior art mentioned above is describing a system for automated configuration of the camera in a video-conference setting.

Summary of the invention

It is an object of the present invention to provide a method and a system solving at least one of the above-mentioned problems in prior art.

The features defined in the independent claims enclosed characterise this method and system.

Brief description of the drawings

In order to make the invention more readily understandable, the discussion that follows will refer to the accompanying drawings.

Figure 1 illustrating a typical video conferencing room,

Figure 2 schematically shows components of "best view" locator according to the present invention,

Figure 3 is a flow chart of the operation of the "best view" locator,

Figure 4 schematically shows a typical video conferencing situation, and exemplary initial orientations of the image pickup device,

Figure 5 illustrates face detection in a image containing two participants,

Figure 6 illustrates one exemplary defined area of interest ("best view"),

Figure 7 illustrates another exemplary defined area of interest ("best view"),

Figure 8 illustrates a camera framing of said defined area in figure 6,

Figure 9 illustrates an audio source detected outside the currently framed image,

Figure 10 illustrates a camera framing including a participant representing said audio source in figure 9,

Figure 11 illustrates a participant leaving the cameras field of view, where

Fig. 11a illustrates that a person leaves the conference;

Fig. 11b illustrates that a person is near the edge of the frame;

Fig. 11c illustrates the remaining two persons; and

Fig. 11d illustrates the optimal view for the remaining persons.

Detailed description the invention

In the following, the present invention will be discussed by describing a preferred embodiment, and by referring to the accompanying drawings. However, people skilled in the art will realize other applications and modifications within the scope of the invention as defined in the enclosed independent claims.

Figure 1 illustrates a typical video conferencing room **10**, with an exemplary video conferencing system **20**. Video conferencing systems **20** usually consist of the following components; a codec **11** (for coding and decoding audio and video information), a user input device **8** (i.e. remote control or keyboard), a image capture device **6** (camera), an audio capture device **4;7** (microphone), a video display **9** (screen) and an audio reproduction device **5** (loudspeakers). Often, high end video conferencing systems (VCS) use high quality cameras **6** with motorized pan-, tilt-, and zoom capabilities.

The present invention uses video detection techniques to detect participants and their respective locations in video frames captured by the camera **6**, and based on the location and sizes of said detected participants automatically

determine and use the optimal camera orientation and zoom for capturing the best view of all the participants.

There may be many opinions on what the "best view" of a set of participants in a video conference is. However, in the following, a "best view" is referred to as a close-up of a group of participants, where the video frame center substantially coincide with the center of the group, and where the degree of zoom gives a tightly fitted image around said group. However, the image must not be too tight, showing at least the participant's upper body, and give room for the participants to move slightly without exiting the video frame.

Figure 2 schematically shows the modules in the "best view" locator **52** according to the present invention. A video detection unit **30** is configured to continuously detect objects, e.g. faces and/or heads, in the frames of a captured video signal. At predefined events (e.g. when the VCS is switched on, when initiated through the user input device **8**, etc.) the camera zooms out to its maximum field of view and moves to a predefined pan- and tilt-orientation (azimuth- and elevation-angle), capturing as much as possible of the room **10** where the system is situated. The video detection unit **30** analyses the frames in the video signal and detects all the faces/heads and their location in the video frame relative to a predetermined and static reference point (e.g. the center of the frame). The face/head location and size (or area) in the video image is transformed into camera coordinates (azimuth- and elevation angles and zoom factors) Information about each detected face/head (e.g. position, size, etc.) is sent to an image processing unit **50** via face tracking unit **35**. Based on said face/head information the image processing unit defines a rectangular area that at least comprises all the detected faces/heads. A predefined set of rules dictate how the area should be defined, and the area represents the best view of the people in the frame (or video conferencing room **10**).

The camera coordinates (azimuth- and elevation angles and zoom factors) for the defined area and its location is sent to a control unit 45. The control unit instructs a camera control unit **12** to move the camera to said camera coordinates, and the camera's **6** pan, tilt and zoom is adjusted to frame an image corresponding to the defined area.

The image pickup device (or camera) **6** includes a camera control unit **12** for positioning the image pickup device. The camera control unit **12** is the steering mechanism, including motors, controlling the pan- and tilt-orientation and the degree of zoom of the image pickup device **6**. The camera control unit **12** can also report back its current azimuth- and elevation angles and zoom factors on demand. The image processing unit **50** and the control unit **45** may supply control signals to the camera control unit **12**. Camera control unit **12** uses a camera coordinate system which indicates a location based on azimuth- and elevation angles and zoom factors which describe the direction of the captured frame relative to camera **6** and the degree of zoom. Video detection unit **30** is configured to convert coordinate measurements expressed in a video (or image) coordinate system to coordinate measurements expressed in the camera coordinate system using the azimuth- and elevation angles and zoom factors of camera **6** when the frame was captured by camera **6**.

Figure 3 is a flow chart of the operation of the "best view" locator **52**. Camera **6** outputs a video signal comprising a series of frames (images). The frames are analyzed by the video detection unit **30**. At predefined events, the camera control unit **12** is instructed to move the camera to an initial orientation (step 60). The object of the initial orientation is to make sure that the camera can "see" all the persons in the meeting room. There are several ways of deciding such an initial orientation.

Referring to figure 4, according to one exemplary embodiment of the invention, the camera zooms out to its maximum field of view and moves to a predefined pan- and tilt-orientation **13**, capturing as much as possible of the room **10a** and/or capturing the part of the room with the highest probability of finding meeting participants. The predefined pan- and tilt-orientation (or initial orientation) is typically manually entered into the system through a set-up function (e.g. move camera manually to an optimal initial position and then save position) or it is a default factory value.

According to another exemplary embodiment of the invention, the camera is configured to capture the entire room by examining a set of initial orientations **(14, 15)** with a maximum field of view, and where the set's fields of view overlaps. In most cases, a set of 2 orientations will suffice. However, the number of orientations will depend on the cameras maximum field of view, and may be 3, 4, 5, 6, etc. For each orientation **(14, 15)** the one or more video frames are analyzed by video detection unit **30** to detect faces and/or heads and their respective locations. After analyzing all the orientations, the image processing unit **50** calculates the pan- and tilt orientation that includes all the detected participants, and defines said calculated orientation as the initial orientation.

A video detection unit **30** analyzes the video signals **25** from the camera **6** to detect and localize faces and/or heads (step 70) in a video frame. The video detection unit **30** measures the offset of the location of the detected faces/heads from some pre-determined and static reference point (for example, the center of the video image).

Different algorithms can be used for the object detection. Given an arbitrary video frame, the goal of face detection algorithms is to determine whether or not there are any faces in the image, and if present, return the image

location and area (size) of each image of a face. Referring to figure 5, according to one exemplary embodiment of the present invention, an analysis window **33** is moved (or scanned) over the image. For each position of the analysis window **33**, the image information within the analysis window **33** is analyzed at least with respect to the presence of typical facial features. However, it should be understood that the present invention is not limited to the use of this type of face detection. Further, head detection algorithms may also be used to detect participants whose heads are not orientated towards the camera.

When an image of a face/head is detected, the video detection unit **30** defines a rectangular segment (or box) surrounding said image of a face/head. According to one embodiment of the invention, said rectangular segment is said analysis window **33**. The location of said segment containing an image of a face/head is measured relative to a video coordinate system which is based on the video frame. The video coordinate system applies to each frame captured by camera **6**. The video coordinate system has a horizontal or x-axis and a vertical or y-axis. When determining a position of a pixel or an image, video detection unit **30** determine that position relative the x-axis and the y-axis of that pixel's or image's video frame. In one exemplary embodiment of the invention, the analysis window **33** center point **31** (pixel in the middle of the window) is the location reference point, and its location is defined by the coordinates x and y in said video coordinate system. When the video detection unit 30 has calculated the location (x,y) and size (e.g. dx=20 dy=24 pixels) of all the faces/heads in a frame, the video detection unit 30 use knowledge of the video frame, optics and mechanics to calculate (step **80**) the corresponding location (α,φ) and size ($\Delta\alpha,\Delta\varphi$) in azimuth and elevation angles in the camera coordinate system for each image of a face/head. The camera coordinates for each face/head are then sent to a face tracking unit **35**.

