IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF ILLINOIS
EASTERN DIVISION

GESTURE TECHNOLOGY PARTNERS, LLC, Plaintiff, v. MOTOROLA MOBILITY LLC, Defendants.	CIVIL ACTION NO. ORIGINAL COMPLAINT FOR PATENT INFRINGEMENT **JURY TRIAL DEMANDED**

ORIGINAL COMPLAINT FOR PATENT INFRINGEMENT

Plaintiff Gesture Technology Partners, LLC ("GTP" or "Plaintiff") files this original complaint against Motorola Mobility LLC ("Motorola" or "Defendant") alleging, based on its own knowledge as to itself and its own actions, and based on information and belief as to all other matters, as follows:

PARTIES

1. Gesture Technology Partners, LLC is a limited liability company filed under the laws of the State of Ohio, with its principal place of business at 2815 Joelle Drive, Toledo, OH 43617.

2. Defendant Motorola Mobility LLC is a Delaware limited liability company, with its principal place of business at 222 W. Merchandise Mart Plaza, Suite 1800, Chicago, Illinois 60654.

3. Motorola may be served through its registered agent, The Corporation Trust Company at Corporation Trust Center, 1209 Orange St., Wilmington, Delaware, 19801.

4. Motorola, an indirect subsidiary of Lenovo Group Limited, is involved in the development and sale of hardware and software relating to mobile products, such as smartphones.

5. Motorola designs, manufactures, makes, uses, imports into the United States, sells, and/or offers for sale in the United States smartphones. Motorola's smartphones are marketed, used, offered for sale, and/or sold throughout the United States, including within this district.

JURISDICTION AND VENUE

6. GTP repeats and re-alleges the allegations in Paragraphs 1-5 as though fully set forth in their entirety.

7. This is an action for infringement of United States patents arising under 35 U.S.C. §§ 271, 281, and 284–85, among others. This Court has subject matter jurisdiction of the action under 28 U.S.C. § 1331 and § 1338(a).

8. Venue is proper in this district pursuant to 28 U.S.C. §§ 1400(b) and 1391(c).

9. This Court has personal jurisdiction over Motorola pursuant to due process because, *inter alia*, (i) Motorola has done and continues to do business in Illinois; (ii) Motorola has committed and continues to commit acts of patent infringement in the State of Illinois, including making, using, offering to sell, and/or selling accused products in Illinois, and/or importing accused products into Illinois, including by Internet sales and/or sales via retail and wholesale stores, inducing others to commit acts of patent infringement in Illinois, and/or committing at least a portion of any other infringements alleged herein in Illinois; and (iii) Motorola regularly places its products within the stream of commerce—directly, through subsidiaries, or through third parties—with the expectation and knowledge that such products will be shipped to, sold, or used in Illinois and elsewhere in the United States. Thus, Motorola has established minimum contacts within Illinois and purposefully availed itself of the benefits of Illinois, and the exercise of personal jurisdiction over Motorola would not offend traditional notions of fair play and substantial justice.

10. Venue is proper in this district as to Motorola under 28 U.S.C. § 1400(b). Venue is further proper as to Motorola because it has committed and continues to commit acts of patent infringement in this district, including making, using, offering to sell, and/or selling accused products in this district, and/or importing accused products into this district, including by Internet sales and/or sales via retail and wholesale stores, and inducing others to commit acts of patent infringement in this district.

11. Furthermore, Motorola has a regular and established place of business in this district, including at least at 222 W. Merchandise Mart Plaza, Suite 1800, Chicago, Illinois, 60654.

THE TECHNOLOGY

12. GTP repeats and re-alleges the allegations in Paragraphs 1-11 as though fully set forth in their entirety.

13. GTP was founded in 2013 by Dr. Timothy Pryor, the sole inventor of the four Asserted Patents. He currently resides in Toledo, Ohio. Dr. Pryor received a B.S. in Engineering Physics from Johns Hopkins University in 1962, where he was also a member of the Army Reserve Officer in Training (ROTC) program. Upon graduation, he was commissioned as a Second Lieutenant in the United States Army. Dr. Pryor continued his education, obtaining an M.S. in Physics from the University of Illinois (1964) and a Ph.D. in Mechanical Engineering from the University of Windsor (1972).

14. Dr. Pryor rose to the rank of Captain in the U.S. Army before his honorable discharge in 1967. Dr. Pryor served at the U.S. Army Aberdeen Proving Ground and in Italy, commanding missile teams supporting the Italian armed forces on a NATO anti-aircraft missile site, charged with guarding nuclear warheads and providing technical assistance to NATO.

15. Dr. Pryor is a named inventor on over 200 patents and patent applications. For the past four decades, he has been a pioneer in laser sensing technology, motion sensing technology, machine vision technology and camera-based interactive technology.

16. Since the 1970's, Dr. Pryor has founded and led three other companies: two small operating companies in the automotive parts inspection and robotics businesses, one company that developed new forms of vehicle instrument panel controls, and co-founded another company that utilized camera-based sensors for physical therapy. Dr. Pryor is responsible for a significant amount of the research and development for the technologies at these companies.

17. The patents-in-suit, U.S. Patent Nos. 8,194,924 (the "'924 patent"), 7,933,431 (the "'431 patent"), 8,878,949 (the "'949 patent"), and 8,553,079 (the "'079 patent") (collectively, the "Asserted Patents"), are generally directed to innovations in using mobile phone cameras to assist a user to interact with their smartphone, including, for example, but not limited to unlocking their phone, taking and using photos or videos, and providing other functions.

18. Dr. Pryor conceived of the inventions embodied in the Asserted Patents in the mid- to late-1990s, when he was working on a variety of different projects related to imaging and computer control. Dr. Pryor describes the process as a brainstorm that led to several breakthrough moments, ultimately resulting in the Asserted Patents.

THE ACCUSED PRODUCTS

19. GTP repeats and re-alleges the allegations in Paragraphs 1-18 as though fully set forth in their entirety.

20. Motorola infringed the asserted patents by making, using, selling, offering to sell, and importing its smartphones and tablets including, but not limited to the Motorola One Fusion+, the Motorola One 5G, the Motorola One Zoom, the Motorola One Action, the Motorola One

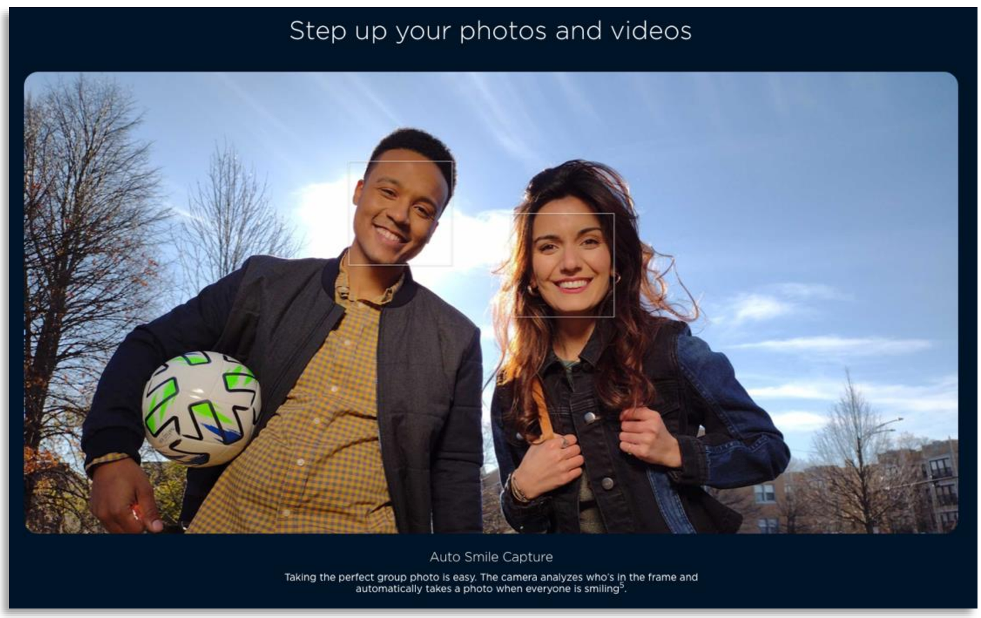
Hyper, the Motorola G Stylus, the Motorola G Power, the Motorola G Fast, and the Motorola E (collectively the "Accused Products").

EXAMPLES OF MOTOROLA'S MARKETING OF THE FEATURES

21. The Accused Products have features including, but not limited to, at least the following: Auto Smile Capture, Shot Optimization, Smart Composition, Portrait Mode, Cutout, Live Filter, Best Shot, Google Lens Integration, AR Sticker, Electronic Image Stabilization, Face Beauty, Attentive Display, Group Selfie, Gesture Selfie, and facial recognition (the "Features").

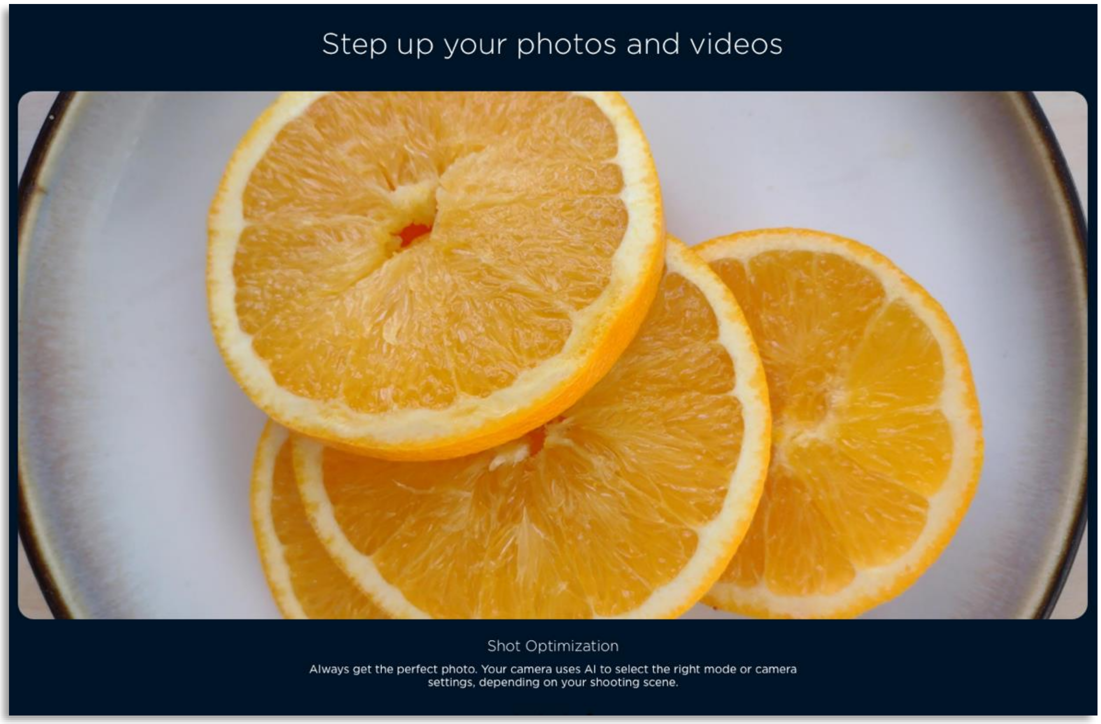
22. The Features drive the popularity and sales of the Accused Products.

23. For example, Motorola has marketed the Accused Products using Auto Smile Capture to automatically take photos when everyone is the frame is smiling, as described in the following screenshot from Motorola's website:[1]



[1] Motorola Electronics Co., Ltd., *moto g Power*, available at https://www.motorola.com/us/smartphones-moto-g-power-gen-2/p (last accessed February 7, 2022).

24. Motorola has marketed its Accused Products using Shot Optimization to obtain the best quality photos, as described in the following screenshot from Motorola's website:[2]



25. Motorola has marketed its Accused Products using functions in taking photos and videos, such as the gesture control in its front-facing camera, as described in the following screenshot from Motorola's website:[3]

[2] Motorola Electronics Co., Ltd., *moto g Power*, available at https://www.motorola.com/us/smartphones-moto-g-power-gen-2/p (last accessed February 7, 2022).
[3] Motorola Electronics Co., Ltd., *Taking selfies - motorola one 5G*, available at https://support.motorola.com/us/en/Solution/MS153524 (last accessed February 7, 2022).

Show palm to take selfie

You can have the camera start a countdown timer to take a selfie by raising your palm.

To enable this gesture:

1. Switch to the selfie camera by touching
2. On the viewfinder, touch **Settings**
3. Turn **Gesture Selfie** on

To use it:

1. Open the selfie camera.
2. Without covering your face in the viewfinder, show your open hand (palm) to the camera. Three dots at the top of the viewfinder indicate a countdown timer.
3. Smile and wait for the timer.

26. Motorola has marketed its Accused Products using attentive display to keep the Accused Product's screen on when looking at it, as described in the following screenshot from Motorola's website:[4]

In certain lighting conditions, your phone may have difficulty detecting when you've looked away.
To turn on Attentive Display:

1. Go to **Settings** > **Display** > **Advanced**.
   Or, open the **Moto** app and touch > **Display**.
2. Touch **Attentive Display** and turn the feature on or off.

