(19) DANMARK

(10) DK/EP 2965172 T3

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: *G 06 F 3/01 (2006.01)* *G 03 B 5/06 (2021.01)* *H 04 N 13/395 (2018.01)*

(45) Oversættelsen bekendtgjort den: **2022-06-20**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2022-06-01**

(86) Europæisk ansøgning nr.: **14723500.6**

(86) Europæisk indleveringsdag: **2014-02-24**

(87) Den europæiske ansøgnings publiceringsdag: **2016-01-13**

(86) International ansøgning nr.: **IT2014000048**

(87) Internationalt publikationsnr.: **WO2014136131**

(30) Prioritet: **2013-03-05 IT PE20130004**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **Blue Cinema TV s.r.l., Via Carlo Armellini 16, 00153 Roma, Italien**

(72) Opfinder: **BALDACCI, Daniele, c/o Blue Cinema TV S.A.S. Di Daniele Baldacci & C., Via Benedetto Croce 35/1, I-65015 Montesilvano (PE), Italien**

(74) Fuldmægtig i Danmark: **Gotapatent AB, Box 3127, S-550 03 Jönköping, Sverige**

(54) Benævnelse: **FREMGANGSMÅDE TIL FREMSTILLING AF ET AUDIOVISUELT INTERFACE**

(56) Fremdragne publikationer:
WO-A1-2009/155926
WO-A2-98/10584
US-A1- 2011 007 079
US-A1- 2011 304 632
US-B1- 7 844 467

DESCRIPTION

[0001] The present invention refers to a process for making an audio-visual interface.

[0002] The most modern prior art in terms of interface provides for voice- or gesture-type interaction modes (Apple Iphone Siri, last-generation video game consoles) whose communication protocol is based on the ratio between a human user and an apparatus (telephone, computer, console).

[0003] The recent patents filed by Samsung and Apple and other companies operating in the field of electronics and telephony, involve the human perception in a more and more complete way (recognition of the eye movement).

[0004] The great change in the media evolution therefore comes from the new modes for interacting with people, for example from different user interface forms.

[0005] US-B1-7 844 467 and US-A1-2011/007079 disclose processes according to the prior art.

[0006] Object of the present invention is providing a process which proposes as result a simple and natural interface which can be used by everybody and, at the same time, is most involving: an interactive real human being (filmed and not generated with computer graphics), reproduced with a three-dimensional effect to increase the visual involvement.

[0007] Such interface allows passing from a third generation of electronic apparatuses and computers which are invisible, cannot even be seen, and are naturally adapted to the ways in which one works and lives his everyday life.

[0008] There are multiple potential uses and applications of the process: as interface of new-generation electronic apparatuses and computers, in which always more and more interactivity is required with a user; as info points in museums, showrooms, stations, airports, towns, parks, etc. where, differently from the past, multi-sensor and interactive aspects are preferred in order to increse the viewer's involvement, who, seeing the human interface, will be able to access to information through a simple talk.

[0009] The invention allows the chance of making a visual interface in the language of signs (LIS) towards apparatuses, and a simultaneous translator from verbal language into LIS.

[0010] Such process brings about a result that, in addition to having a value per se, is a paradigmatic change in the history of relationships between user and apparatus.

[0011] The extreme easiness of such relationship (with simple "talk") makes the users of any type of apparatus extended also to all those that have difficulties in interfacing with modern

technologies, in addition to handicapped people.

[0012] The above and other objects and advantages of the invention, as will appear from the following description, are obtained with a process for making an audio-visual interface as claimed in claim 1. Preferred embodiments and non-trivial variations of the present invention are the subject matter of the dependent claims.

[0013] It is intended that all enclosed claims are an integral part of the present description.

[0014] The present invention will be better described by some preferred embodiments thereof, provided as a non-limiting example, with reference to the enclosed drawings, in which:

- Figure 1 is a schematic view of the operating environment in which the process according to the present invention operates;
- Figures 2 and 3 are flow diagrams of the main steps of the process of the present invention;
- Figures 4 to 7 are some possible results of the process of the present invention; and
- Figure 8 shows the final image which is obtained with the process of the present invention.

[0015] With reference to the Figures, a preferred embodiment of the process for making an audio-visual interface of the present invention is shown and described.

[0016] The inventive process for making an audio-visual interface is developed in three steps:

FIRST STEP: interactivity analysis.

SECOND STEP: image acquisition

THIRD STEP: reproduction through optical (or television) system

FIRST STEP: INTERACTIVITY ANALYSIS

[0017] The process provides for the use of an optical apparatus which reproduces with three-dimensional effect the human presence (described in the THIRD STEP), equipped with a sensor/controller 3 (with its related software) capable of voice recognising, able to interpret the human language.

[0018] The interpretation of the human language allows identifying the information semantics.

[0019] The system is able to trigger the appropriate event depending on the received voice command, and communicates with a computer through an interface to connect the external

devices, such as a USB port, a parallel port, a standard serial port (Figure 1).

[0020] The reproduction system can access a series of "n" actions by the reproduced subject through a logic programming. A volumetric sensibility (with suitable sensor 3) allows the system to perceive the presence, absence and related status change (arrival, good-bye), in addition to a series of movements by the observer/user 1. Consequent actions by the system can be made and programmed. For example, upon programming, the reproduction of the answer segment: "The book is used to fetch information", could be associated with the questions "what is the book?", "what is the book used for?", or simply "book" and many others.

[0021] Once having outlined the specific interactive project, visual and audio shootings are made (SECOND STEP).

[0022] The chosen actor will recite all provided segments (ACTIONS), in addition to a whole series of non-verbal actions (IDLE, PRE-IDLE and BRIDGE) which will be used to provide the interface with such presence effect due to a simulation of the human behaviour during the dialogue.