Face tracking unit **35** correlates the detected faces from the current video frame to the detected faces in the previous video frames and hence tracks the detected faces through a series of frames. Only if a face/head is detected at substantially the same location throughout a series of frames, the detection is validated as a positive detection. First of all this prevents false face detections, unless the same detection occurs in several consecutive video frames. Also, if the face detection unit fails to detect a face in the substantially same coordinates as a face has been detected before, the image tracking unit does not consider the face as absent from the image unless the detection has failed in several consecutive frames. This is to prevent false negative detections. Further, the tracking allows for obtaining a proper position of a participant who may be moving in a video frame. To perform this tracking, face tracking unit **35** creates and maintains a track file for each detected face. The track file can for example be stored in a memory device.

In step **90**, the image processing unit **50** defines an area of interest **34** (best view). The area of interest **34** is shown in the in figure 6, where said area 34 at least comprises all the detected images of faces in that frame.

According to one embodiment of the invention, based on the location (α,φ) of each face and their corresponding sizes $(\Delta\alpha,\Delta\varphi)$, the image processing unit **50** may calculate a first area restricted by a set of margins (M_1, M_2, M_3 and M_4), where said margins are derived from the left side of the leftmost face segment (M_1), upper side of the uppermost face segment (M_3), right side of the rightmost face segment (M_2) and the bottom side of the bottommost face segment (M_4). The location of the center $(\alpha_{fa},\varphi_{fa})$ of said first area can now be calculated in camera coordinates based on said margins. The location of said first area is relative to a reference point (α_0,φ_0), typically the direction of the camera when azimuth and elevation angle is zero.

Further, the width and height of the first area is transformed into a zoom factor (Z_{fa}).

This first area is very close to the participants' faces and may not represent the most comfortable view (best view) of the participants, especially when only two participants are present as shown in this exemplary embodiment. Therefore, when said margins (M_1, M_2, M_3 and M_4) has been calculated, a second area (best view frame **34**) is defined by expanding said margins by a set of offset values **a,b,c** and **d**. These offset values may be equal, or they may differ, e.g. to capture more of the table in front of the participants than above a participants head. The offset values may be preset and static, or they may be calculated to fit each situation.

According to another exemplary embodiment, the best view frame **34** is defined by just substracting a compensation value Z_c from the calculated zoom factor Z_{fa}, making the camera zoom out an additional distance. The compensation value Z_c may be static, or vary linearly depending on the size of the first area zoom factor Z_{fa}.

Figure 7 schematically shows an examplary video frame taken from an initial camera orientation. 3 faces have been detected in the video frame, and the image processing unit **50** has defined a best view frame **34**, and calculated the location (α_{fa},φ_{fa}) of the best view frame.

Most image pickup devices 6 for video conferencing systems operate with standard television image aspect ratios, e.g. 4:3 (1.33:1) or 16:9 (1.78:1). Since most calculated best view frames **34** as described above have aspect ratios that deviate from the standards, e.g. 4:3 or 16:9, some considerations must be made when deciding the zoom coordinate. Since **Aφ** is the shortest edge of area 34, if the camera zooms in to capture the exact height **Aφ**, large parts of the area will miss the light sensitive area (e.g.

image sensor) in the camera because its aspect ratio is different than the defined area. If the camera zooms in to capture the exact width $\mathbf{A}\alpha$ of the defined area 34, no information is lost.

Therefore, according to one exemplary embodiment of the present invention, the two sides $\mathbf{A}_\varphi$ and $\mathbf{A}_\alpha$ of the best view frame 34 are compared. Each of the two sides defines a zoom factor needed to fit the area of interest in the image frame, in the horizontal and vertical direction respectively. Thus, the degree of zoom is defined by the smallest of the two calculated zoom factors, ensuring that the area of interest is not cropped when zooming to the area of interest.

In step **100**, the image processing unit **50** supplies the camera control unit **12** with the camera positioning directives $(\alpha_{fa}, \varphi_{fa}, Z)$ derived in step 90, via control unit 45. Once the camera positioning directives is received, the camera moves and zooms to the instructed coordinates, to get the best view of the participants in the video conference. Figure 8 shows the best view of participant **1** and **2** from meeting room **10a** in figure 6.

When the camera has moved to the new orientation, it will stay in that orientation until an event is detected (step 110). As mention earlier, the camera is only instructed to move the camera to an initial orientation (step 60) on certain predefined events. Such predefined events may include when the video conferencing system is started, when it is awaken from a sleep mode, when it receives or sends a conference call initiation requiest, when initiated by a user via e.g. remote control or keyboard, etc. Usually when an optimal view of the participants has been found, there is usually little need to change the orientation of the camera. However, situations may arise during a video conference that creates a need to reconfigure the orientation, e.g. one of the participants may leave, a new

participant may arrive, one of the participants change his/her seat, etc. Upon such situations, one of the users may of course initiate the repositioning (step 60) by pushing a button on the remote control. However, an automatic detection of such events is preferable.

Therefore, according to one embodiment of the present invention, audio source localization is used as an event trigger in step 110. As discussed above, figure 8 shows an optimal view of the 2 participants **1** and **2** in the large meeting room **10a**. As can be seen in figure 8, in this view the camera has been zoomed in quite extensively, and if a person was to enter the conference late and sit down in one of the chairs **12**, he/she would not be captured by the camera. When entering a meeting, it is natural to excuse yourself and/or introduce yourself. This is a matter of politeness, and to alert the other participants (that may be joining on audio only) that a new participant has entered the conference. By using known audio source localization arrangements **7;40**, the video conferencing system can detect that an audio source (participant) **200** has been localized outside of the current field of view of the camera. The audio source locator **40** operates in camera coordinates. When an audio source has been detected and localized by the audio source locator 40, it sends the audio source coordinates to the control unit **45**. Nothing is done if the audio source coordinates is within the current field of view of the camera. However, if the audio source coordinates is outside the current field of view, it indicates that the current field of view is not capturing all the participants, and the detection process according to steps 60-100 is repeated. The result can be seen in figure 10. Therefore, according to one embodiment of the invention, such detection of at least one audio source outside of the current field of view of the camera is considered as an event in step 110 triggering repetition of steps **60-100**.

Audio source localization arrangements are known and will not be discussed here in detail. They generally are a plurality of spatially spaced microphones **7**, and are often based on the determination of a delay difference between the signals at the outputs of the microphones. If the positions of the microphones and the delay difference between the propagation paths between the source and the different microphones are known, the position of the source can be calculated. One example of an audio source locator is shown in U.S. patent number 5,778,082.

According to another embodiment of the present invention another predefined event is when a participant is detected leaving the room (or field of view). Such detection depends on the tracking function mentioned earlier. As shown in figure 11a, when a participant leaves the room, the track file or track history will show that the position/location (α,φ) of a detected face changes from a position (α_3, φ_3) to a position (α_4,φ_4) close to the edge of the frame, over a sequence of frames (figure 11a-11b). If the same face detection suddenly disappears (no longer detecting a face) and do not return within a certain time frame (figure 11c), the face detection is considered as a participant leaving the conference. Upon detection of such an event, step 60-100 is repeated to adjust the field of view of the camera to a new optimal view as shown in figure 11d.

According to yet another embodiment of the present invention, another predefined event is when movement is detected near the edge of the video frame. Not everybody entering a video conferencing room will start speaking at once. This will depend on the situation, seniority of the participant, etc. Therefore, it may take some time before the system detects the new arrival and acts accordingly. Referring back to figure 9, parts 38 of a participant may be captured in the video frame, even though most of the person is outside the camera's field of view. Since people rarely sit completely still, relative to the static

furniture, the parts 38 can easily be detected as movement in the image by video detection unit 35. Upon detection of such an event (movement is detected near the image/frame edge), step 60-100 is repeated to adjust the field of view of the camera to a new optimal view.

The systems according to the present invention provide a novel way of automatically obtaining the best visual representation of all the participants in a video conferencing room. Further, the system is automatically adapting to new situations, such as participants leaving or entering the meeting room, and changing the visual representation accordingly. The present invention provides a more user friendly approach to a superior visual experience.