COUNT I

INFRINGEMENT OF U.S. PATENT NO. 8,194,924

27. GTP repeats and re-alleges the allegations in Paragraphs 1-26 as though fully set forth in their entirety.

28. GTP owns all substantial rights, interest, and title in and to the '924 patent, including the sole and exclusive right to prosecute this action and enforce the '924 patent against infringers, and to collect damages for all relevant times. The United States Patent and Trademark

[4] Motorola Electronics Co., Ltd., *Display settings - motorola one 5G ACE*, available at https://support.motorola.com/us/en/Solution/MS156753 (last accessed February 7, 2022).

Office duly issued the ’924 patent on June 5, 2012. A copy of the ’924 patent is attached as Exhibit A.

29. The ’924 patent is titled “Camera Based Sensing in Handheld, Mobile, Gaming or Other Devices.” The ’924 patent describes using a camera output such that the handheld device’s computer performs a control function on the device, such as acquiring or taking images, reading things, determining data, transmitting data, printing data, and actuating a vehicle or function.

30. The claims of the ’924 patent are not directed to an abstract idea.

31. Motorola has directly infringed (literally or under the doctrine of equivalents) at least Claim 1 of the ’924 patent.

32. Motorola has infringed the ’924 patent by making, using, selling, offering for sale, and importing the Accused Products.

33. The Accused Products are handheld devices with a housing and a computer, including but not limited to one or more System-on-Chips.

34. The Accused Products have at least one first camera oriented to view a user of the Accused Product. The first camera has an output when used.

35. The Accused Products have at least one second camera oriented to view an object other than the user. The second camera has an output when used.

36. The first and second cameras of the Accused Products have non-overlapping fields of view.

37. The computer of the Accused Products is adapted to perform a control function, such as the control functions associated with the Features, based on an output of either the first camera or the second camera.

38. Plaintiff has been damaged as a result of the infringing conduct by Motorola alleged above. Thus, Motorola is liable to Plaintiff in an amount that compensates it for such infringements, which by law cannot be less than a reasonable royalty, together with interest and costs as fixed by this Court under 35 U.S.C. § 284.

39. Plaintiff has satisfied all statutory obligations required to collect pre-filing damages for the full period allowed by law for infringement of the ’924 patent.

40. Plaintiff has not offered for sale nor sold any product implicated by 35 U.S.C. § 287 with respect to the ’924 patent.

41. Motorola had knowledge of the ’924 patent at least as of the filing of this Complaint.

42. Motorola has also indirectly infringed one or more claims of the ’924 patent by inducing others to directly infringe the ’924 patent. Motorola has induced end-users and other third-parties to directly infringe (literally or under the doctrine of equivalents) the ’924 patent by using the Accused Products. Motorola took active steps, directly or through contractual relationships with others, with the specific intent to cause them to use the Accused Products in a manner that infringes one or more claims of the ’924 patent, including, for example, Claim 1 of the ’924 patent. Such steps by Motorola included, among other things, advising or directing end-users and other third-parties to use the Accused Features in the Accused Products in an infringing manner; advertising and promoting the use of the Accused Products in an infringing manner; or distributing instructions that guide end-users and other third-parties to use the Accused Products in an infringing manner. Motorola performed these steps, which constitute induced infringement with the knowledge of the ’924 patent and with the knowledge that the induced acts constitute infringement. Motorola was aware that the normal and customary use of the Accused Products by

others would infringe the ’924 patent. Motorola’s direct infringement of the ’924 patent was willful, intentional, deliberate, or in conscious disregard of Plaintiff’s rights under the patent.

COUNT II

INFRINGEMENT OF U.S. PATENT NO. 7,933,431

43. GTP repeats and re-alleges the allegations in Paragraphs 1-42 as though fully set forth in their entirety.

44. GTP owns all substantial rights, interest, and title in and to the ’431 patent, including the sole and exclusive right to prosecute this action and enforce the ’431 patent against infringers, and to collect damages for all relevant times. The United States Patent and Trademark Office duly issued the ’431 patent on April 26, 2011. A copy of the ’431 patent is attached as Exhibit B.

45. The ’431 patent is titled “Camera Based Sensing in Handheld, Mobile, Gaming, or Other Devices.” The ’431 patent describes a method for a user to control a handheld device using gestures that are observed by a sensor on the handheld device.

46. Motorola has directly infringed (literally or under the doctrine of equivalents) at least Claim 7 of the ’431 patent. Motorola has infringed the ’431 patent by making, using, selling, offering to sell, and importing the Accused Products.

47. The claims of the ’431 patent are not directed to an abstract idea.

48. The Accused Products are handheld computers.

49. The Accused Products have a housing.

50. The Accused Products have one or more cameras associated with their housing. The one or more cameras obtain images of objects using reflected light from the objects.

51. A computer, including but not limited to at least one System on Chip, resides within the housing of the Accused Products. The computer analyzes images obtained by the one or more images to determine information about a position or movement of the object.

52. The Accused Products use information about the object to control a function of the Accused Products, such as the functions associated with the Features.

53. Plaintiff has been damaged as a result of the infringing conduct by Motorola alleged above. Thus, Motorola is liable to Plaintiff in an amount that adequately compensates it for such infringements, which by law cannot be less than a reasonable royalty, together with interest and costs as fixed by this Court under 35 U.S.C. § 284.

54. Plaintiff has satisfied all statutory obligations required to collect pre-filing damages for the full period allowed by law for infringement of the ’431 Patent.

55. Plaintiff has not offered for sale nor sold any product implicated by 35 U.S.C. § 287 with respect to the ’431 patent.

56. Motorola had knowledge of the ’431 patent at least as of the filing of this Complaint.

57. Motorola has also indirectly infringed one or more claims of the ’431 patent by inducing others to directly infringe the ’431 patent. Motorola has induced end-users and other third-parties to directly infringe (literally or under the doctrine of equivalents) the ’431 patent by using the Accused Products. Motorola took active steps, directly or through contractual relationships with others, with the specific intent to cause them to use the Accused Products in a manner that infringes one or more claims of the ’431 patent, including, for example, Claim 7 of the ’431 patent. Such steps by Motorola included, among other things, advising or directing end-users and other third-parties to use the Accused Products in an infringing manner; advertising and

promoting the use of the Accused Products in an infringing manner; or distributing instructions that guide end-users and other third-parties to use the Accused Products in an infringing manner. Motorola performed these steps, which constitute induced infringement with the knowledge of the ’431 patent and with the knowledge that the induced acts constitute infringement. Motorola was aware that the normal and customary use of the Accused Products by others would infringe the ’431 patent. Motorola’s direct and indirect infringement of the ’431 patent was willful, intentional, deliberate, or in conscious disregard of Plaintiff’s rights under the patent.

COUNT III

INFRINGEMENT OF U.S. PATENT NO. 8,878,949

58. GTP repeats and re-alleges the allegations in Paragraphs 1-57 as though fully set forth in their entirety.

59. GTP owns all substantial rights, interest, and title in and to the ’949 patent, including the sole and exclusive right to prosecute this action and enforce the ’949 patent against infringers, and to collect damages for all relevant times. The United States Patent and Trademark Office duly issued the ’949 patent on November 4, 2014. A copy of the ’949 patent is attached as Exhibit C.

60. The ’949 Patent is titled “Camera Based Interaction and Instruction.” The ’949 patent describes a device that allows a user to control the device using gestures registered by the front-facing camera and an electro-optical sensor.

61. The claims of the ’949 patent are not directed to an abstract idea.

62. Motorola has directly infringed (literally or under the doctrine of equivalents) at least Claim 1 of the ’949 patent. Motorola infringed the ’949 patent by making, using, selling, offering for sale, and importing the Accused Products.

63. The Accused Products are portable devices.

64. The Accused Products have a housing. The housing has a forward-facing portion that includes an electro-optical sensor that has a field of view and a digital camera.

65. Within the housing is a processing unit including, but not limited to, at least one System on Chip. The processing unit is coupled to the electro-optical sensor.

66. The processing unit in the Accused Products has been programmed to determine if a gesture has been performed in the electro-optical sensors field of view based on an output from the electro-optical sensor.

67. The processing unit of the Accused Products controls the digital camera in response to the gesture performed. Such gestures are used by the Features.

68. Plaintiff has been damaged as a result of the infringing conduct by Motorola alleged above. Thus, Motorola is liable to Plaintiff in an amount that adequately compensates it for such infringements, which by law cannot be less than a reasonable royalty, together with interest and costs as fixed by this Court under 35 U.S.C. § 284.

69. Plaintiff has satisfied all statutory obligations required to collect pre-filing damages for the full period allowed by law for infringement of the ’949 Patent.

70. Plaintiff has not offered for sale nor sold any product implicated by 35 U.S.C. § 287 with respect to the ’949 patent.

71. Motorola had knowledge of the ’949 patent at least as of the filing of this Complaint.

72. Motorola has also indirectly infringed one or more claims of the ’949 patent by inducing others to directly infringe the ’949 patent. Motorola has induced end-users and other third-parties to directly infringe (literally or under the doctrine of equivalents) the ’949 patent by

using the Accused Products. Motorola took active steps, directly or through contractual relationships with others, with the specific intent to cause them to use the Accused Products in a manner that infringes one or more claims of the ’949 patent, including, for example, Claim 1 of the ’949 patent. Such steps by Motorola included, among other things, advising or directing end-users and other third-parties to use the Accused Products in an infringing manner; advertising and promoting the use of the Accused Products in an infringing manner; or distributing instructions that guide end-users and other third-parties to use the Accused Products in an infringing manner. Motorola performed these steps, which constitute induced infringement with the knowledge of the ’949 patent and with the knowledge that the induced acts constitute infringement. Motorola was aware that the normal and customary use of the Accused Products by others would infringe the ’949 patent. Motorola’s direct and indirect infringement of the ’949 patent was willful, intentional, deliberate, or in conscious disregard of Plaintiff’s rights under the patent.

COUNT IV

INFRINGEMENT OF U.S. PATENT NO. 8,553,079

73. GTP repeats and re-alleges the allegations in Paragraphs 1-72 as though fully set forth in their entirety.

74. GTP owns all substantial rights, interest, and title in and to the ’079 patent, including the sole and exclusive right to prosecute this action and enforce the ’079 patent against infringers, and to collect damages for all relevant times. The United States Patent and Trademark Office duly issued the ’079 patent on October 8, 2013. A copy of the ’079 patent is attached as Exhibit D.

75. The ’079 patent is titled “More Useful Man Machine Interfaces and Applications.” The ’079 patent describes methods and apparatuses related to determining gestures illuminated by a light source of a computer by using a camera housed in the computer.

76. The claims of the ’079 patent are not directed to an abstract idea.

77. Motorola has directly infringed (literally or under the doctrine of equivalents) at least Claim 11 of the ’079 patent. Motorola has infringed the ’079 patent by making, using, selling, offering for sale, and importing the Accused Products.

78. The Accused Products are computer apparatuses.

79. The Accused Products contain a light source that will illuminate a human body part within a work volume.

80. The Accused Products have one or more cameras. The one or more cameras have a fixed relation to the light source. The one or more cameras of the Accused Products are oriented to observe gestures performed by a human body part.

81. The Accused Products have one or more processors including, but not limited to, one or more System on Chips, that have been programmed to determine a gesture performed based on output from the one or more cameras.

82. Plaintiff has been damaged as a result of the infringing conduct by Motorola alleged above. Thus, Motorola is liable to Plaintiff in an amount that adequately compensates it for such infringements, which by law cannot be less than a reasonable royalty, together with interest and costs as fixed by this Court under 35 U.S.C. § 284.

83. Plaintiff has satisfied all statutory obligations required to collect pre-filing damages for the full period allowed by law for infringement of the ’079 patent.

84. Plaintiff has not offered for sale nor sold any product implicated by 35 U.S.C. § 287 with respect to the ’079 patent.

85. Motorola had knowledge of the ’079 patent at least as of the filing of this Complaint.

86. Motorola has also indirectly infringed one or more claims of the ’079 patent by inducing others to directly infringe the ’079 patent. Motorola has induced end-users and other third-parties to directly infringe (literally or under the doctrine of equivalents) the ’079 patent by using the Accused Products. Motorola took active steps, directly or through contractual relationships with others, with the specific intent to cause them to use the Accused Products in a manner that infringes one or more claims of the ’079 patent, including, for example, Claim 11 of the ’079 patent. Such steps by Motorola included, among other things, advising or directing end-users and other third-parties to use the Accused Products in an infringing manner; advertising and promoting the use of the Accused Products in an infringing manner; or distributing instructions that guide end-users and other third-parties to use the Accused Products in an infringing manner. Motorola performed these steps, which constitute induced infringement with the knowledge of the ’079 patent and with the knowledge that the induced acts constitute infringement. Motorola was aware that the normal and customary use of the Accused Products by others would infringe the ’079 patent. Motorola’s inducement is ongoing.

87. Motorola’s direct and indirect infringement of the ’079 patent was willful, intentional, deliberate, or in conscious disregard of Plaintiff’s rights under the patent.

JURY DEMAND

Plaintiff hereby requests a trial by jury on all issues so triable by right.