[0023] The visual sequences to shoot will be segmented according to an interactivity logic to obtain the sensation of talking with a human being.

[0024] The logic sequences are divided into four main categories (to which the sub-categories for more complex projects are added):

1. 1) IDLE sequence (indifference status in which the interface "lives" waiting for a status change command);
2. 2) PRE-IDLE sequence (a behaviour preluding to the action);
3. 3) BRIDGE sequence (an image transition which helps to give visual continuity to the interface between the IDLE/PRE-IDLE status and the ACTION status)
4. 4) ACTION sequence (in which the interface performs or makes the action performed)

[0025] For example, when the interface does not perform actions, it will have to be in an indifference (IDLE) status, for example looking around or whistling. When a question is inserted, its attention status changes (PRE-IDLE) looking at a user with its face, in order to then analyse the performed question and proceed with the related answer or programmed action (ACTION).

[0026] An invitation action to repeat the question could be provided for non-programmed questions.

[0027] Statuses detected by the volumetric sensor 3 will also be analysed and the interface behaviour, reaction or action will be programmed for interacting with the utente 1 according to the specific interactive project. Figure 1 also graphically shows the audio-volumetric sensibility

of the microphone and of the sensor.

[0028] An accurate choice of the IDLE, BRIDGE and ACTION statuses is then performed, together with their logic presentation within a behaviour simulated by the reproduced human subject.

[0029] Like in a flow diagram, everything is studied, made, programmed according to a logic whose result is displaying a human being careful to give a series of "n" replies on "n" specific subjects in "n" languages or perform commands, everything with a "real" behaviour (Figurr 2). In such Figure, in step F1 IDLE 01, 02, 03 etc. are performed, while in F2 the question is listened to and in F3 the answer is analysed ("it is possible to reply?"). If the outcome is negative, in F4 an invitation is send to rephrase the question, while if it is positive, in F5 BRIDGE is performed and in F6 ACTION 01, 02, 03, ecc. are performed.

[0030] The "human" interface allows driving peripherals outside the optical reproduction system: domotic peripherals, DMX controlled lights systems for shows or events, the videoprojection of an audio-visual contribution on a large screen external to the system or any other electric or electronic apparatus equipped with standard communication protocols (for example USB, serial, parallel, Wifi, Bluetooth).

[0031] An inverse interactive situation can be provided, in which the interface, after having shown a certain subject required by the user, can also make questions to the user and analyse the answers to understand and communicate the user's degree of preparation (FIGURE 03). In such Figure, in step F7, IDLE 01, 02, 03 etc. are performed, while in F8 a subject is selected and in F9 the subject is analysed ("is it possible to go on?"). If the outcome is negative, in F10 an invitation is made to perform again the subject selection, while if it is positive, in F11 BRIDGE is performed and in F12 ACTION 01, 02, 03, ecc. are performed, related to the selected subject. Then, in F13 n questions are made on the selected subject, in F14 the answers are listened to, in F15 the answers themselves are analysed (for all n questions - F15A) and in F16 the evaluation is shown related to the analyses of the answers.

SECOND STEP: IMAGE ACQUISITION

[0032] The process implies the acquisition of images by using a particular kinematographic technique described below.

[0033] The subject 7 (aimed to become the protagonist of the interface described in this process) has to be shot with a high-definition digital shooting machine 5 with interlaced scanning in vertical, instead of horizontal, position (Figure 4).

[0034] The need for a vertical orientation stems from the fact that the subject 7 to be shot in the interface is a standing or sitting human being whose orientation is vertical.

[0035] All the shooting space is then developed in height and such arrangement of the shooting machine 5 allows acquiring images with a vertical resolution of at least 1920 pixel instead of 1080.

[0036] As result, there will be the perception of the general definition equal almost to twice with respect to a normal high definition FULL HD image (bottom of Figure 4).

[0037] In order to obtain the best realistic effect similar to the real visual perception, on the digital shooting machine 5 oscillating tilt and shift optics are assembled to obtain field depth according to the Scheimpflug rule or condition.

[0038] It states that, for an optical system, the focal plane generated by the objective and the one of the subject 7 meet on the same straight line.

[0039] As limit case, when the three planes are mutually parallel, they meet at infinite.

[0040] When the plane on which the subject 7 rests, the rear nodal plane and the focal plane meet in a single point, the full focusing of the subject 7 is obtained, independently from the used diaphragm.

[0041] Such wide field depth favours the visual perception of a three-dimensionality on the edges of the shot subject 7.

[0042] The herein described shooting technique assumes an illumination scheme with well defined contrasts.

[0043] Figura 5 shows, on the right, the scheme in which the shooting machine 5 frames the subject 7 on a black background 9 through a projector 11 (preferably of the type with 500 W and Fresnel lens) and softbox 13 (preferably 1000 W).

[0044] The target is obtaining images of the subject 7 with widespread and uniform illumination on the front to obtain a general opacity of the subject 7 who otherwise will result transparent in the THIRD STEP (reproduction).

[0045] An illumination of well accurate side cuts and backlights will have to draw the subject 7 on his whole perimeter to increase the three-dimensionality effect and have well defined edges detached in the empty space (Figure 5).

[0046] The subject 7 will have to be in total absence of lighting in the areas surrounding his edge and thereby the effect called "edge cutting" will be obtained, necessary for reproducing the images described in the third step.

[0047] In the optical system, the image is projected by a LED monitor matrix 26 in a reverse vertical position on a plane made of transparent polycarbonate 28 placed at 45° with respect

thereto (Figure 8).

[0048] All the image portion not containing lightness is non reflecting and therefore transparent, obtaining the effect of suspension of the illuminated subject 7 in the empty space with a final "edge cutting" effect (Figure 6, in which on the right there is an image generated by a LED monitor matrix and on the left the same image reflected on transparent polycarbonate).