Patent claims

1. A method for automatically steering the orientation and zoom of an image pickup device associated with a video conferencing system, wherein said method comprises the steps of

generating, at said image pickup device, an image signal representative of an image framed by said image pickup device, and

processing the image signal to identify objects in said image, and, if predetermined events occur:

steering the image pickup device to an initial orientation,

determining the location of all identified objects relative to a reference point and their respective sizes,

defining an area of interest in said image, where said area of interest at least comprises all identified objects,

steering the image pickup device to frame said defined area of interest.

2. A method according to claim 1, where said step of steering the image pickup device comprises the sub steps of

varying the azimuth angle and elevation angle of the image pickup device, and

varying the zoom of the image pickup device.

3. A method according to claim 1, where said step of steering the image pickup device to an initial orientation further comprise the sub steps of

zooming the image pickup device out to a maximum field of view and moving the image pickup device according to a predefined pan- and tilt-sequence, framing as much as possible of a room in which it is situated.

4. A method according to claim 1, wherein the image signals represent frames of video images, and the step of identifying objects further comprise the sub step of

detecting images of the faces and/or heads in said frames of video,

tracking the detected faces/heads through a series of frames,

identify a detection as a face/head only if said detection occurs in all of a predefined number of succeeding frames.

5. A method according to claim 4, where the step of defining an area of interest further comprise the sub steps of

defining a set of margins for a first area, where said first area is the smallest definable area enclosing all of said detected images of faces and/or heads, and

defining said area of interest by expanding said margins by a set of offset values.

6. A method according to claim 5, where the area of interest is further expanded to fit a standard image aspect ratio.

7. A method according to claim 1, wherein said area of interest represents a close-up view of an object or a group of objects.

8. A method according to claim 1, where said predefined events comprise,

switching on the video conferencing system, receiving or sending a conference call initiation request, and/or receiving a command from a user.

9. A method according to claim 1, where said method further comprises,

processing an audio signal from a set of audio pickup devices, to determine the location of an audio source relative to a reference point.

10 A method according to claim 8, where the said predefined events comprises,

detecting the presence of an audio source outside the framed area of interest.

11. A method according to claim 1, where the said predefined events comprises,

detecting the disappearance of one or more of the participants from the framed area of interest.

12. A system for automatic steering of the orientation and zoom of an image pickup device associated with a video conferencing system, wherein said image pickup device generates image signals representative of an image framed by said image pickup device, where the system comprises a video detection unit configured to process the image signal to identify objects in said image, and determine the

location of all identified objects relative to a reference point and their respective sizes,

c h a r a c t e r i z e d i n that it further comprises

an image processing unit configured to define an area of interest in said image, where said area at least comprises all identified objects,

a control unit configured to, on predefined events,

steer the image pickup device to an initial orientation,

receive camera coordinates from said image processing unit corresponding to said area of interest,

steer the image pickup device to frame said defined area of interest.

13. A system according to claim 12, wherein the image signals represent frames of video images, and where the identified objects are detected images of faces and/or heads in said frames of video.

14. A system according to claim 13, further comprising

a face tracking unit configured to track the detected faces/heads through a series of frames, and

identify a detection as a face/head only if said detection occurs in all of a predefined number of succeeding frames.

15. A system according to claim 13, wherein said image processing unit is further configured to

define a set of margins for a first rectangular area, where said first area is the smallest definable area enclosing all of said detected images of faces and/or heads, and

define said area of interest by expanding said margins by a set of offset values.

16. A system according to claim 15, where the area of interest is further expanded to fit a standard image aspect ratio.

17 A system according to any of claims 13 - 16, the system further comprising,

an audio source locator configured to process an audio signal from a set of audio pickup devices, to determine the location of an audio source in camera coordinates.

18. A system according to any of claims 13 - 17, wherein said control unit is further configured to,

receive audio source coordinates from said audio source locator,

compare said audio source coordinates to current field of view.

19. A system according to any of claims 13 - 18, wherein said image pickup device comprises a camera control unit for positioning said image pickup device, wherein the control unit supplies control signals to said camera control unit for the orientation and zoom of the image pickup device, the control signal being generated based on said area of interest.

20. A system according to one of the proceeding claims, where the said predefined events comprises,

detecting the presence of an audio source outside the current field of view.

21. A system according to one of the proceeding claims, where the said predefined events comprises,

detecting the disappearance of one or more of the participants from the framed area of interest.

22. A system according to one of the proceeding claims, where the said predefined events comprises,

detecting the presence of an audio source outside the currently framed area of interest.