PRAYER FOR RELIEF

GTP requests that the Court find in its favor and against Motorola, and that the Court grant GTP the following relief:

a. Judgment that one or more claims of the Asserted Patents have been infringed, either literally or under the doctrine of equivalents, by Motorola or all others acting in concert therewith;

b. Judgment that Motorola accounts for and pays to GTP all damages to and costs incurred by GTP because of Motorola's infringing activities and other conduct complained of herein;

c. Judgment that Motorola's infringements be found willful, and that the Court award treble damages for the period of such willful infringement pursuant to 35 U.S.C. § 284;

d. Pre-judgment and post-judgment interest on the damages caused by Motorola's infringing activities and other conduct complained of herein;

e. That this Court declare this an exceptional case and award GTP its reasonable attorneys' fees and costs in accordance with 35 U.S.C. § 285; and

f. All other and further relief as the Court may deem just and proper under the circumstances.

Dated: July 7, 2022

Respectfully submitted,

By: *<u>/s/ Brian E. Martin</u>*
Brian E. Martin
STAMOS & TRUCCO LLP
One East Wacker Drive, 3rd Floor
Chicago, Il 60601
Tel: 312-630-7979
bmartin@stamostrucco.com

Fred I. Williams
Texas State Bar No. 00794855
Michael Simons
Texas State Bar No. 24008042
WILLIAMS SIMONS & LANDIS PLLC
The Littlefield Building
601 Congress Ave., Suite 600
Austin, TX 78701
Tel: 512-543-1354
fwilliams@wsltrial.com
msimons@wsltrial.com

Todd E. Landis
State Bar No. 24030226
WILLIAMS SIMONS & LANDIS PLLC
2633 McKinney Ave., Suite 130 #366
Dallas, TX 75204
Tel: 512-543-1357
tlandis@wsltrial.com

John Wittenzellner
Pennsylvania State Bar No. 308996
WILLIAMS SIMONS & LANDIS PLLC
1735 Market Street, Suite A #453
Philadelphia, PA 19103
Tel: 512-543-1373
johnw@wsltrial.com

Attorneys for Plaintiff Gesture Technology Partners, LLC

CERTIFICATE OF SERVICE

The undersigned hereby certifies that on July 7, 2022, the undersigned caused a copy of the foregoing document to be served on Motorola through a process server.

/s/ Brian E. Martin
Brian E. Martin

EXHIBIT A

US008194924B2

(12) **United States Patent**
Pryor

(10) **Patent No.: US 8,194,924 B2**
(45) **Date of Patent: Jun. 5, 2012**

(54) **CAMERA BASED SENSING IN HANDHELD, MOBILE, GAMING OR OTHER DEVICES**

(76) Inventor: **Timothy R. Pryor**, Sylvania, OH (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **13/051,698**

(22) Filed: **Mar. 18, 2011**

(65) **Prior Publication Data**

US 2011/0170746 A1 Jul. 14, 2011

Related U.S. Application Data

(63) Continuation of application No. 12/834,281, filed on Jul. 12, 2010, now Pat. No. 7,933,431, which is a continuation of application No. 11/980,710, filed on Oct. 31, 2007, now Pat. No. 7,756,297, which is a continuation of application No. 10/893,534, filed on Jul. 19, 2004, now Pat. No. 7,401,783, which is a continuation of application No. 09/612,225, filed on Jul. 7, 2000, now Pat. No. 6,766,036.

(60) Provisional application No. 60/142,777, filed on Jul. 8, 1999.

(51) **Int. Cl.**
G06K 9/00 (2006.01)

(52) **U.S. Cl.** **382/103**; 382/154; 382/312

(58) **Field of Classification Search** 382/103, 382/154, 312
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,909,002 A		9/1975	Levy	
4,219,847 A		8/1980	Pinkney et al.	
4,339,798 A		7/1982	Hedges et al.	
4,631,676 A		12/1986	Pugh	
5,008,946 A		4/1991	Ando	
5,088,928 A		2/1992	Chan	
5,227,986 A		7/1993	Yokota et al.	
5,297,061 A		3/1994	Dementhon et al.	
5,388,059 A		2/1995	DeMenthon	
5,491,507 A	*	2/1996	Umezawa et al.	348/14.02
5,581,276 A		12/1996	Cipolla et al.	
5,594,469 A		1/1997	Freeman et al.	
5,616,078 A		4/1997	Oh	
5,624,117 A		4/1997	Ohkubo et al.	
5,781,647 A		7/1998	Fishbine et al.	
5,828,770 A	*	10/1998	Leis et al.	382/103
5,845,006 A	*	12/1998	Sumi et al.	382/154
5,853,327 A		12/1998	Gilboa	
5,878,174 A		3/1999	Stewart et al.	
5,926,168 A		7/1999	Fan	
5,940,126 A	*	8/1999	Kimura	348/294
6,204,852 B1	*	3/2001	Kumar et al.	345/419
6,342,917 B1		1/2002	Amenta	
6,373,472 B1		4/2002	Palalau et al.	
6,442,465 B2		8/2002	Breed et al.	
6,508,709 B1		1/2003	Karmarkar	
6,597,817 B1		7/2003	Silverbrook	
6,775,361 B1	*	8/2004	Arai et al.	379/93.17
6,788,336 B1	*	9/2004	Silverbrook	348/207.2
6,911,972 B2	*	6/2005	Brinjes	345/175
7,489,863 B2	*	2/2009	Lee	396/429

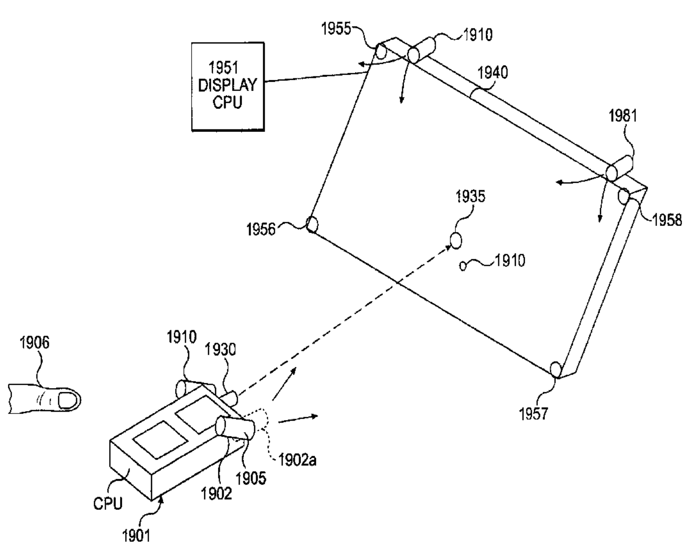
* cited by examiner

Primary Examiner — Tom Y Lu
(74) *Attorney, Agent, or Firm* — Warner Norcross & Judd LLP

(57) **ABSTRACT**

Method and apparatus are disclosed to enable rapid TV camera and computer based sensing in many practical applications, including, but not limited to, handheld devices, cars, and video games. Several unique forms of social video games are disclosed.