[0049] The strong contrast of the edges (area 16) and the special field depth obtained with the oscillating optics help obtaining in the reflection a strong detachment 22 of the subject 7 (with respect to the totally transparent area 20), while the widespread illumination 18 (with respect to the area with total lack of lightness 14) of the subject 7 is translated into an opacity 24 on its whole surface.

[0050] Particular care must be paid to perspective ratios and the accurate reproduction of the actual measures of the shot subject/object 7.

[0051] The "dummy pact" with the user 1, the perception of having a human being in front of you, is constrained to the condition that the images reproduced in the interface, through the optical system, should have the exact sizes of the shot subject 7.

[0052] As can be seen in Figure 7, a person 170 cm tall in reality will have to be reproduced with an image exactly 170 cm tall.

[0053] The perspective of an average user 1 in front of the interface will have to comply with those of a look towards a human being placed in front of him.

[0054] In order to obtain the right perspective, the following parameters will have to be balanced: focus, shooting machine 5/subject 7 distance, view angle (Figure 7).

[0055] The status passages from IDLE (IDLE_01, IDLE_02, IDLE_03...IDLE_n) to PRE-IDLE (PREIDLE_01, PREIDLE_02, PREIDLE_03...PREIDLE_n), to BRIDGE, to ACTION (ACTIC)N_01, ACTION_02, ACTION_03...ACTION_n) are very delicate, and careful techniques for checking the shooting perspectives will have to be observed with the help of reference grids overlapped to the wiewfinder of the shooting machine 5.

[0056] In such a way, the shot subject 7 will have the same proportions in the various segments to be programmed and the related transitions will be homogeneous and will not interfere on the perception, by the user 1, of the action continuity by the subject 7 reproduced at the interface.

THIRD STEP: REPRODUCTION

[0057] For the final reproduction of the process, an optical reproduction system is used, as

described below and shown in Figure 8.

[0058] However, other displaying systems, some simpler (like a HD monitor), or other optical systems (suitable to reproduce the reality effect of the subject 7/interface) can be used as visual output of the interface described in this process.

[0059] The optical reproduction system allows, through the reflection of images generated by a LED monitor matrix 26 or other FULL HD source, having the optical illusion of the total suspension of an image in the empty space.

[0060] The reproduced images are reflected 28 on a slab made of transparent polycarbonate, 1 mm thick (thickness which guarantees the reduction to a minimum of the refraction of the two surfaces of the slab itself, preventing them from overlapping).

[0061] The final image perceived by the viewer is on a vertical plane 30 behind the polycarbonate slab, completely detached from the background 32, with which a strong parallax effect 34 is formed (Figure 8) .

[0062] Let us analyse, for example, the application of the process to the making of a "butler" interactive interface in a hotel suite.

[0063] Entering the suite, a three-dimensional image reproduced in the optical system (described in the THIRD STEP) is activated upon detecting the volumetric controller, which manages the signal through the dedicated software, giving rise to the reproduction of the programmed welcome segment, in order to then perform the action (associated with another specific programmed sequence) of exposing all comforts of the suite.

[0064] While it explains the operation of all elements of the suite (through a reproduction of the specific programmed segment), the related domotic peripherals will be enabled upon its command (tv, lights, curtains, taps...) through programmed commands sent with communication protocols (USB, serial port, parallel port, wifi, Bluetooth).

[0065] After having performed programmed sequence and related output commands, the interface will be put in IDLE status, waiting for voice or gesture commands from a user.

[0066] Every time, the commands will be performed and the butler will communicate the execution of commands verbally (replying with reference to requested information), physically (helping itself with gestures, particularly useful to be exploited by deaf users) and with the output towards external peripherals.

[0067] As regards the process which allows the interface operation for controlling the external peripherals, the idea is talking with an image which reproduces a human being (from a cinematography of a real uman being and not generated in computer graphics or animated) to allow controlling external peripherals.

[0068] The steps of such process are:

providing a reproduction (IDLE) of the human being in a standby status

detecting the presence of one or more users through an input of data provided by the volumetric sensor

in case of negative input, going back in a loop to the reproduction of the IDLE status

in caso of positive input, reproducing PREIDLE and ACTION with a request to make the control request of the peripheral (with the chance of showing a list of possible choices)

performing a verbal input by the user

performing a semantic analysis of the sentence and a logic comparison with the chance of executing updated commands by the peripheral

if the execution is not possible, reproducing ACTION with the request of inputting again a command on the peripheral

if the execution is not, visually executing BRIDGE and ACTION in which the interface describes and confirms the execution of the current command and going back to the initial IDLE status

simultaneously, providing an output through a suitable communication protocol with the peripheral, in which the peripheral is actuated and updated in its status.

REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US7844467B1 **[0005]**
- US2011007079A1 **[0005]**