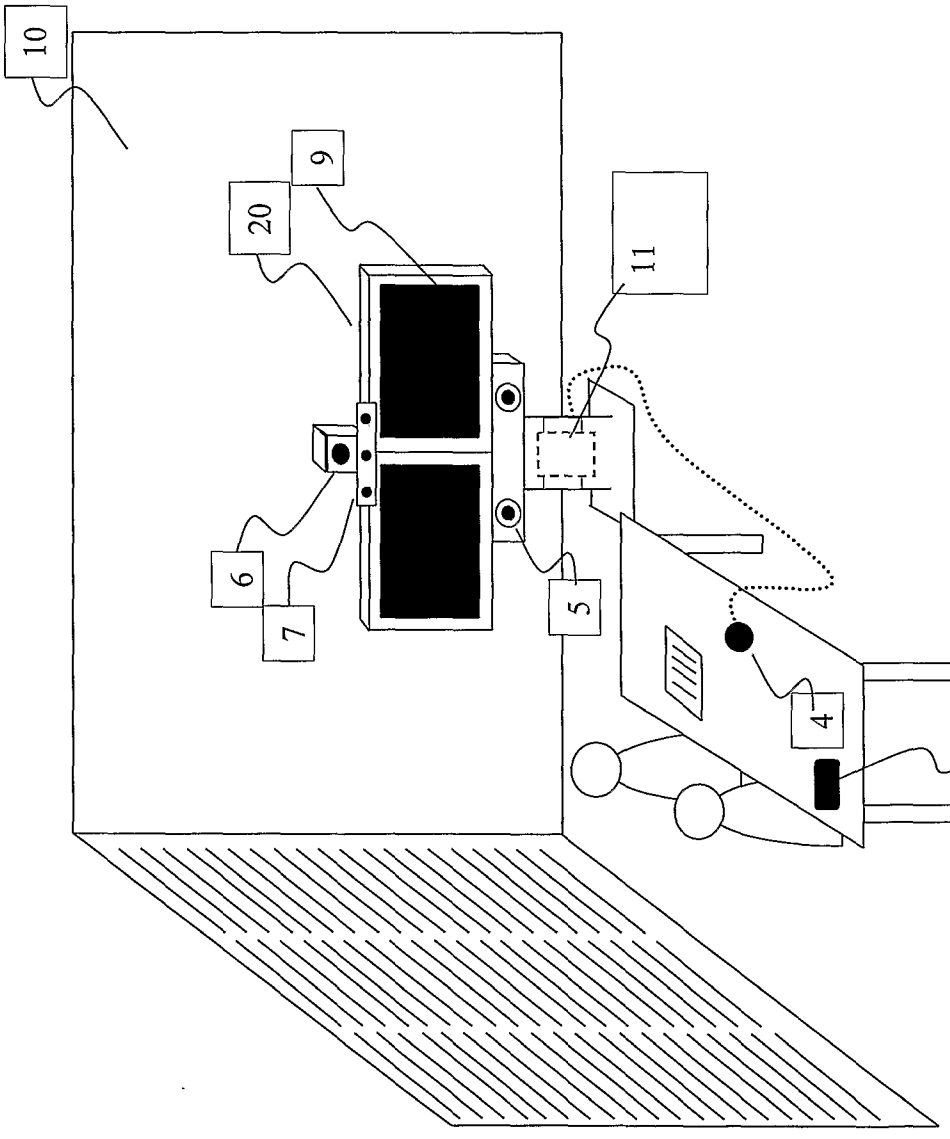

Figure 1

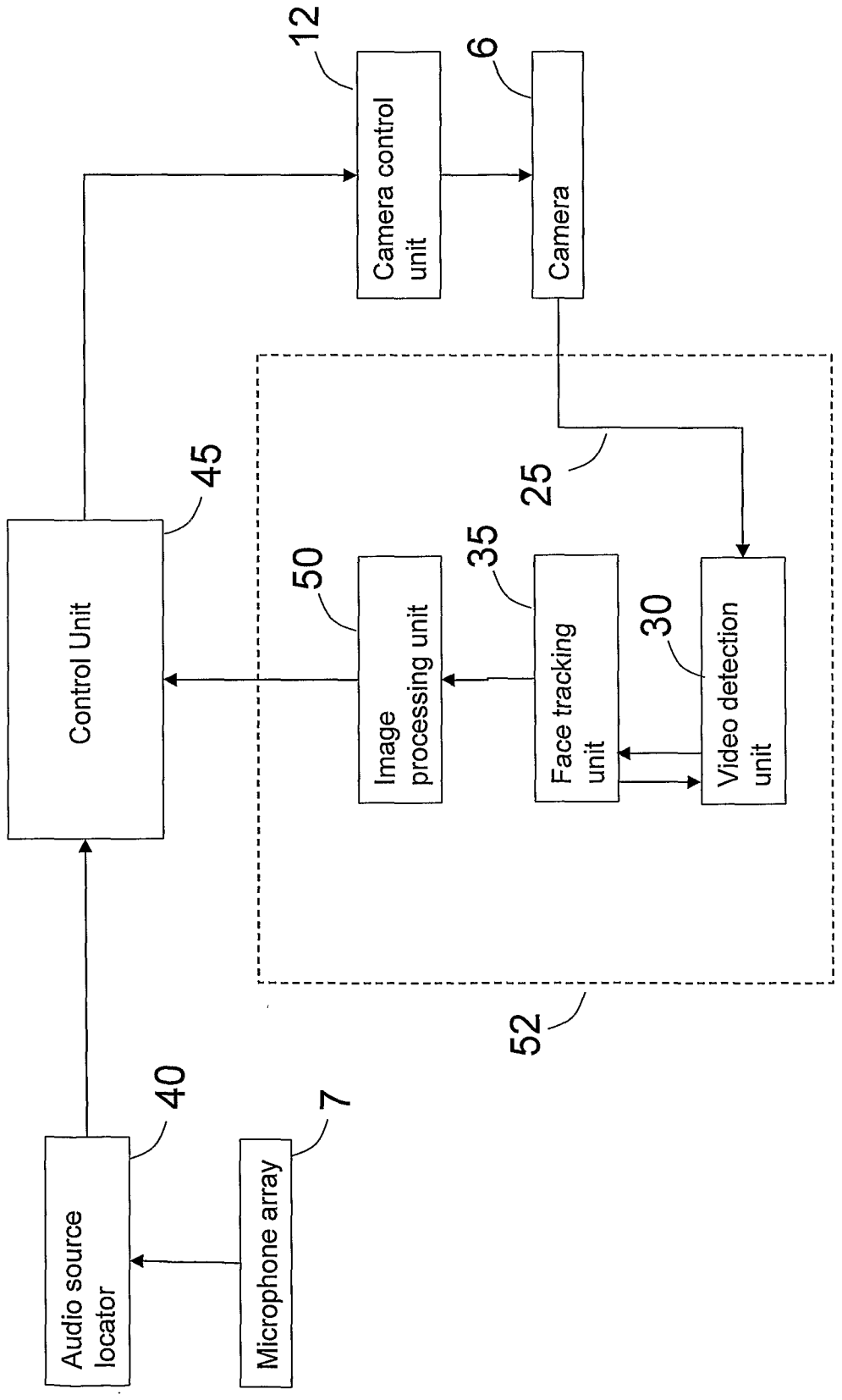
12
Camera control unit
6
Camera
25
45
Control Unit
50
Image processing unit
35
Face tracking unit
30
Video detection unit
52
40
Audio source locator
7
Microphone array

Figure 2

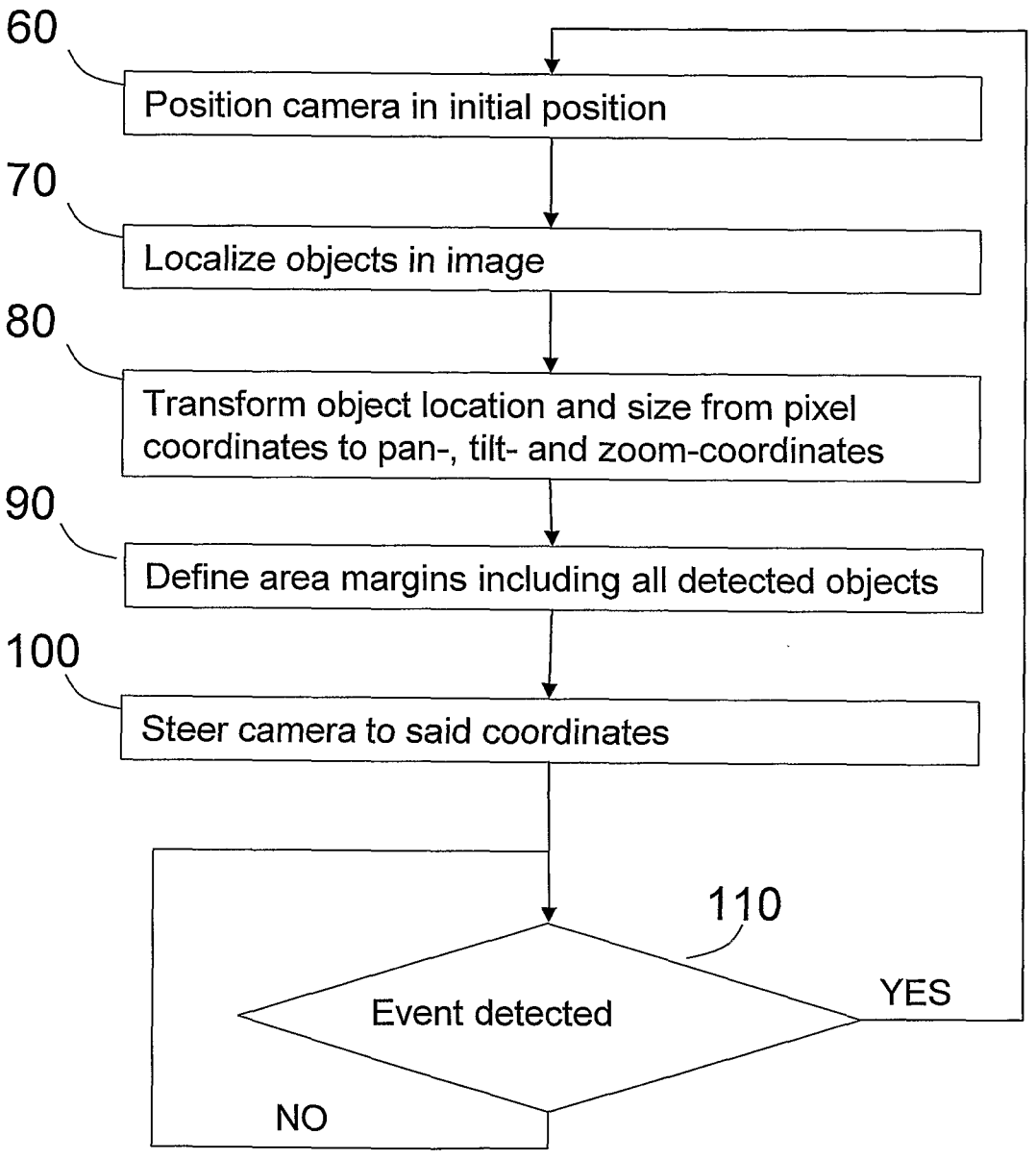
60
Position camera in initial position
70
Localize objects in image
80
Transform object location and size from pixel coordinates to pan-, tilt- and zoom-coordinates
90
Define area margins including all detected objects
100
Steer camera to said coordinates
110
Event detected
YES
NO