14 Claims, 23 Drawing Sheets



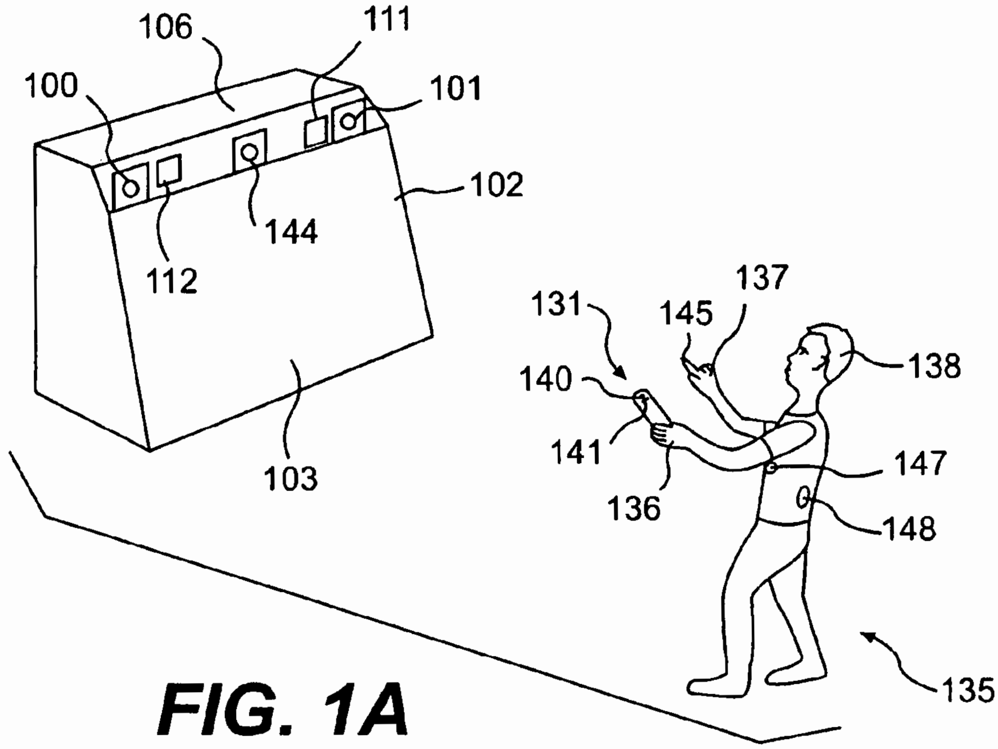


FIG. 1A

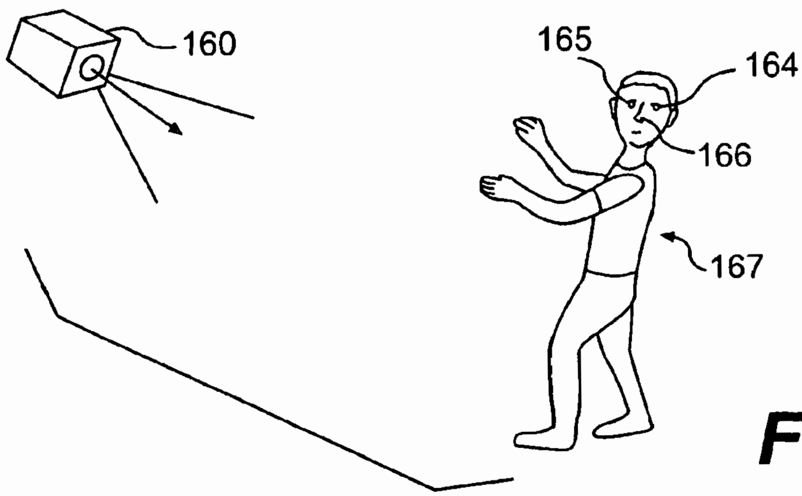


FIG. 1B

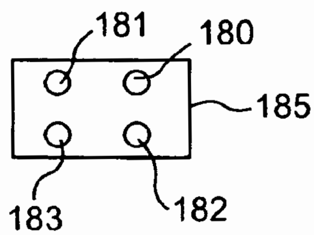


FIG. 1C

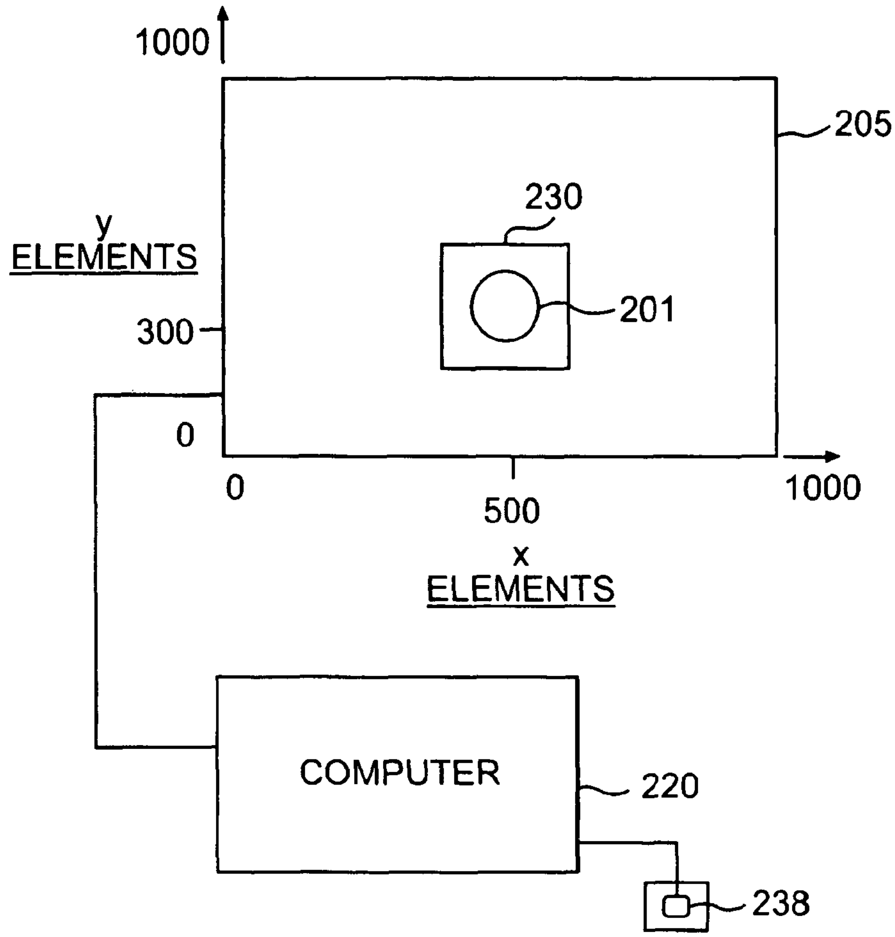

1000
205
230
y
ELEMENTS
201
300
0
0
500
1000
x
ELEMENTS
COMPUTER
220
238

FIG. 2A

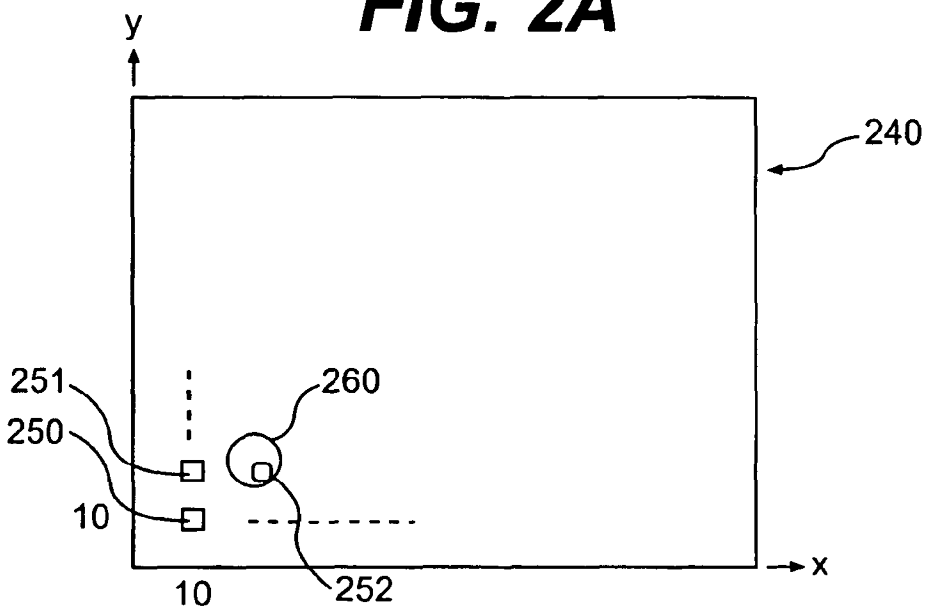

y
240
251
260
250
10
10
252
x

FIG. 2B

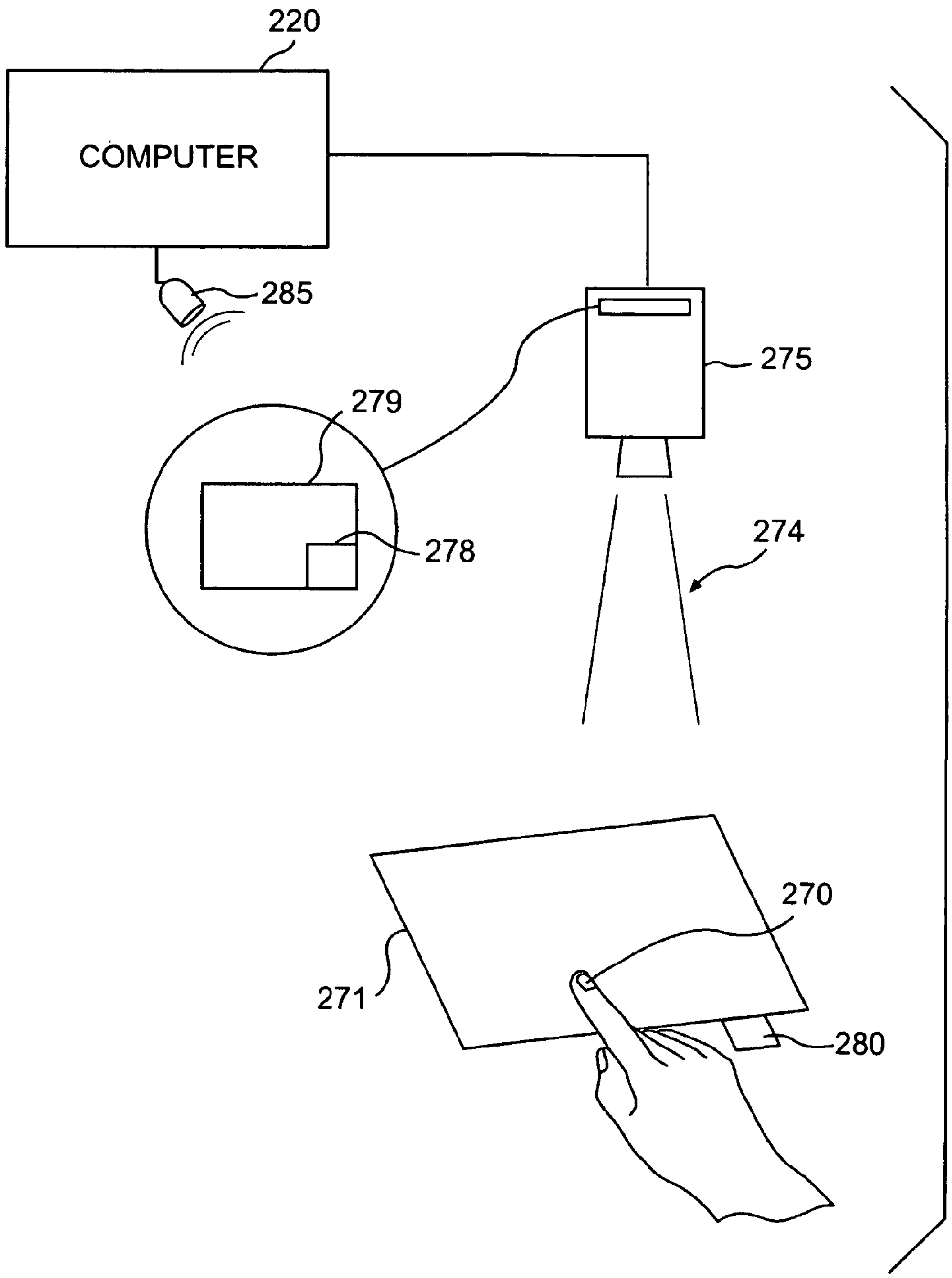