FREMGANGSMÅDE TIL FREMSTILLING AF ET AUDIOVISUELT INTERFACE

PATENTKRAV

1. Fremgangsmåde til fremstilling af et audiovisuelt interface som reproducerer et komplet interaktivt menneske som valgt subjekt (7), fremgangsmåden omfatter følgende trin:
- interaktiv analyse;
- billede erhvervelse; og
- reproduktion, gennem et optisk eller fjernsyns system med stemme, bevægelser, logisk interaktions kapaciteter,

hvori trinnet med interaktiv analyse omfatter følgende undertrin: brug af et optisk apparat som reproducerer med tredimensional effekt mennesket, udstyret med en sensor/controller (3) i stand til udføre en stemme genkendelse, i stand til at fortolke menneskets sprog, fortolkningen af menneskets sprog gør det muligt at identificere informationens semantik; udløsende den passende hændelse afhængig af den modtagne stemme kommando, kommunikerende med en computer gennem et interface for at forbinde eksterne anordninger; opfattende tilstedeværelsen, fraværelse og relateret status ændring, og en serie af bevægelser af en seer/bruger (1), og

hvori det valgte subjekt (7) reciterer alle tilvejebragte segmenter, nemlig ACTION, og en serie af nonverbale handlinger, nemlig IDLE, PRE-IDLE og BRIDGE, som forsyner interfacet med en nærværs effekt på grund af simuleringen af menneskets opførsel under dialogen, de visuelle sekvenser som optages er segmenterede i henhold til en interaktivitets logik for at tilvejebringe en følelse af at tale med et menneske, de logiske sekvenser bliver delt i fire hoved kategorier: 1) IDLE sekvens (ligegyldigheds status i hvilken interfacet "lever" ventende på en status ændrings kommando); 2) PRE-IDLE sekvens (en adfærd som forekommer inden handlingen); 3) BRIDGE sekvens (en billede overgang som hjælper med at give visuel kontinuitet til interfacet mellem IDLE/PRE-IDLE statussen); og 4) ACTION sekvens (i hvilken interfacet udfører handlingen eller får handlingen udført),

hvori:

- trinnet for billede erhvervelse sørger for brugen af en kinematografisk teknik omfattende følgende undertrin: optage subjektet (7) med en high-definition digital optage maskine (5) med skanning sammenflettet i en vertikal, i stedet for horisontal, position, i den digitale optage maskine (5) oscillerende hældning og skiftende optik bliver monteret for at opnå dybdeskarphed i

henhold til Scheimpflug reglen eller vilkår, som fastslår, for et optisk system, at det fokale plan genereret af objektivet og det fokale plan af subjektet (7) mødes på den samme lige linje; fremskaffelse af billeder af subjektet (7) med udbredt og ensartet belysning af forsiden for at opnå en generel opacitet af subjektet (7) og en belysning af gode præcise sidesnit og baggrundslys for at tegne subjektet (7) i dets hele omkreds for at forøge den tredimensionelle effekt og have nøjagtig definerede kanter løsrevet i et tomt område, subjektet (7) som er totalt uden lys i området omgivende dets kant, herved opnås "kantskærings" effekten som er nødvendig for at reproducere billederne;

- fremgangsmåden omfatter endvidere undertrinene til fremvisning af billedet fra en LED monitor matrix (26) i en omvendt vertikal position i et plan fremstillet af gennemsigtigt polycarbonat (28) placeret i 45° i forhold til matrixen (26);

- trinnet til reproduktion omfatter undertrinene til: tilvejebringelse af et optisk reproduktions system tilpasset, gennem refleksion af billederne genererede af en LED monitor matrix (26) eller anden FULL HD kilde, for at få den optiske illusion af den totale suspension af et billede i et tomt område; reflekterende (28) det reproducerede billede på en plade fremstillet af gennemsigtigt polycarbonat; tilvejebringelse af et endeligt billede opfattet af seeren (1) på et vertikalt plan (30) bagved polycarbonat pladen, helt løsrevet fra en baggrund (32), med hvilket en parallakse effekt (34) opstår; og

- fremgangsmåden omfatter trinnet til aktivering af interfacet til kontrol af eksterne perifere enheder udenfor det optiske reproduktions system, nemlig domotiske perifere enheder, DMX styrede lys systemer til shows og events, en video projektion af et audiovisuelt bidrag på en ekstern storskærm eller enhver andet elektrisk eller elektronisk apparat udstyret med standard kommunikations protokoller, omfattende følgende trin:

tilvejebringelse af en reproduktion (IDLE) af en person i en standby status;

påvisning af tilstedeværelsen af en eller flere brugere gennem et input af data tilvejebragt af den volumetriske sensor;

i tilfælde af negativt input, nemlig hvor ingen brugere er til stede, går den tilbage i en sløjfe til reproduktionen af IDLE statussen;

i tilfælde af positivt input, nemlig hvor brugere er til stede, reproduceres PRE-IDLE sekvensen og ACTION sekvensen med en anmodning om at lave en kontrol anmodning til den perifere enhed, muligvis visende en liste over mulige valg;

gennemførelse af et verbalt input af brugeren;

gennemførelse af en semantisk analyse af sætningen og en logisk sammenligning, muligvis udførelse af opdaterede kommandoer af den perifere enhed;

hvis udførelsen ikke er mulig, reproducere ACTION sekvensen med en anmodning om indtastning igen af en kommando på den perifere enhed;

hvis udførelsen er mulig, visuel udførelse af BRIDGE sekvensen og ACTION sekvensen i hvilken interfacet beskrives og bekræfter udførelsen af den aktuelle kommando og går tilbage til den oprindelige IDLE status;

samtidigt, tilvejebringe et output gennem en passende kommunikations protokol med den perifere enhed, i hvilken den perifere enhed er aktiveret og opdateret i dens status.

2. Fremgangsmåde i henhold til patentkrav 1, tilpasset til at styre kommunikations protokoller, for eksempel WiFi og Bluetooth, på forskellige typer af elektriske og elektroniske apparater.

3. Fremgangsmåde i henhold til patentkrav 1, omfattende fremvisning af den menneskelige tilstedeværelse gennem et high-definition HD televisions signal.

4. Fremgangsmåde i henhold til patentkrav 1, omfattende fremvisning af den menneskelige tilstedeværelse gennem ethvert billede reproduktions system.