Figure 3

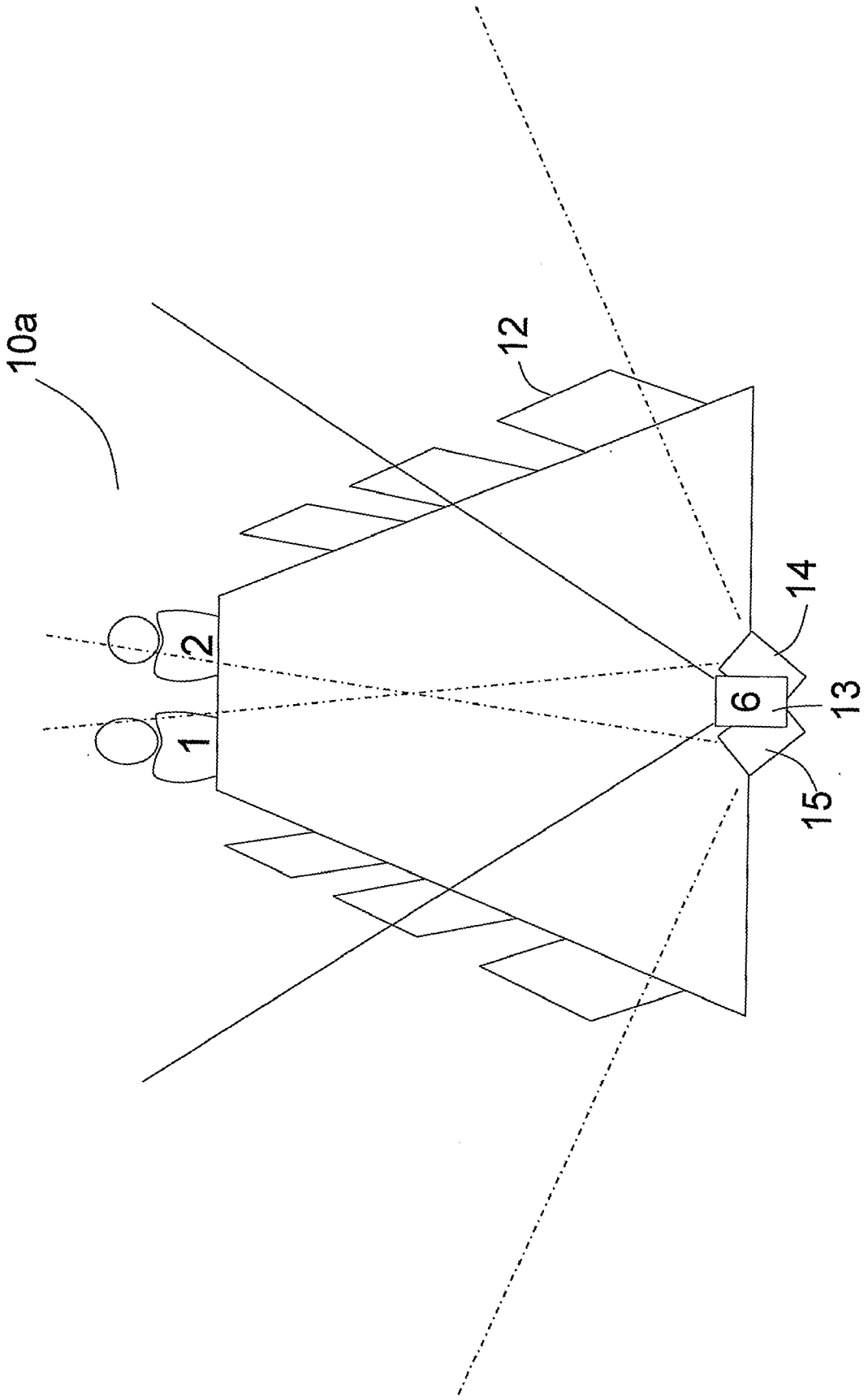

Figure 4

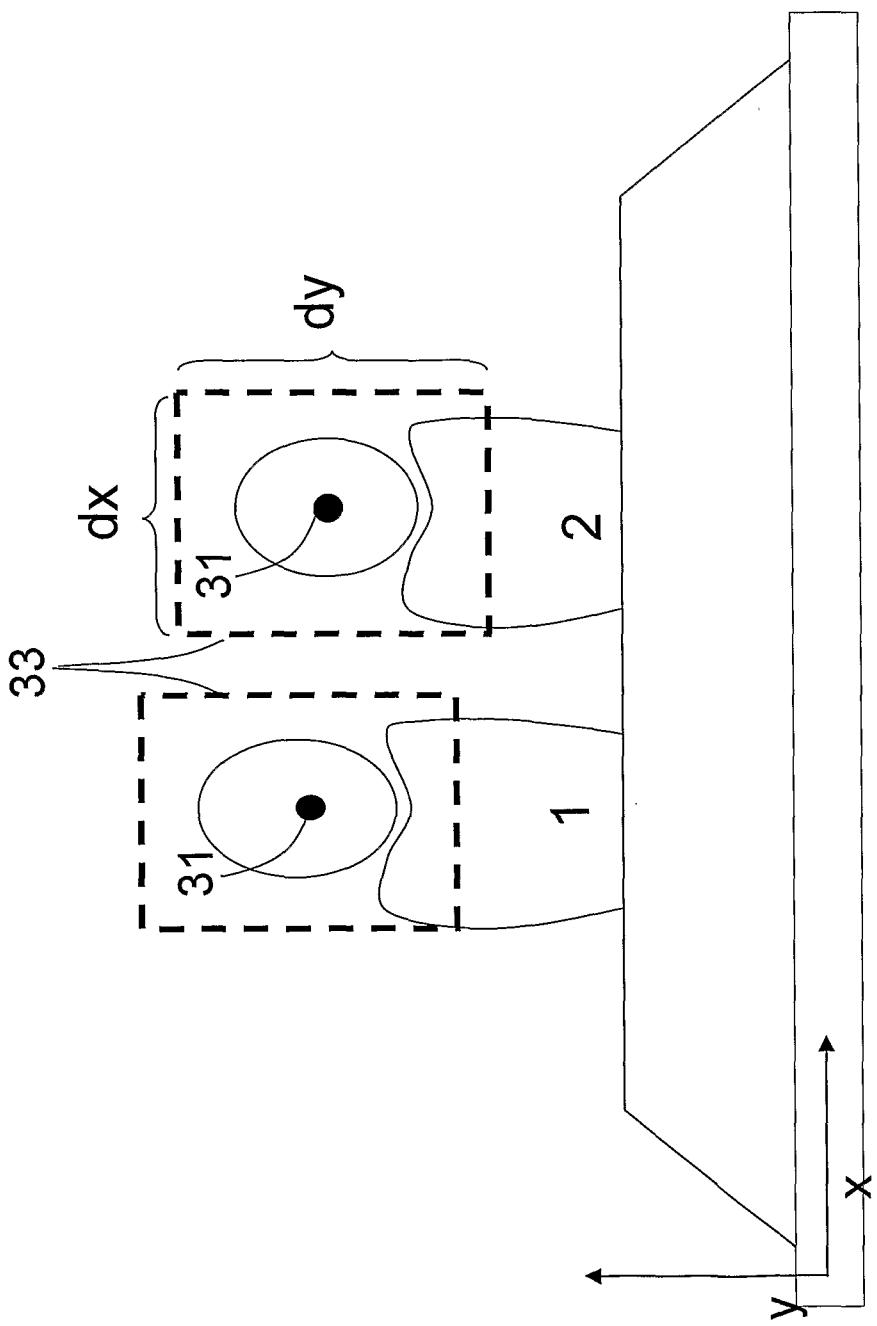

Figure 5

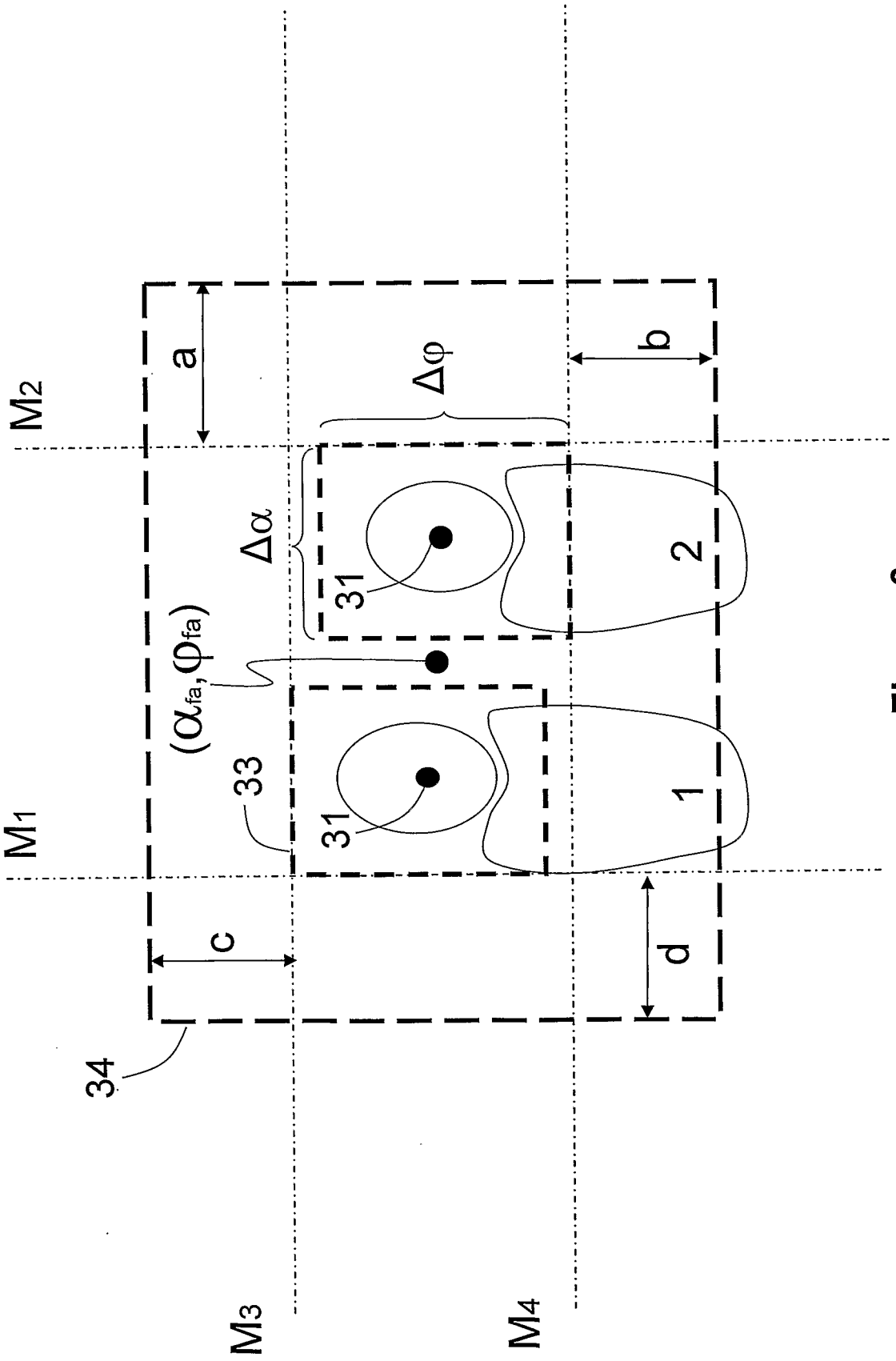

Figure 6

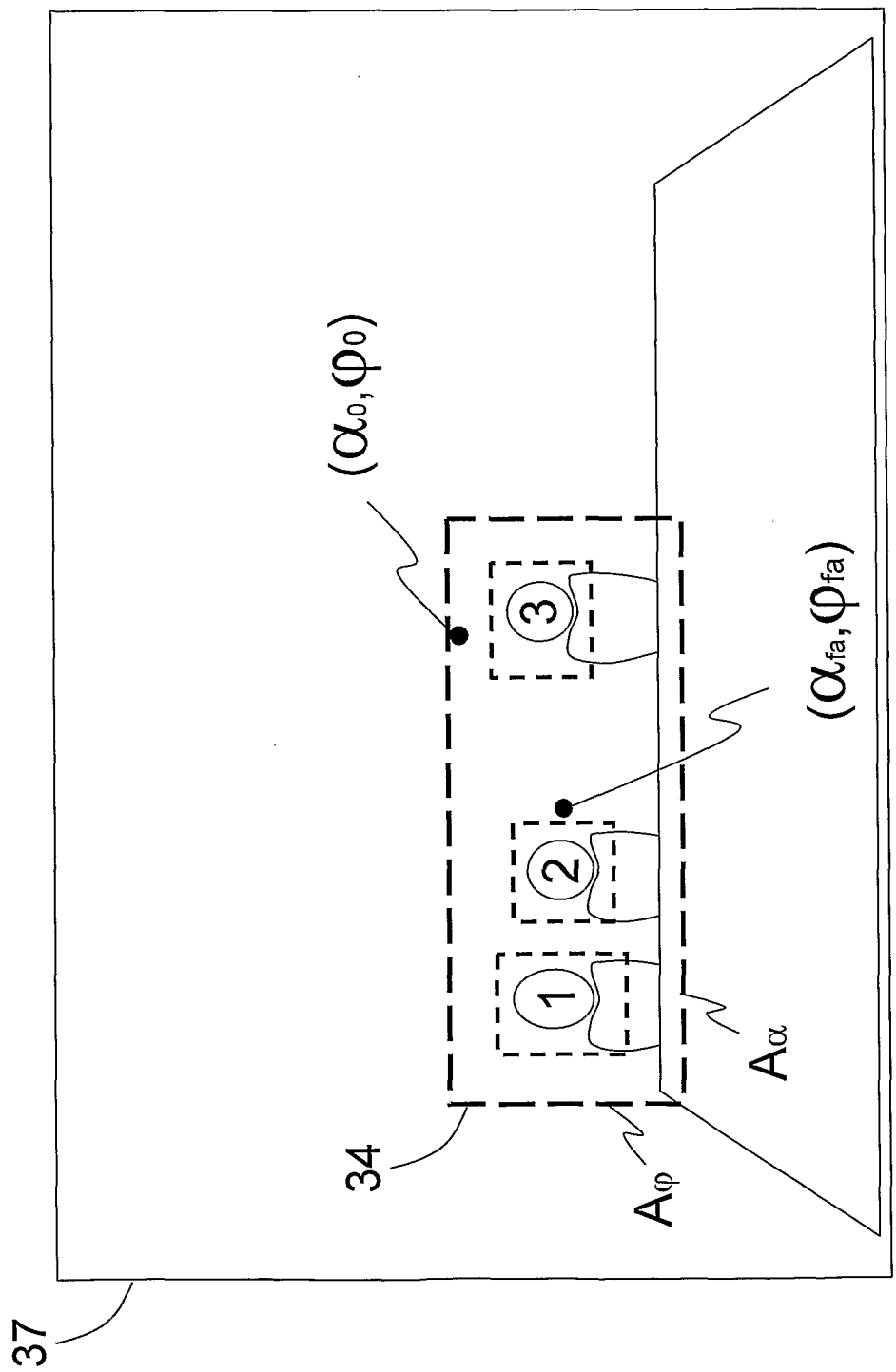

Figure 7

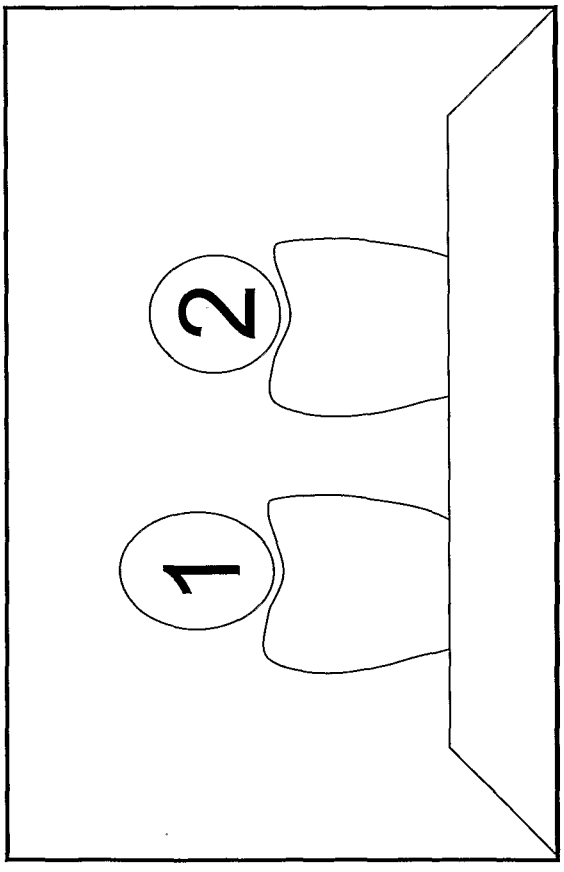

Figure 8

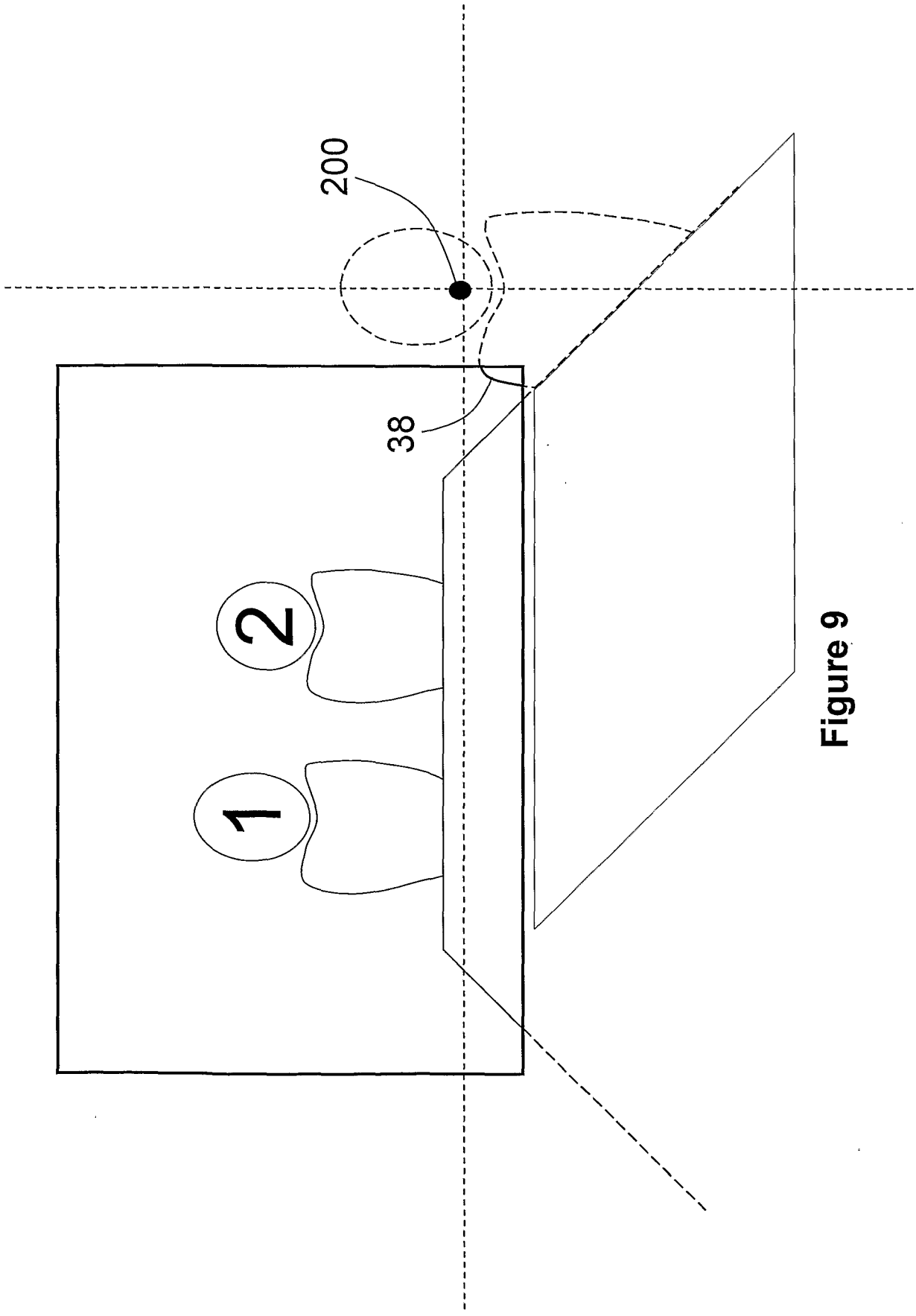

Figure 9

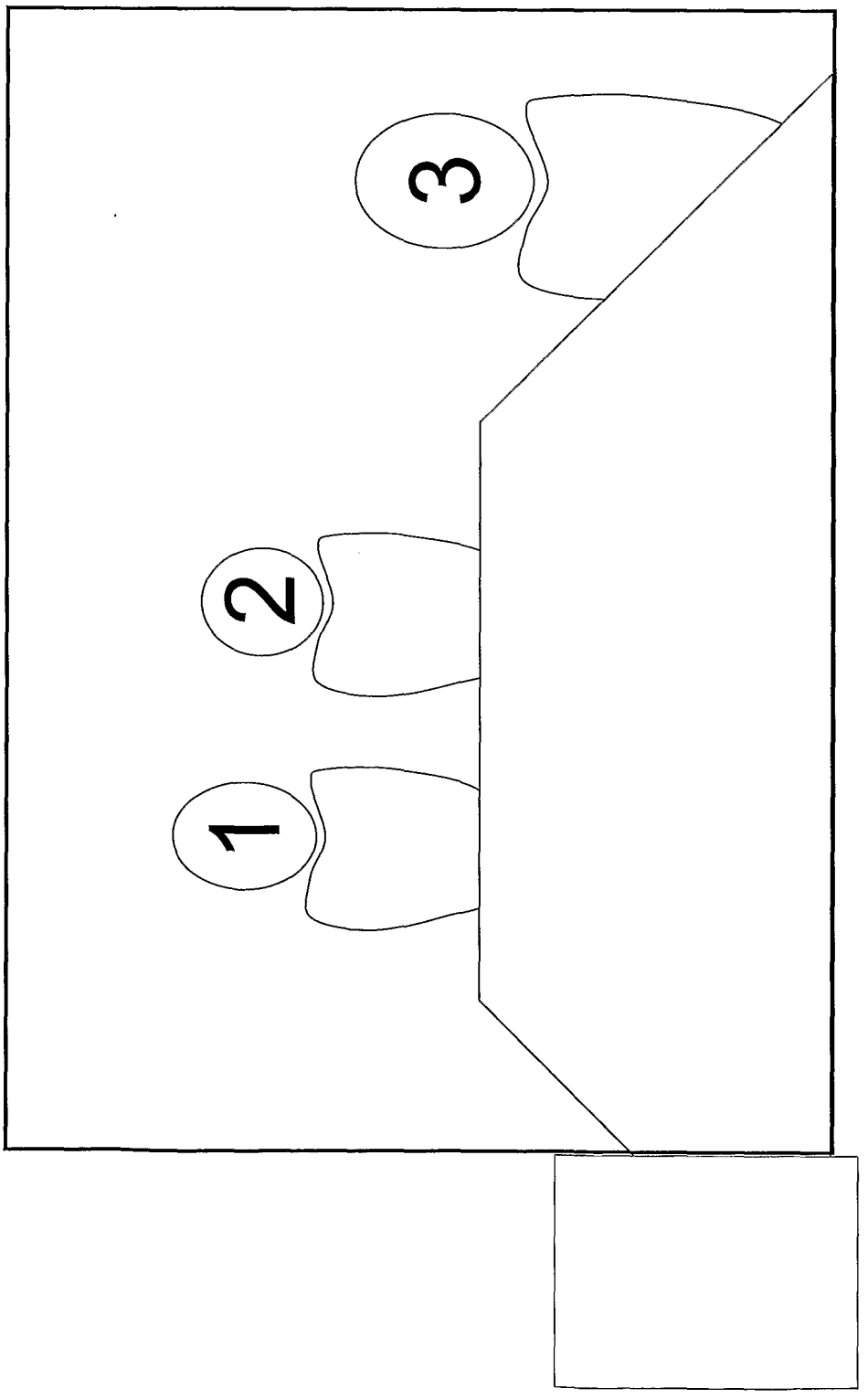

Figure 10

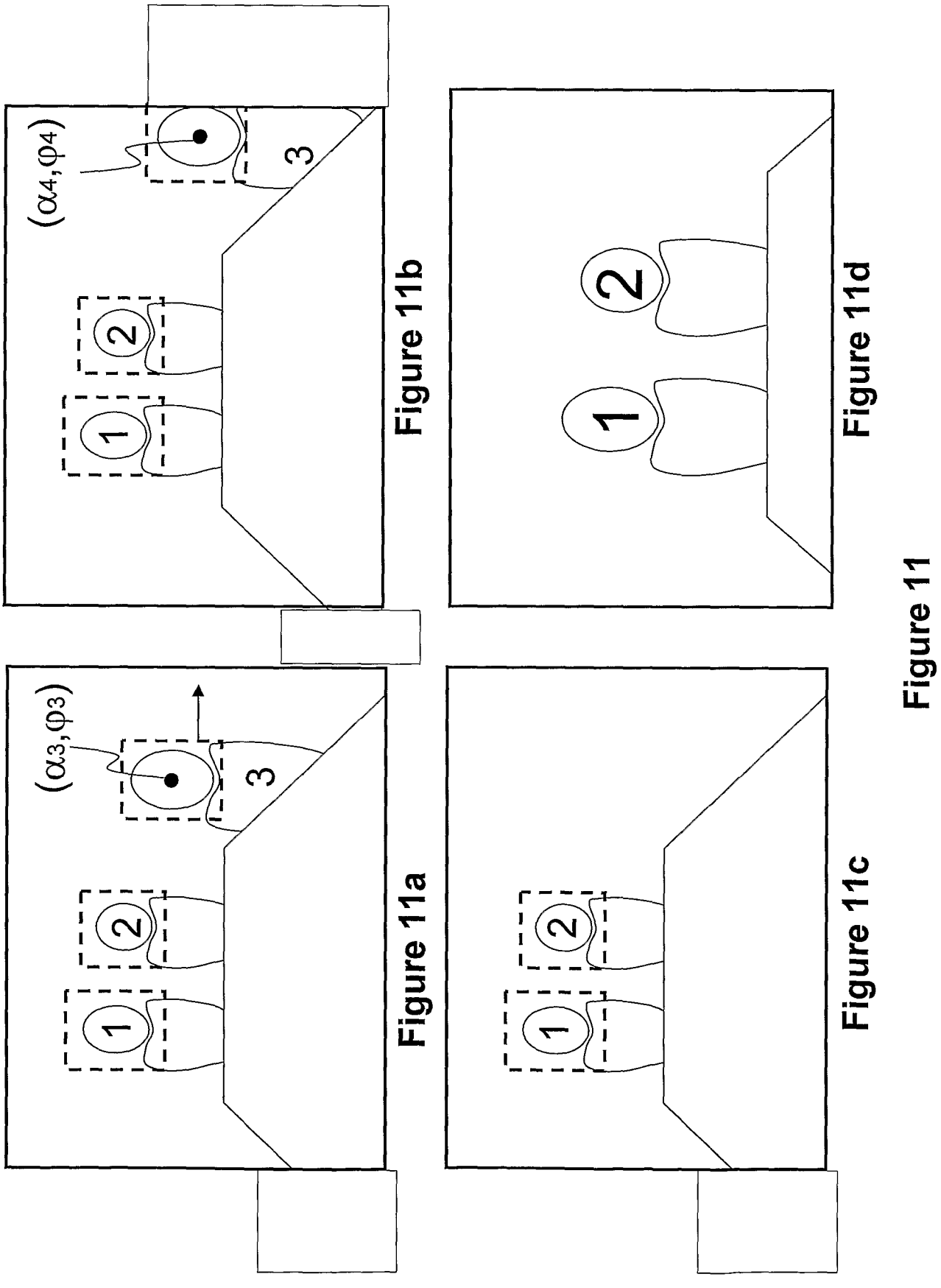

Figure 11a

Figure 11b

Figure 11c

Figure 11d

Figure 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NO2008/000249

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, WPI DATA, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 9960788 A1 (PICTURETEL CORP), 25 November 1999 (25.11.1999) --	1-22
A	US 20040257432 A1 (GIRISH, M K ET AL), 23 December 2004 (23.12.2004) --	1-22
A	US 20020140804 A1 (COLMENAREZ, A J ET AL), 3 October 2002 (03.10.2002) --	1-22
A	US 20030103647 A1 (RUI, Y ET AL), 5 June 2003 (05.06.2003) --	1-22