FIG. 2C

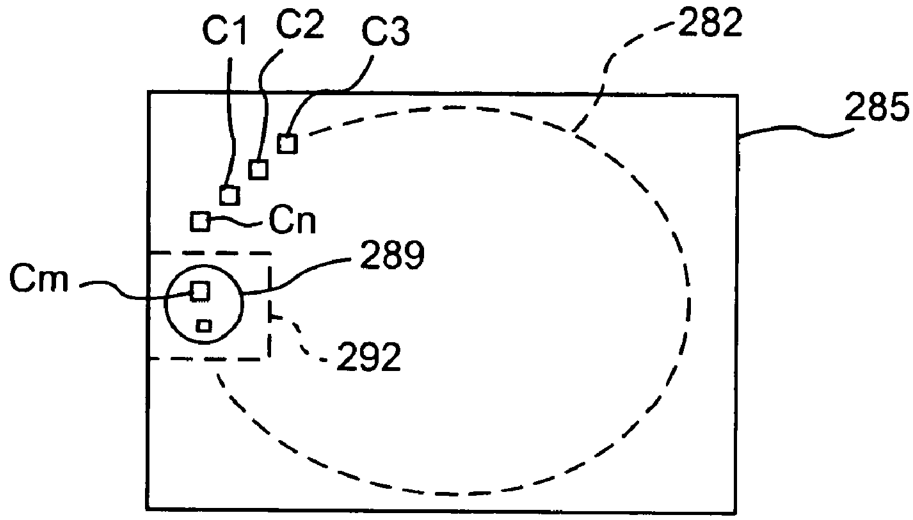


FIG. 2D

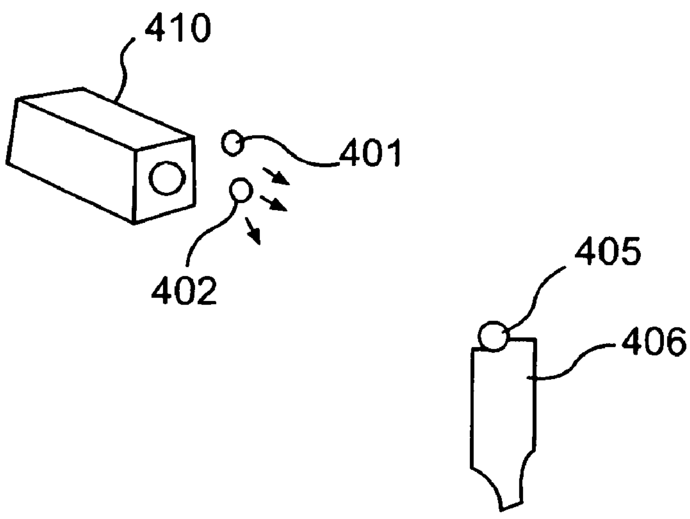


FIG. 4A

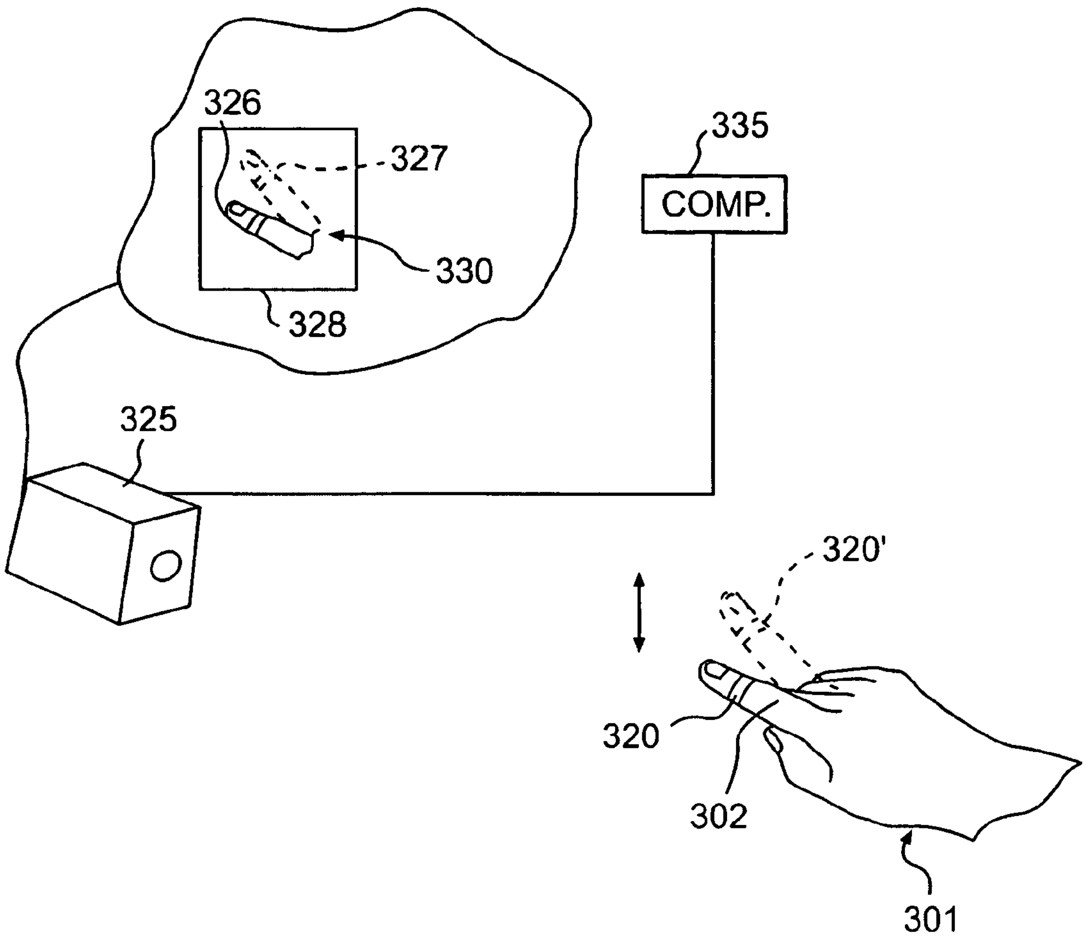


FIG. 3A

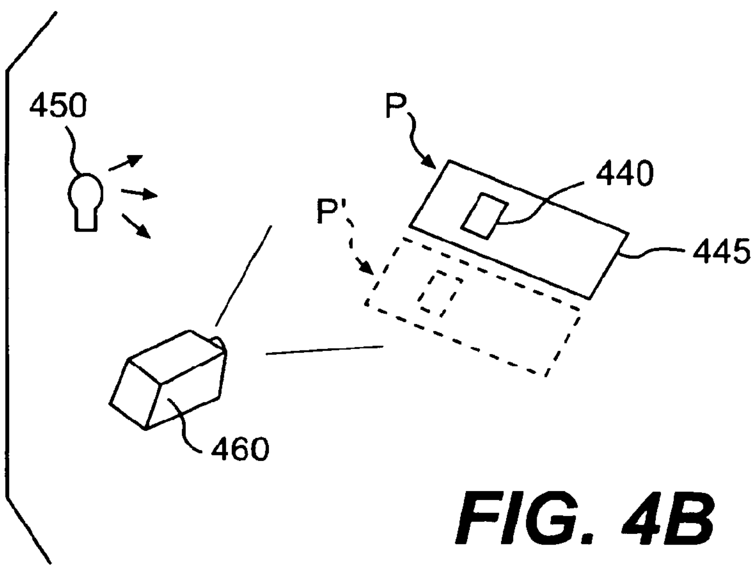


FIG. 4B

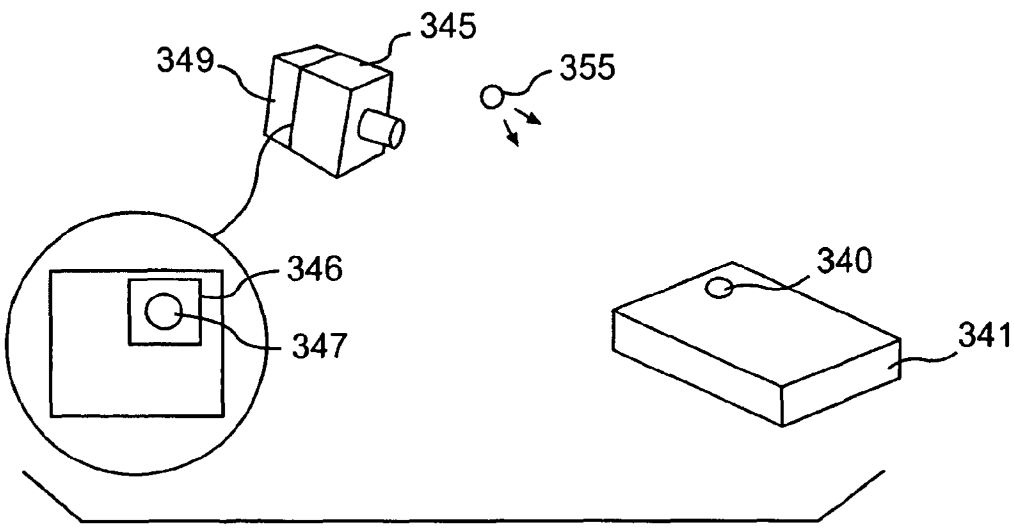


FIG. 3B

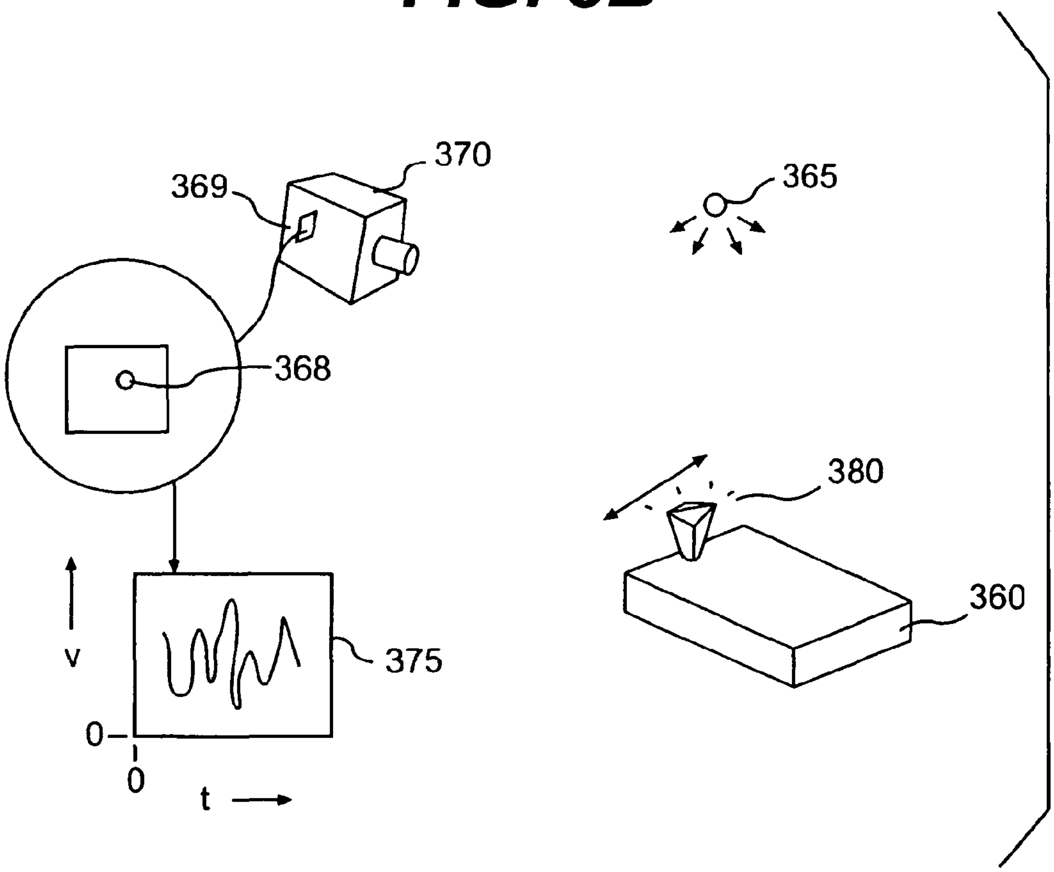


FIG. 3C

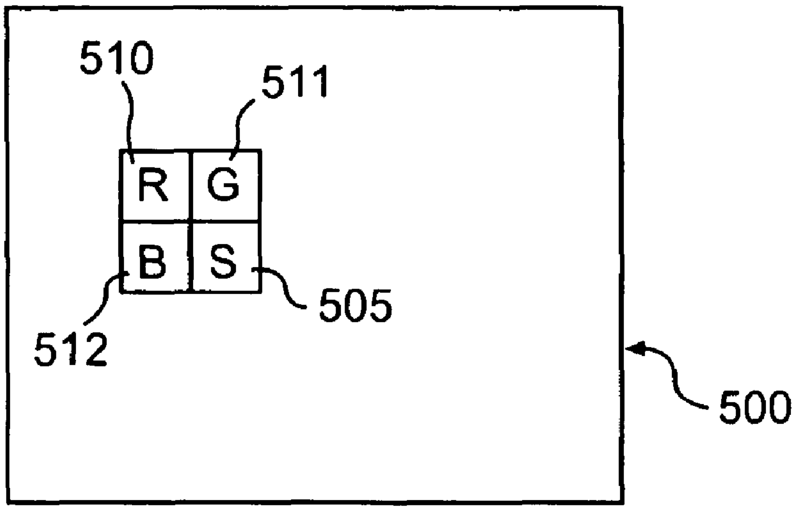


FIG. 5A

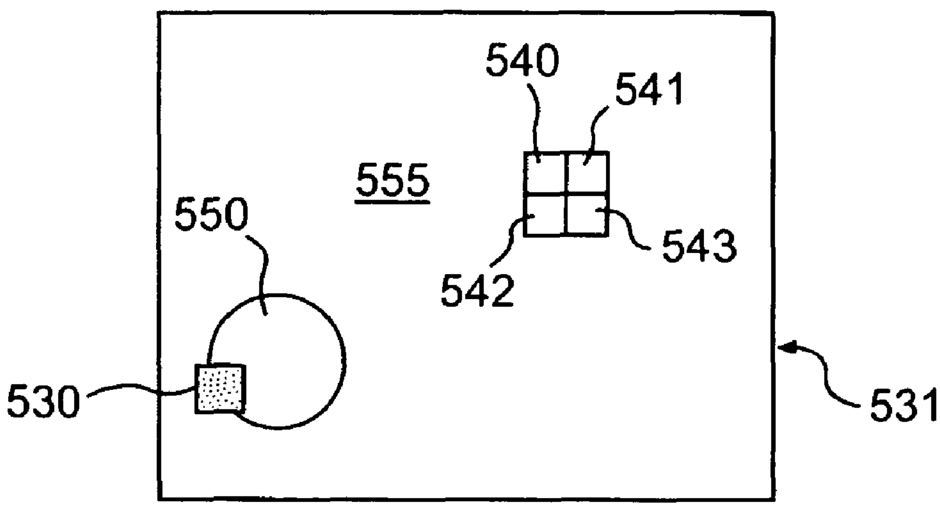


FIG. 5B