DRAWINGS

FIG. 1

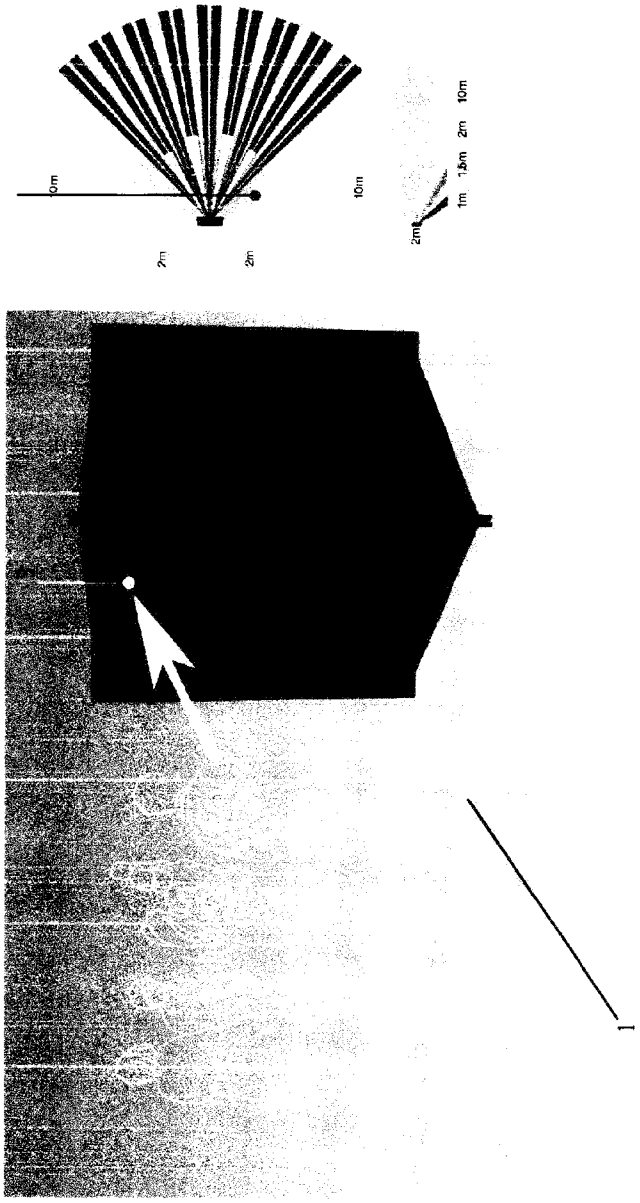


FIG. 2

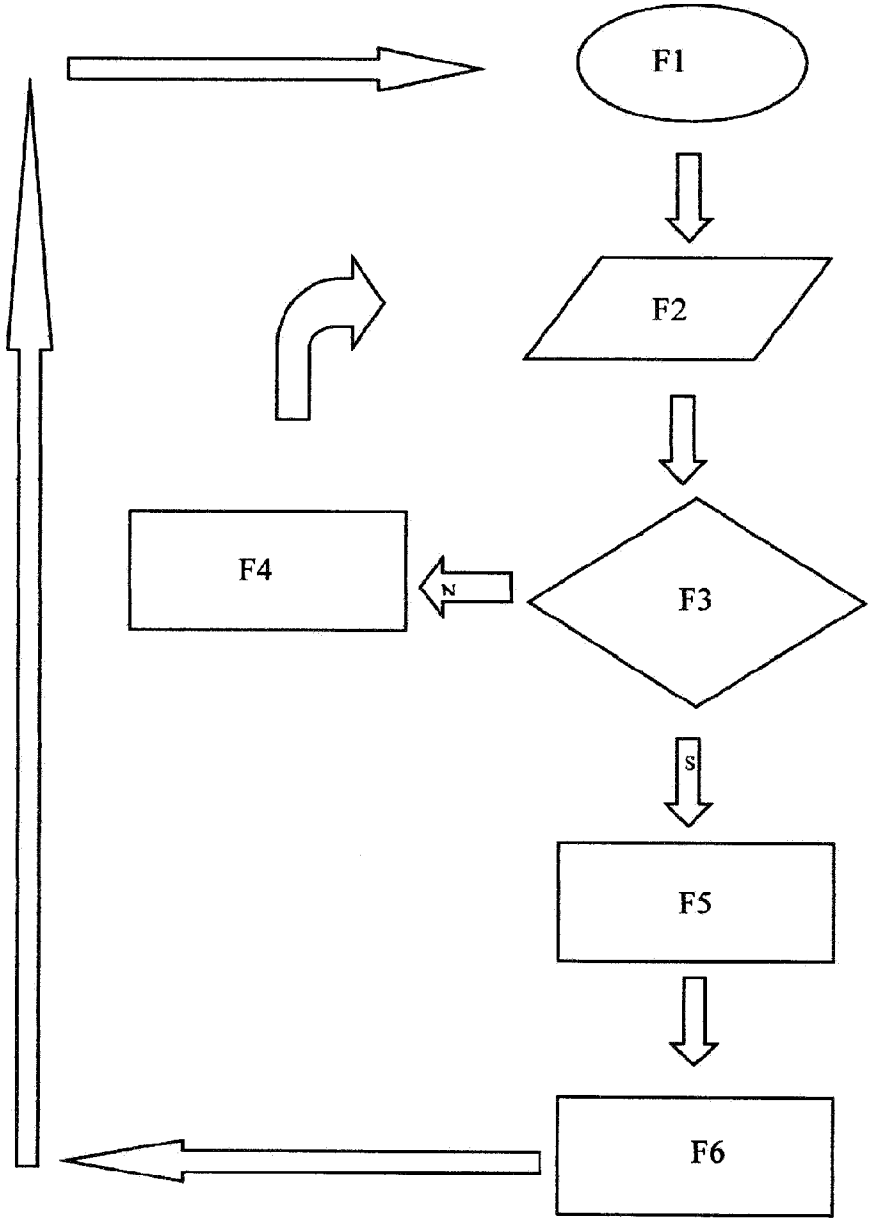


FIG. 3