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 October 2008	28 -10- 2008
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Henrik Andersson /PR Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NO2008/000249

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5852669 A (ELEFTHERIADIS, A ET AL), 22 December 1998 (22.12.1998) --	1-22
A	WO 9906940 A1 (INTERVAL RESEARCH CORP), 11 January 1999 (11.01.1999) -- --------	1-22

International patent classification (IPC)
H04N 7/15 (2006.01)

Download your patent documents at www.prv.se
The cited patent documents can be downloaded at www.prv.se by following the links:

- In English/Searches and advisory services/Cited documents (service in English) or
- e-tjänster/anförda dokument (service in Swedish).

Use the application number as username.
The password is **JURFVFCLAE**.

Paper copies can be ordered at a cost of 50 SEK per copy from PRV InterPat (telephone number 08-782 28 85).

Cited literature, if any, will be enclosed in paper form.

WO	9960788	A1	25/11/1999	AU	3367899	A	18/10/1999
				AU	6308799	A	05/03/2001
				CA	2323754	A	07/10/1999
				DE	69920138	D,T	03/02/2005
				EP	1004204	A,B	31/05/2000
				EP	1068342	A	17/01/2001
				JP	2002516535	T	04/06/2002
				JP	2002529049	T	10/09/2002
				US	6495738	B	17/12/2002
				US	6593956	B	15/07/2003
				WO	9950430	A	07/10/1999
US	20040257432	A1	23/12/2004	WO	2005002225	A	06/01/2005
				US	20040257431	A	23/12/2004
				US	20060163625	A	27/07/2006
				US	20080218583	A	11/09/2008
				WO	2005002193	A	06/01/2005
US	20020140804	A1	03/10/2002	CN	1460185	A,T	03/12/2003
				CN	100370830	C	20/02/2008
				EP	1377847	A	07/01/2004
				JP	2004528766	T	16/09/2004
				WO	02079792	A	10/10/2002
US	20030103647	A1	05/06/2003	AT	397354	T	15/06/2008
				CN	1423487	A	11/06/2003
				CN	100334881	C	29/08/2007
				CN	101093541	A	26/12/2007