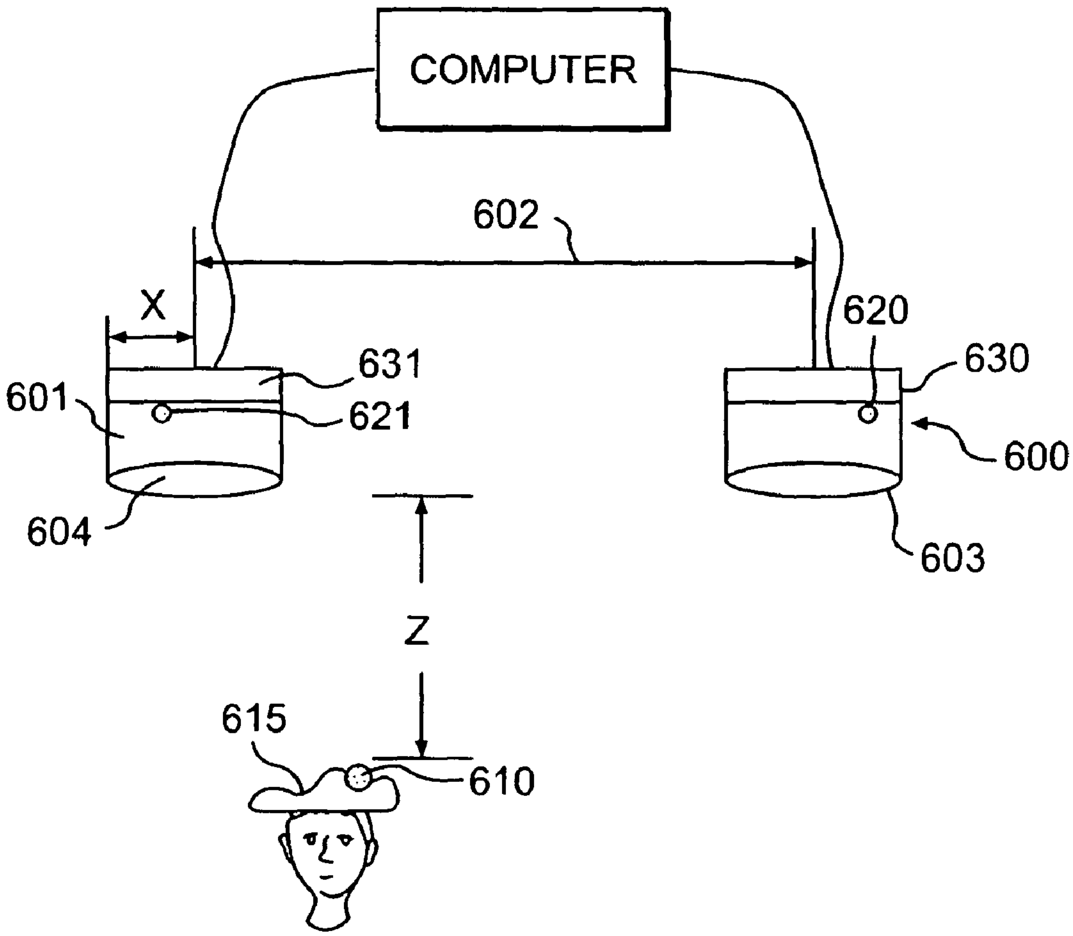


FIG. 6

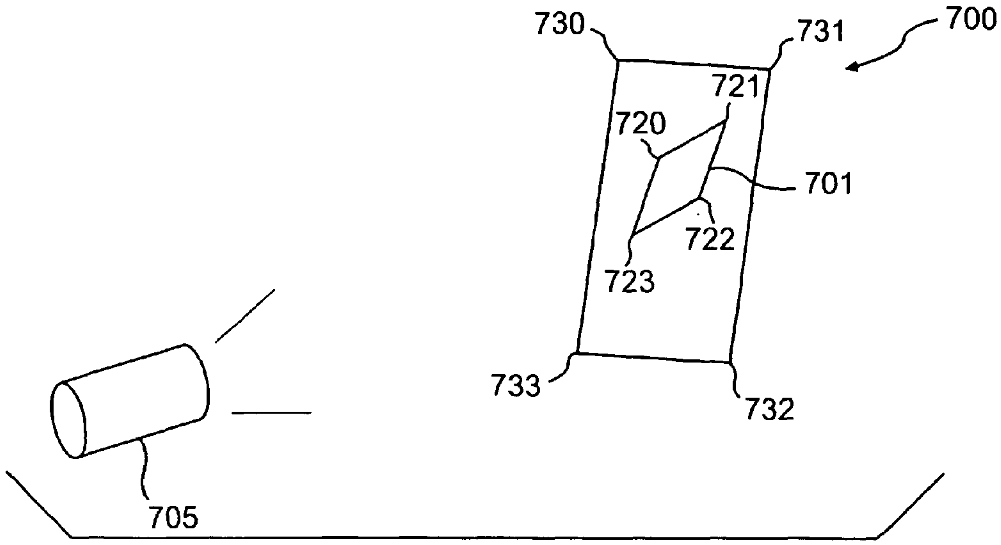

700
730
731
721
720
701
722
723
733
732
705

FIG. 7

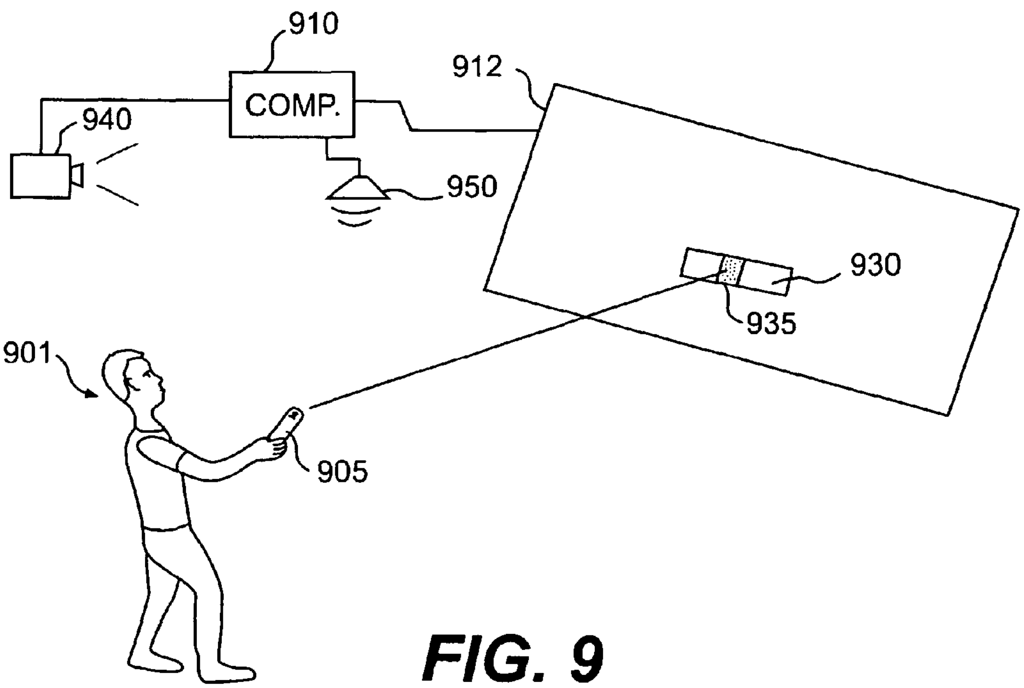

910
COMP.
912
940
950
930
935
901
905

FIG. 9

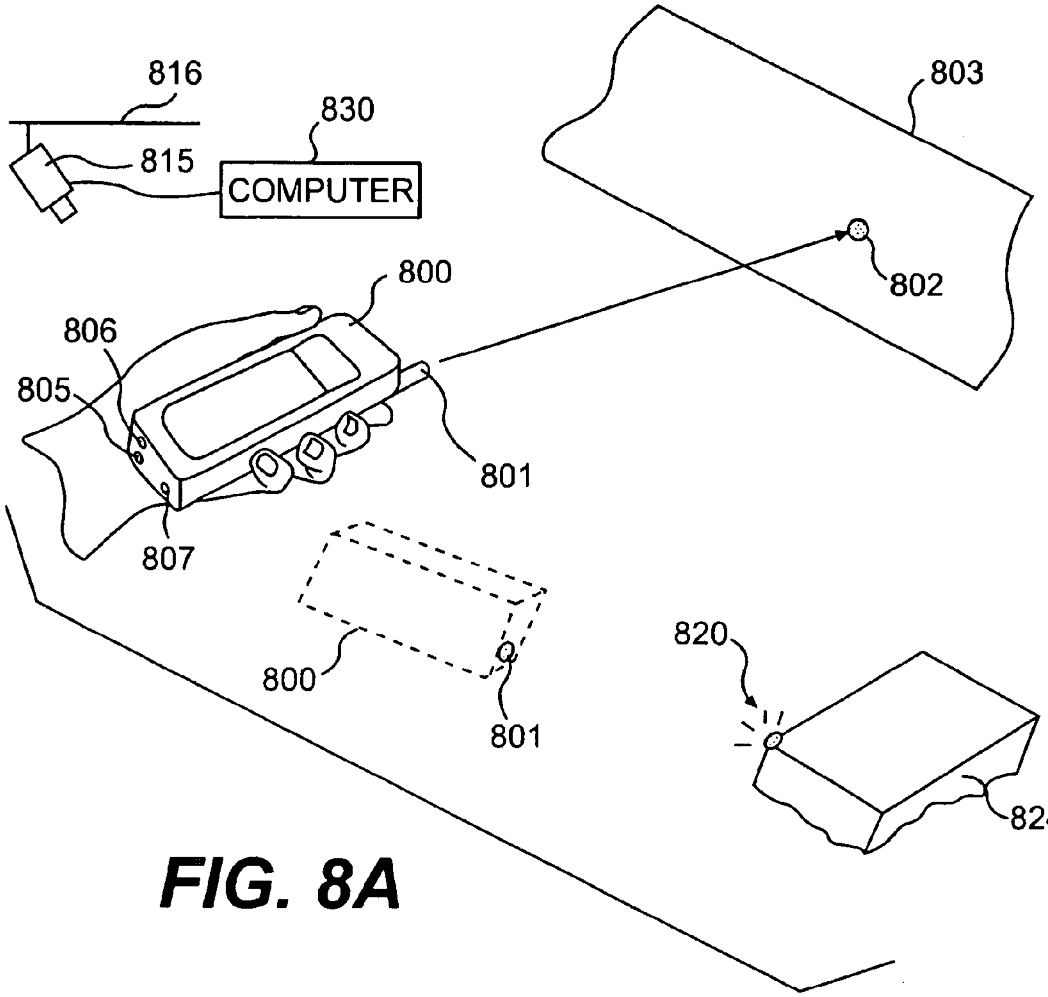

816
830
815
COMPUTER
803
802
800
806
805
801
807
800
801
820
824

FIG. 8A

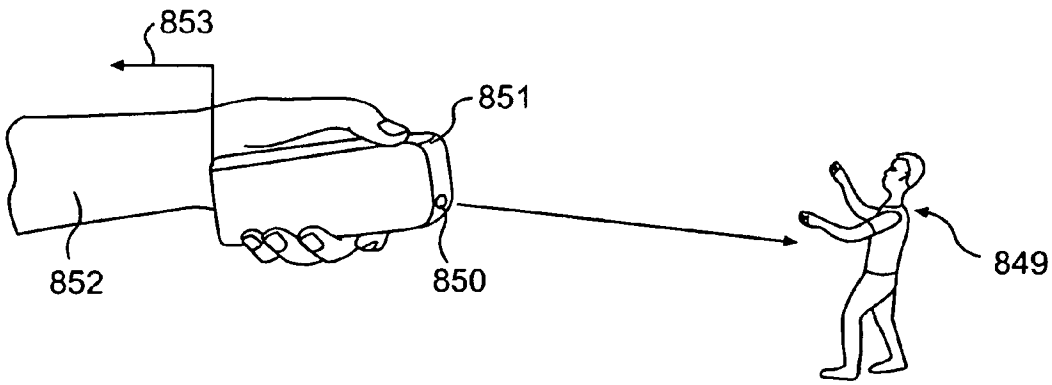
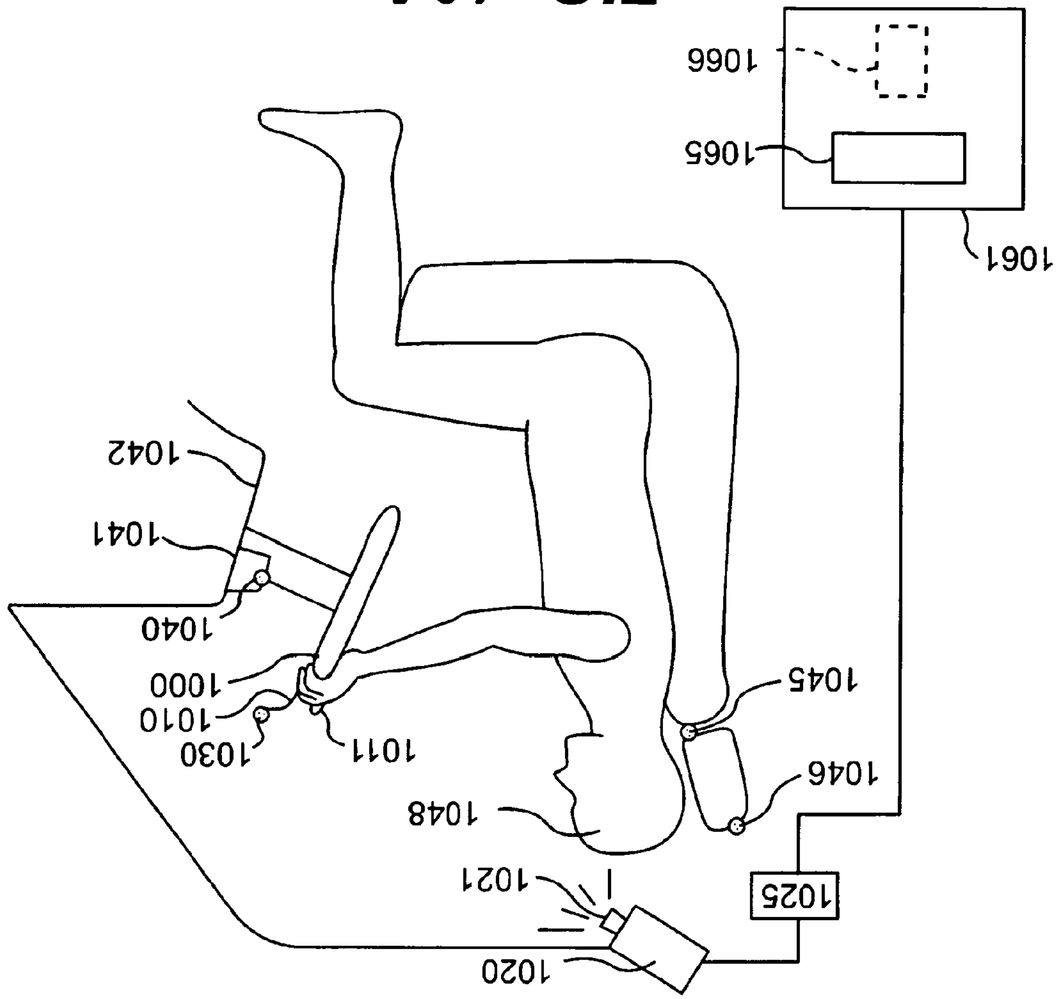