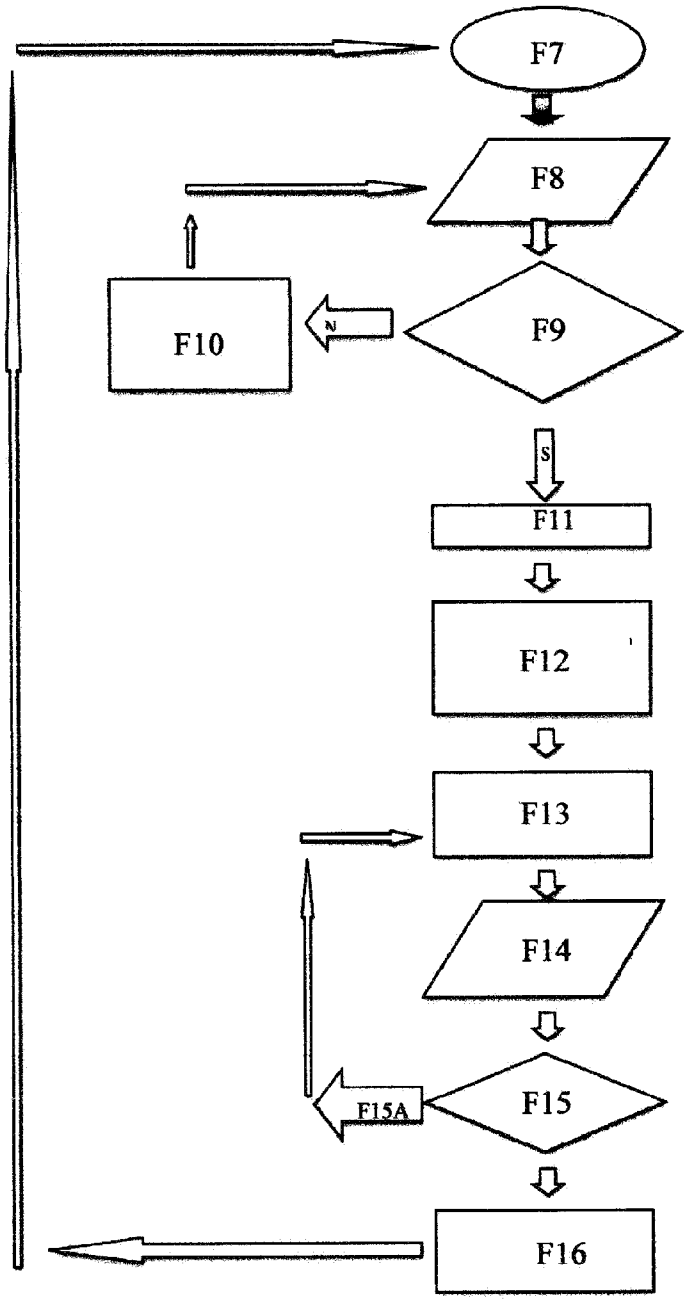
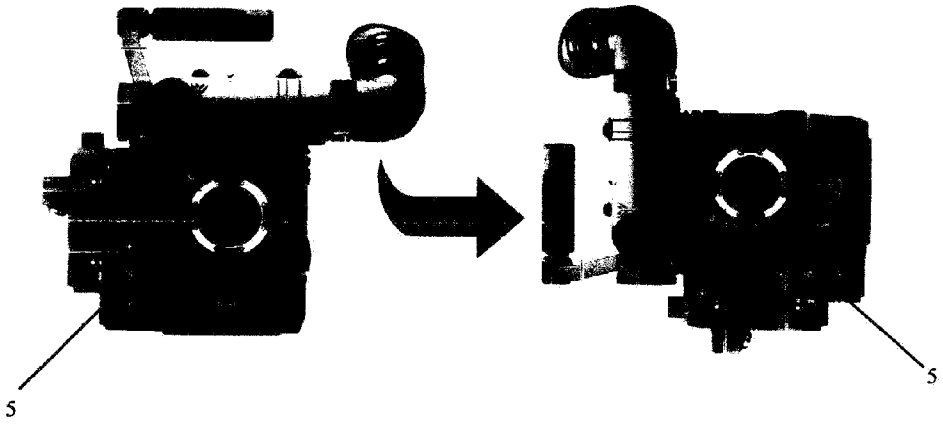
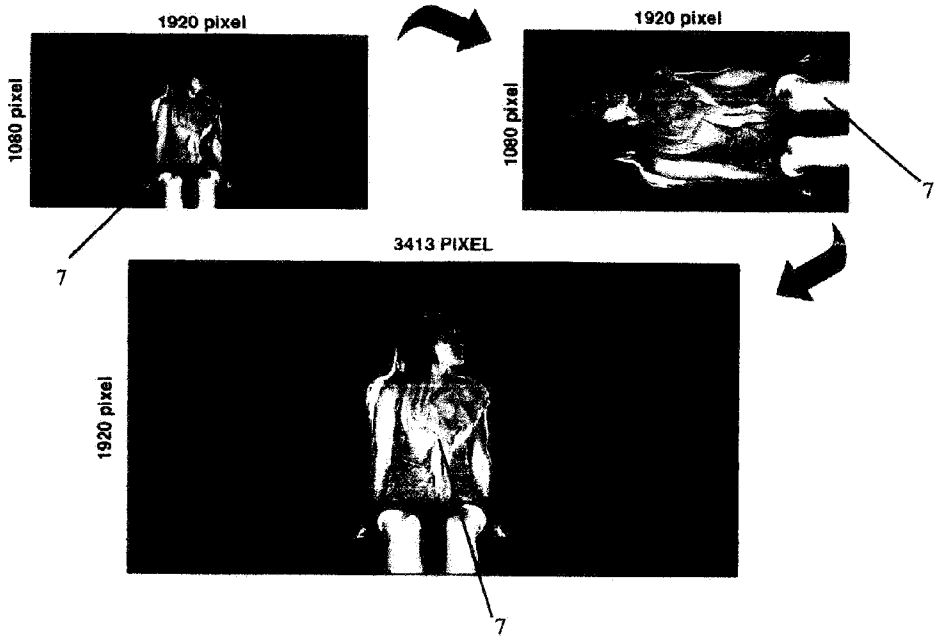