				EP	1330128	A,B	23/07/2003
				EP	1838104	A	26/09/2007
				JP	2003216951	A	31/07/2003
				KR	20030045624	A	11/06/2003
				TW	222031	B	11/10/2004
				US	7130446	B	31/10/2006
				US	7151843	B	19/12/2006
				US	7171025	B	30/01/2007
				US	7428315	B	23/09/2008
				US	20050129278	A	16/06/2005
				US	20050147278	A	07/07/2005
				US	20050188013	A	25/08/2005
				US	20050210103	A	22/09/2005
US	5852669	A	22/12/1998	CN	1118961	A	20/03/1996
				DE	69523503	D,T	11/07/2002
				EP	0676899	A,B	11/10/1995
				US	5500673	A	19/03/1996
				US	5512939	A	30/04/1996
				US	5548322	A	20/08/1996
				US	5550580	A	27/08/1996
				US	5550581	A	27/08/1996
				US	5596362	A	21/01/1997
				CA	2177866	A	11/01/1997
				EP	0753969	A	15/01/1997
				JP	9035069	A	07/02/1997

WO	9906940	A1	11/01/1999	AU	8584898	A	22/02/1999
				EP	0998718	A	10/05/2000
				US	6188777	B	13/02/2001
				US	6445810	B	03/09/2002
				US	20010000025	A	15/03/2001