853
851
852
850
849

FIG. 8B

FIG. 10A


1066
1065
1061
1042
1041
1040
1000
1010
1030
1011
1045
1046
1048
1021
1025
1020

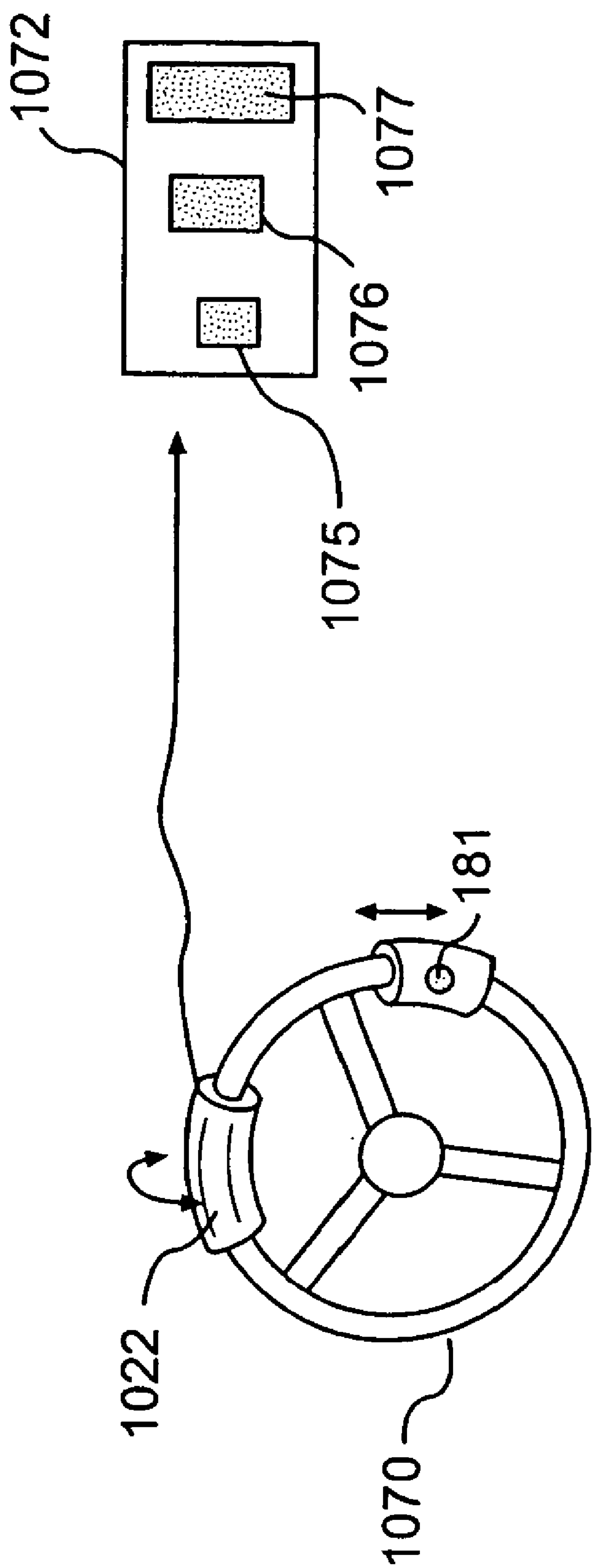


FIG. 10B

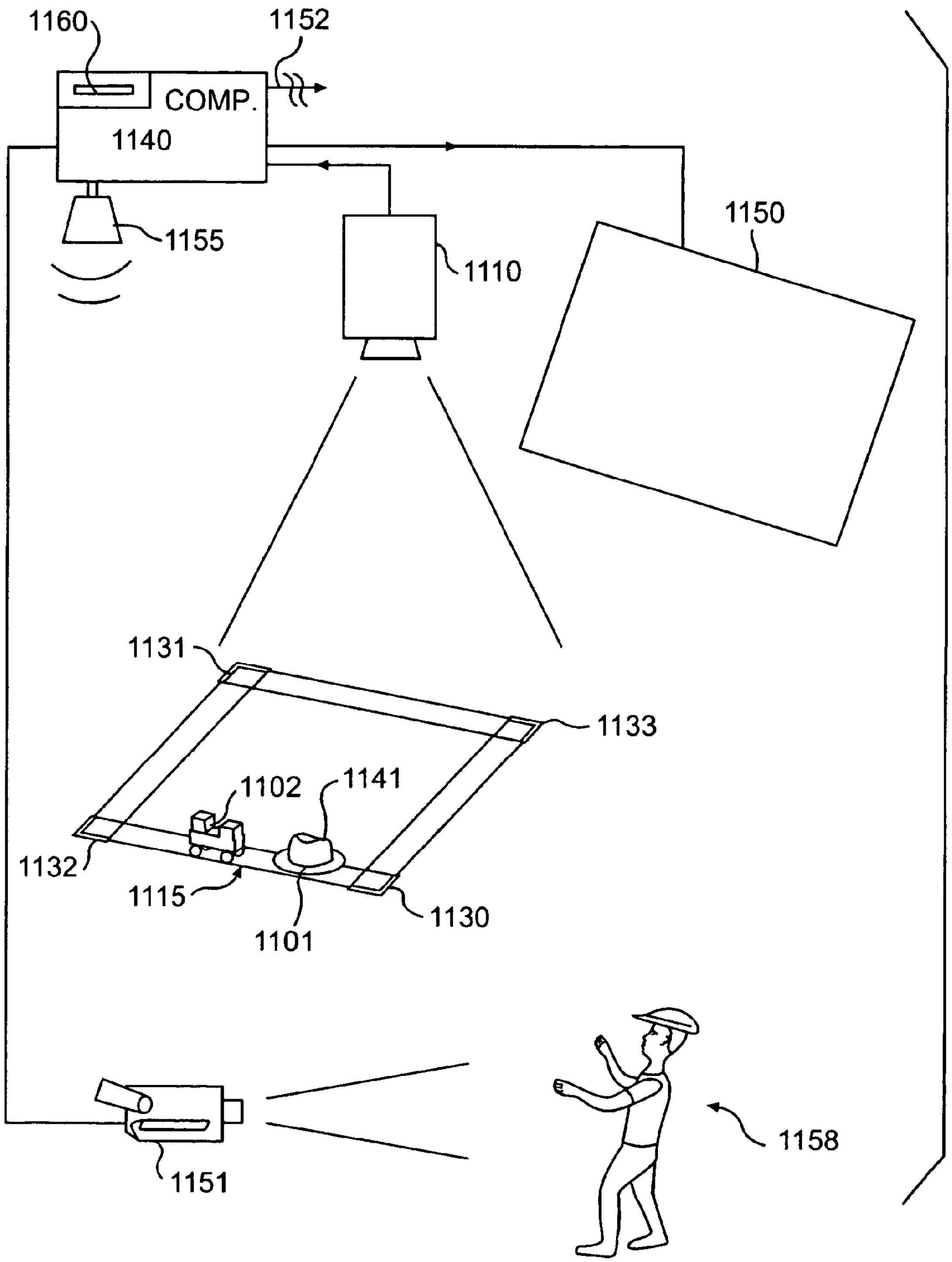

1160
1152
COMP.
1140
1155
1110
1150
1131
1133
1102
1141
1132
1115
1101
1130
1151
1158

FIG. 11A

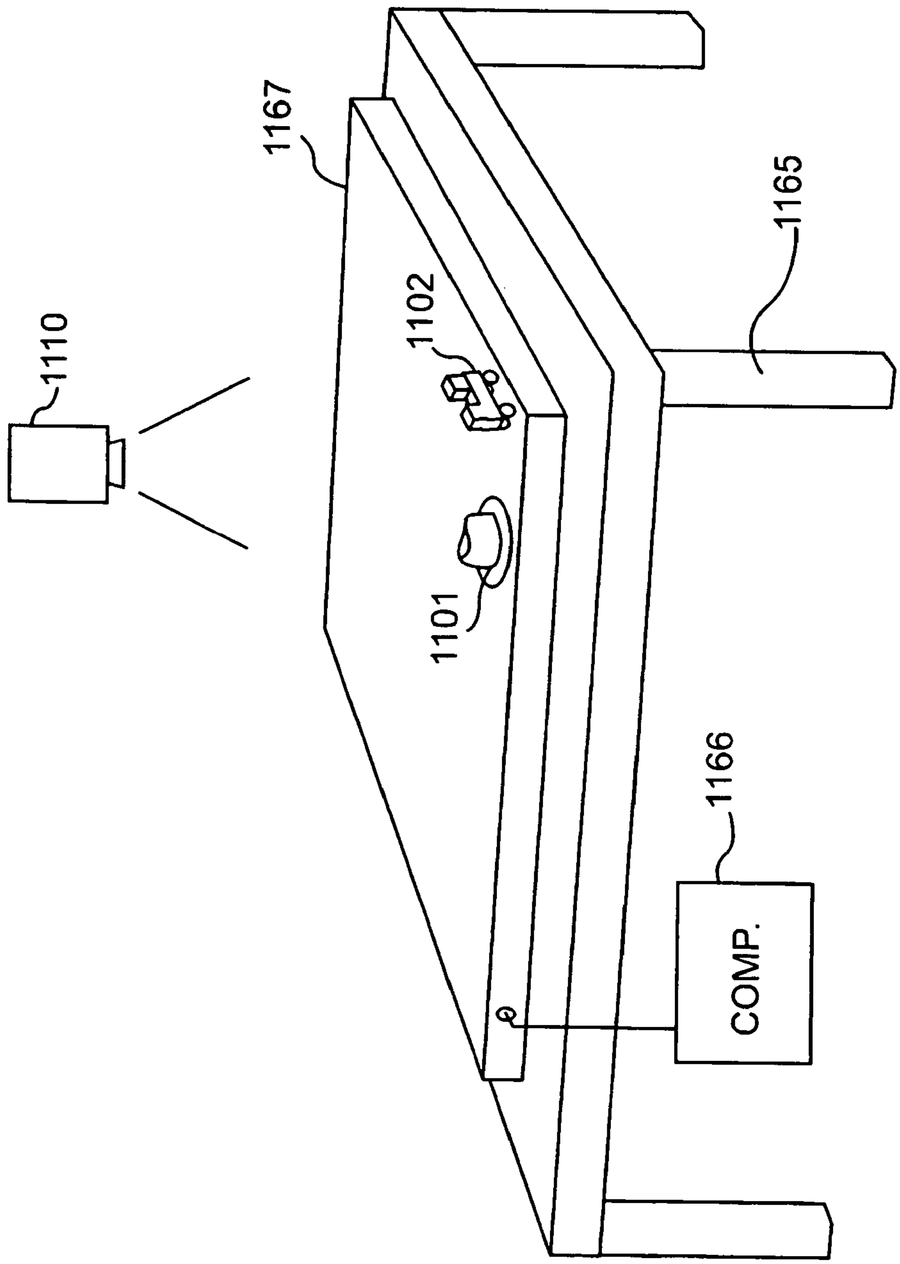


FIG. 11B

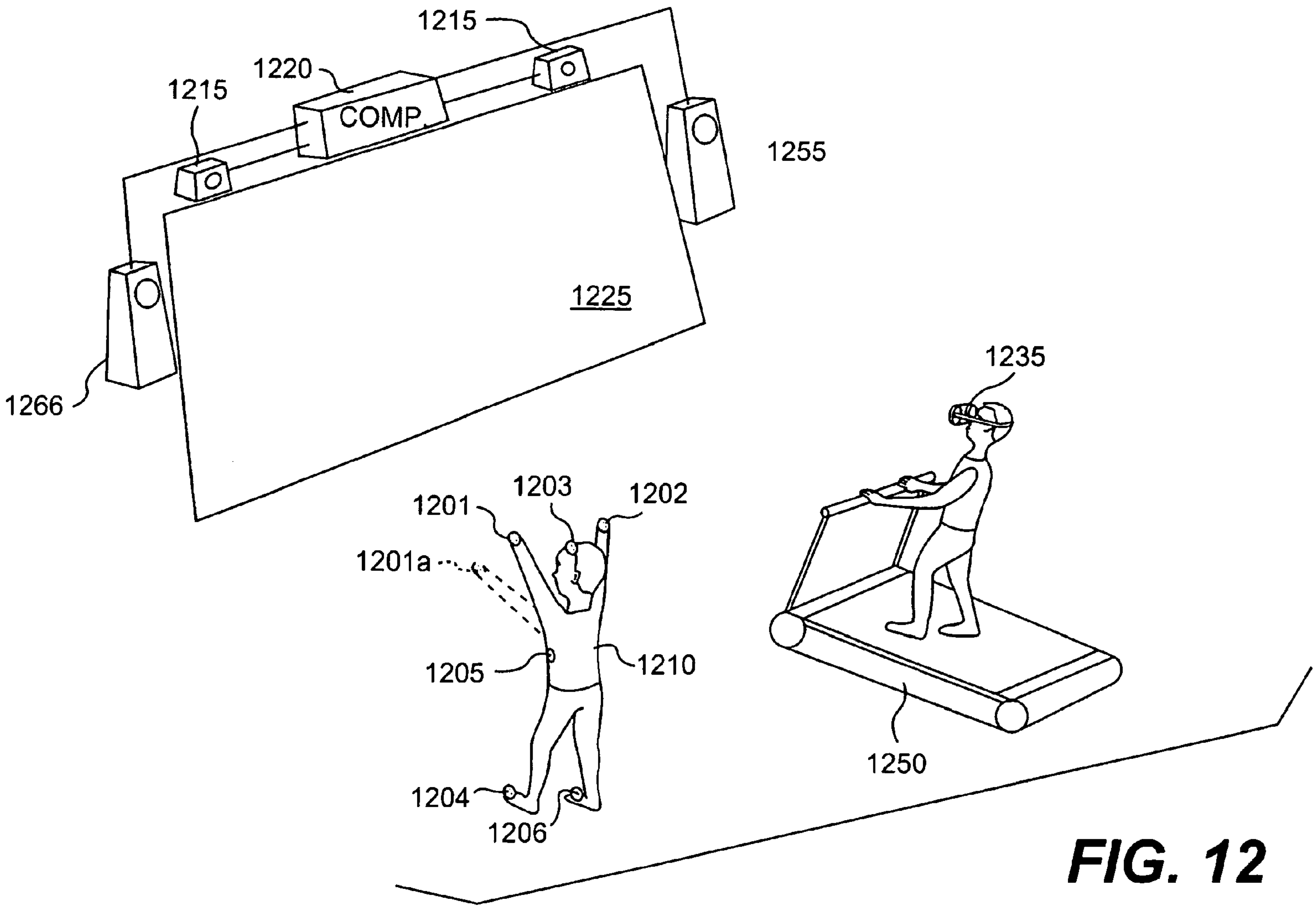

1215
1220
COMP.
1215
1255
1225
1266
1235
1201
1203
1202
1201a
1205
1210
1204
1206
1250