FIG. 4


5
5


1920 pixel
1080 pixel
1920 pixel
1080 pixel
7
7
3413 PIXEL
1920 pixel
7

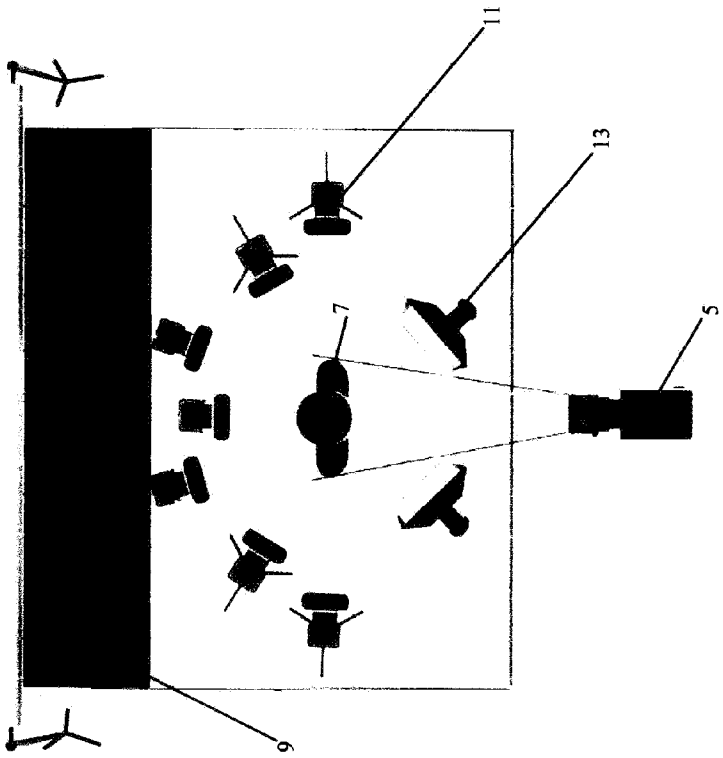


FIG. 5

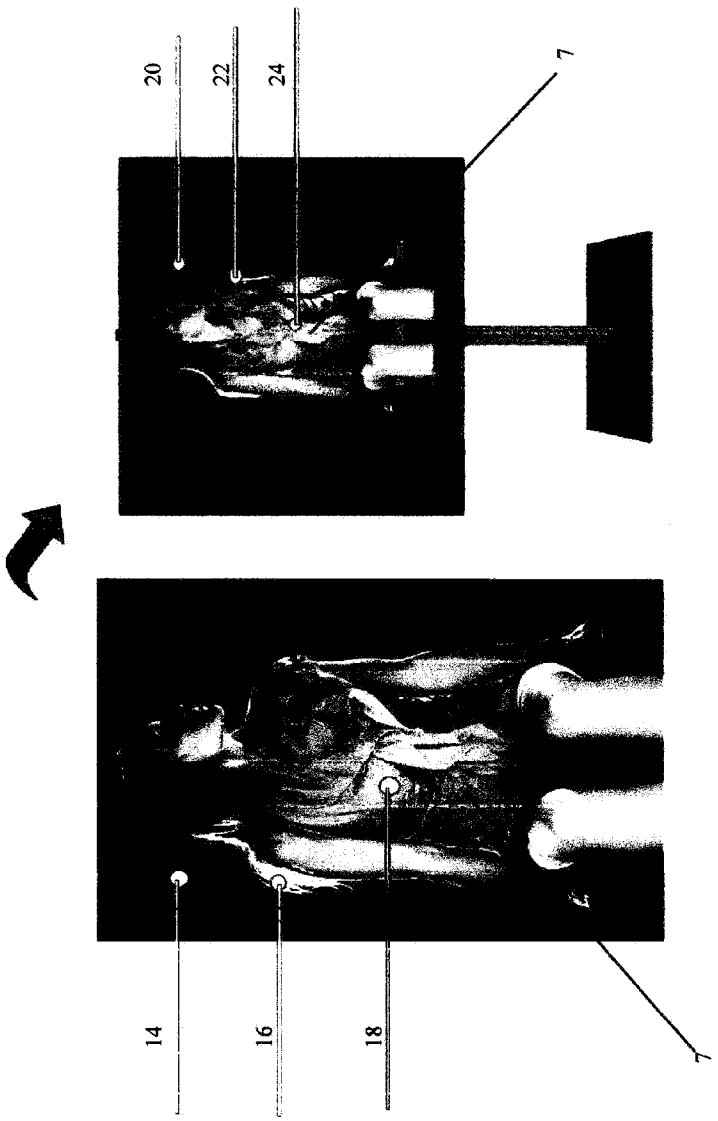


FIG. 6

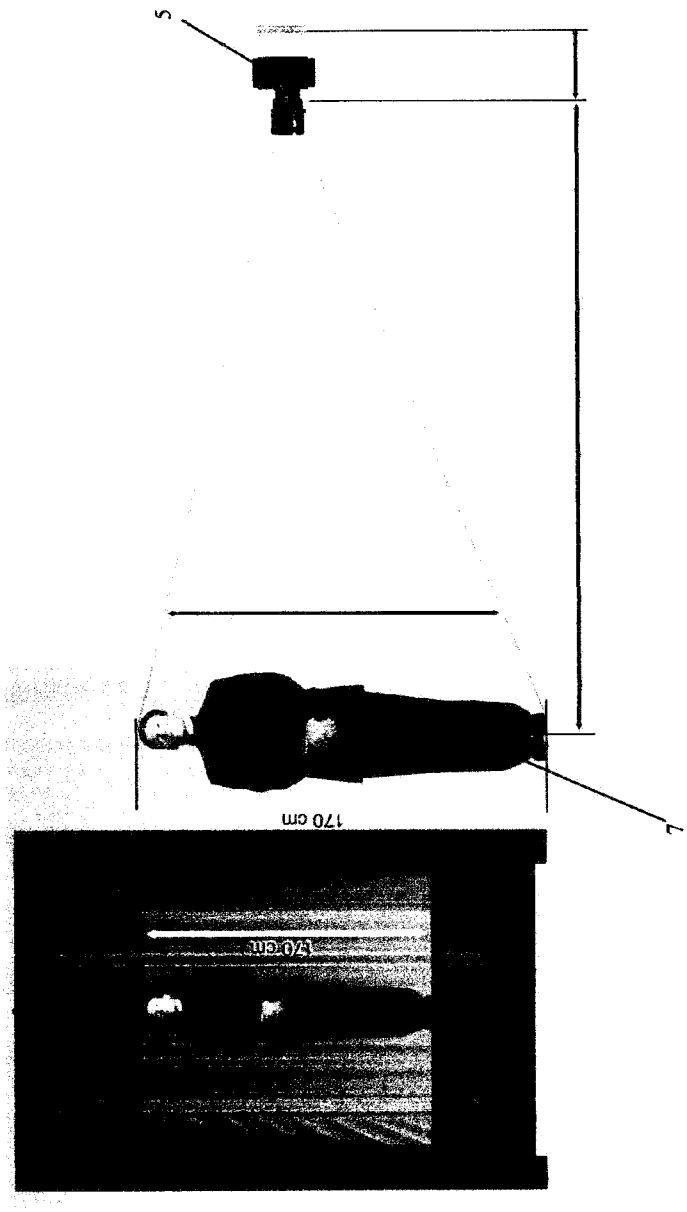


FIG. 7

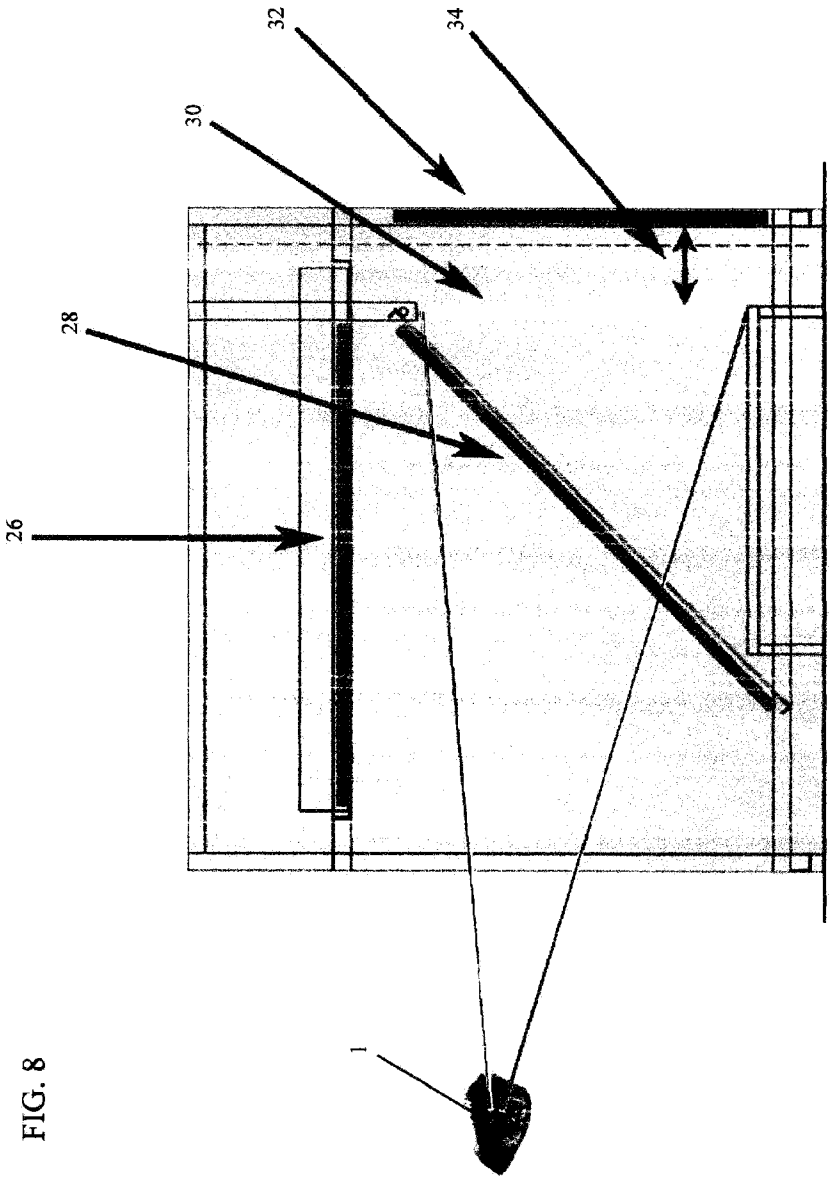


FIG. 8