FIG. 12

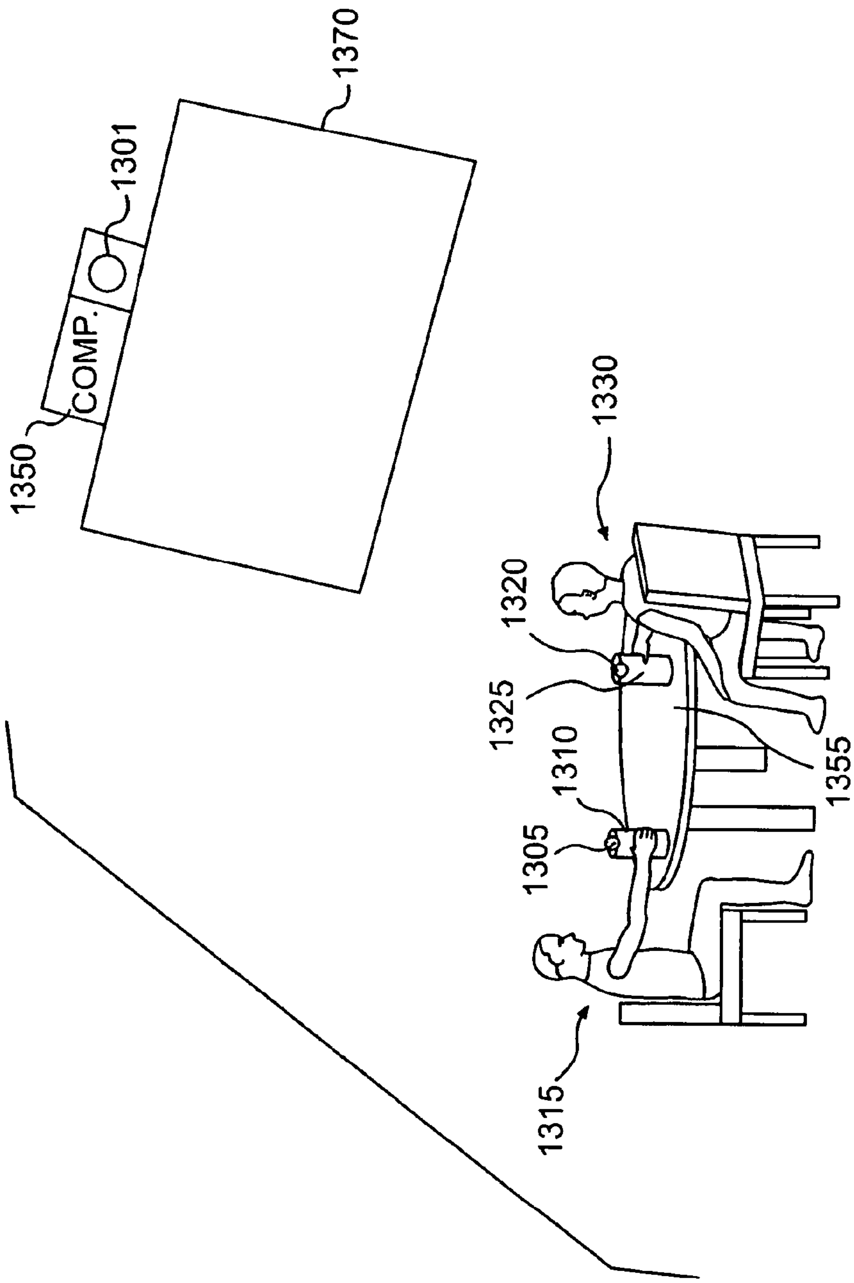


FIG. 13

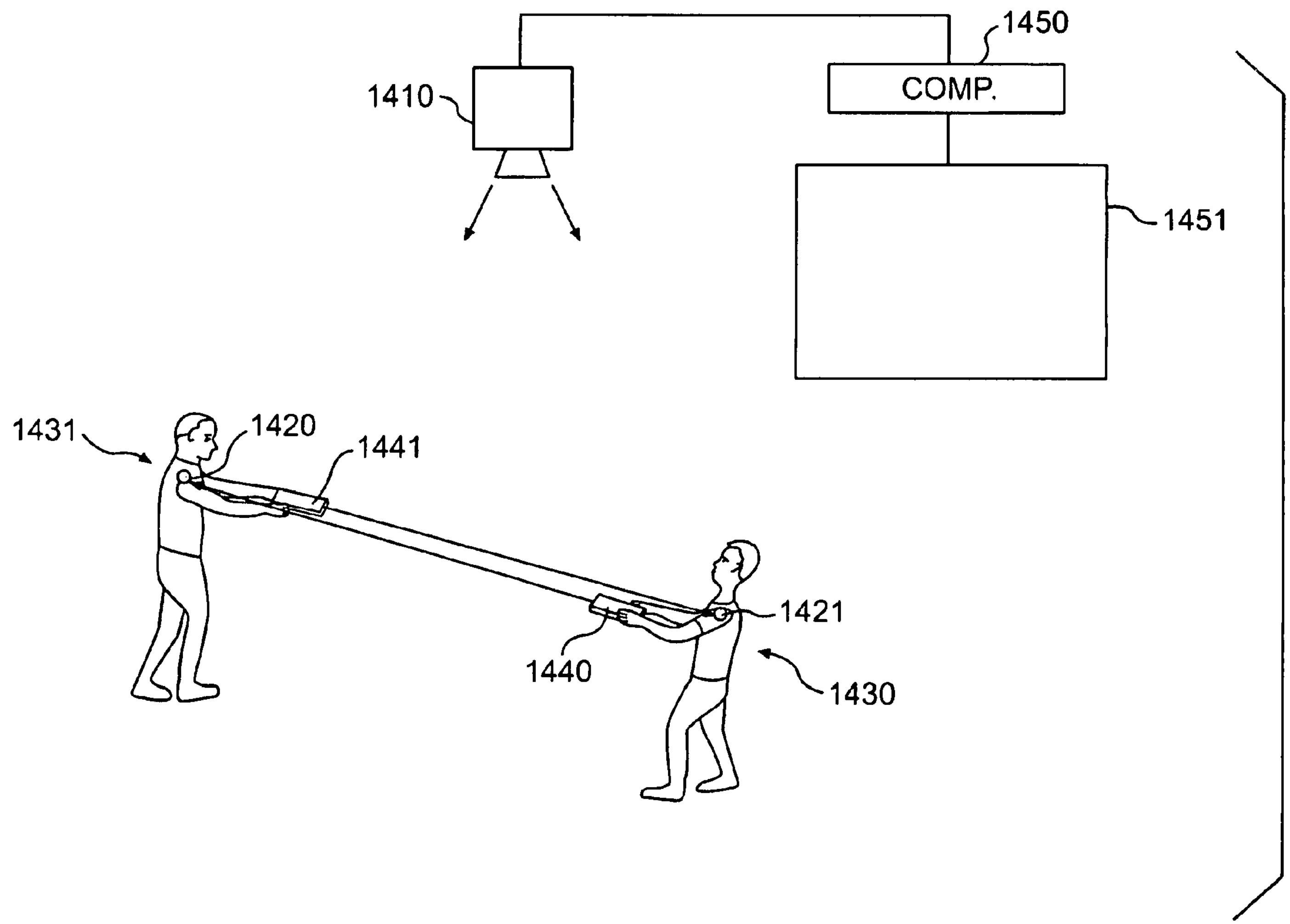


FIG. 14A

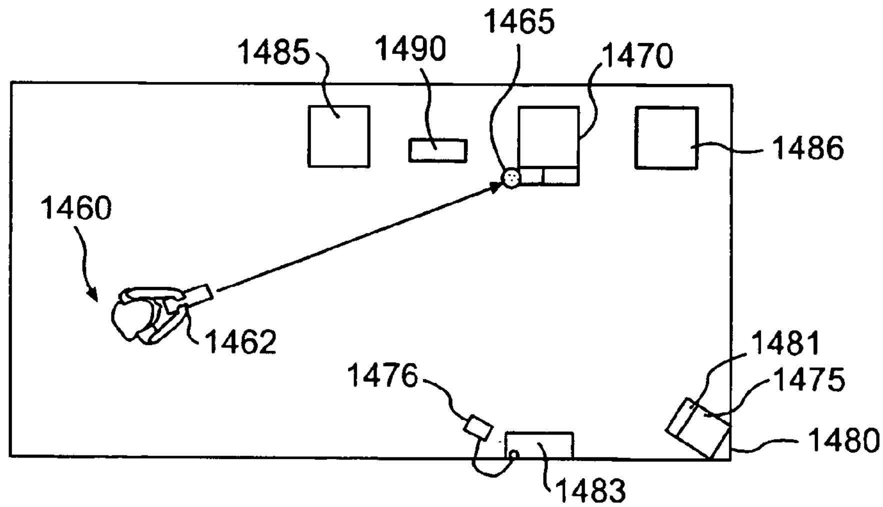


FIG. 14B

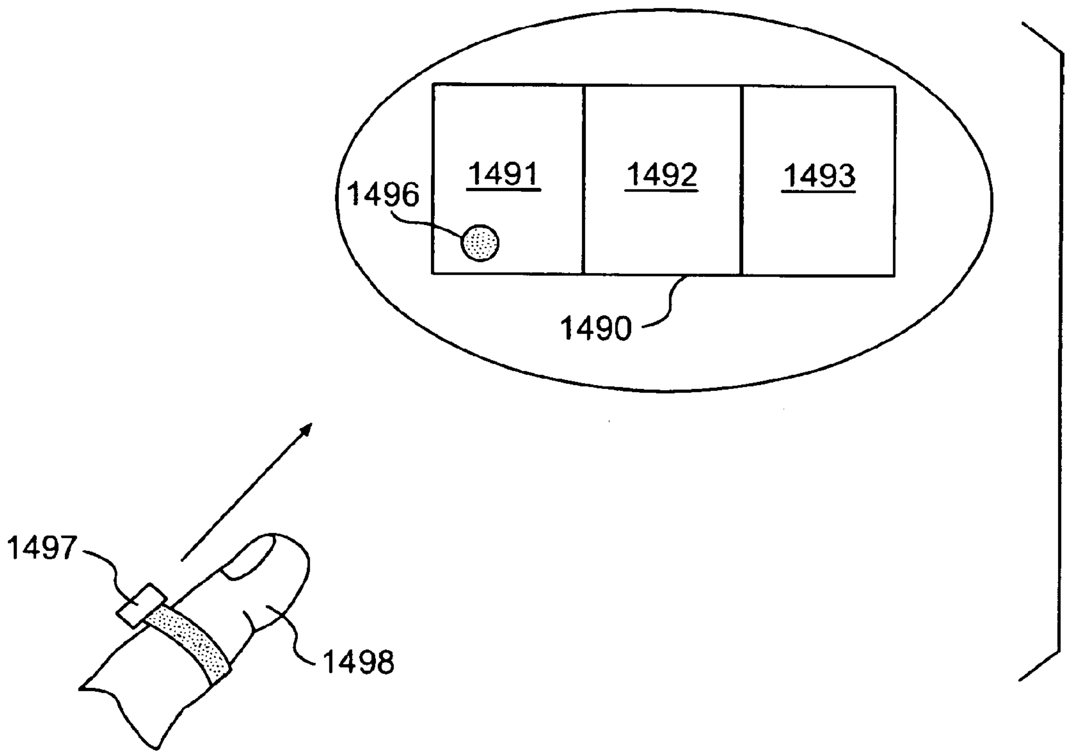


FIG. 14C

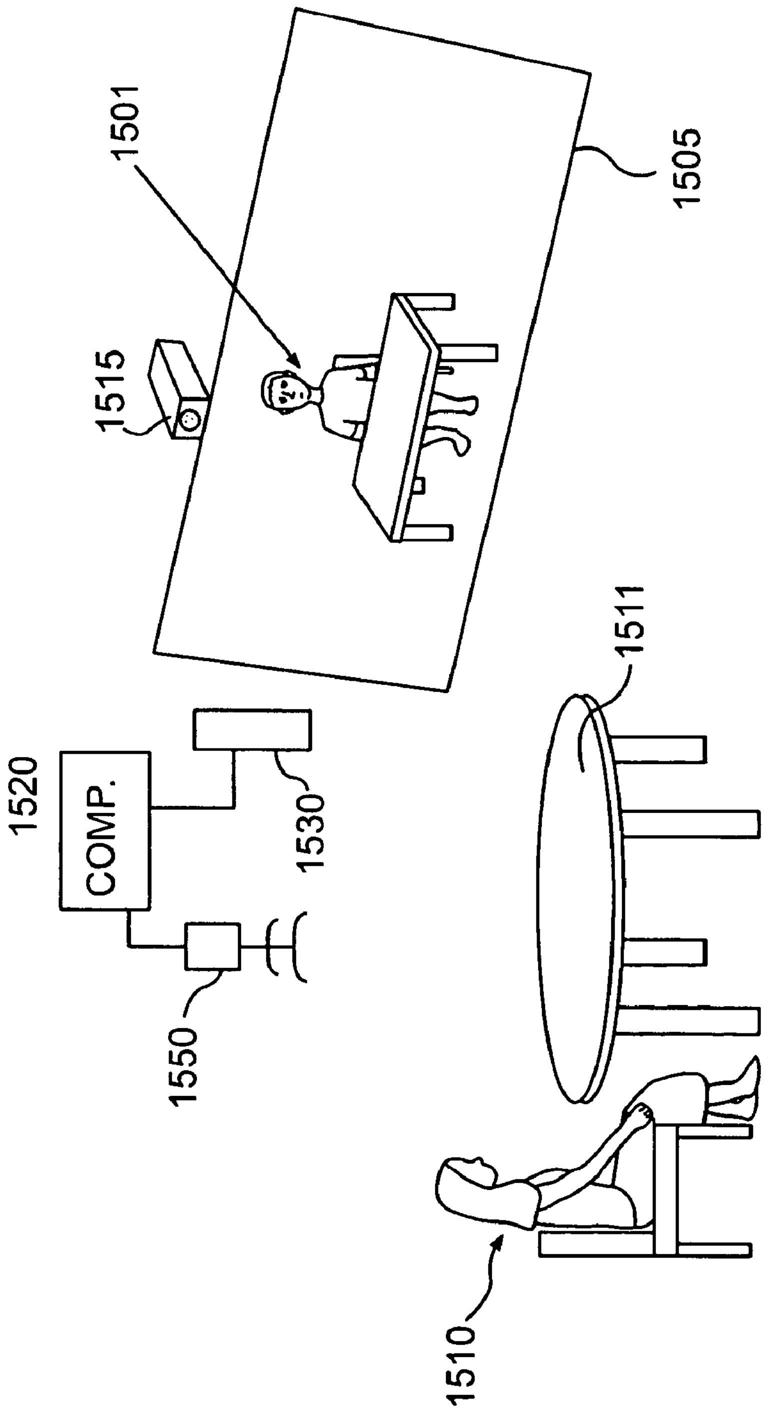


FIG. 15

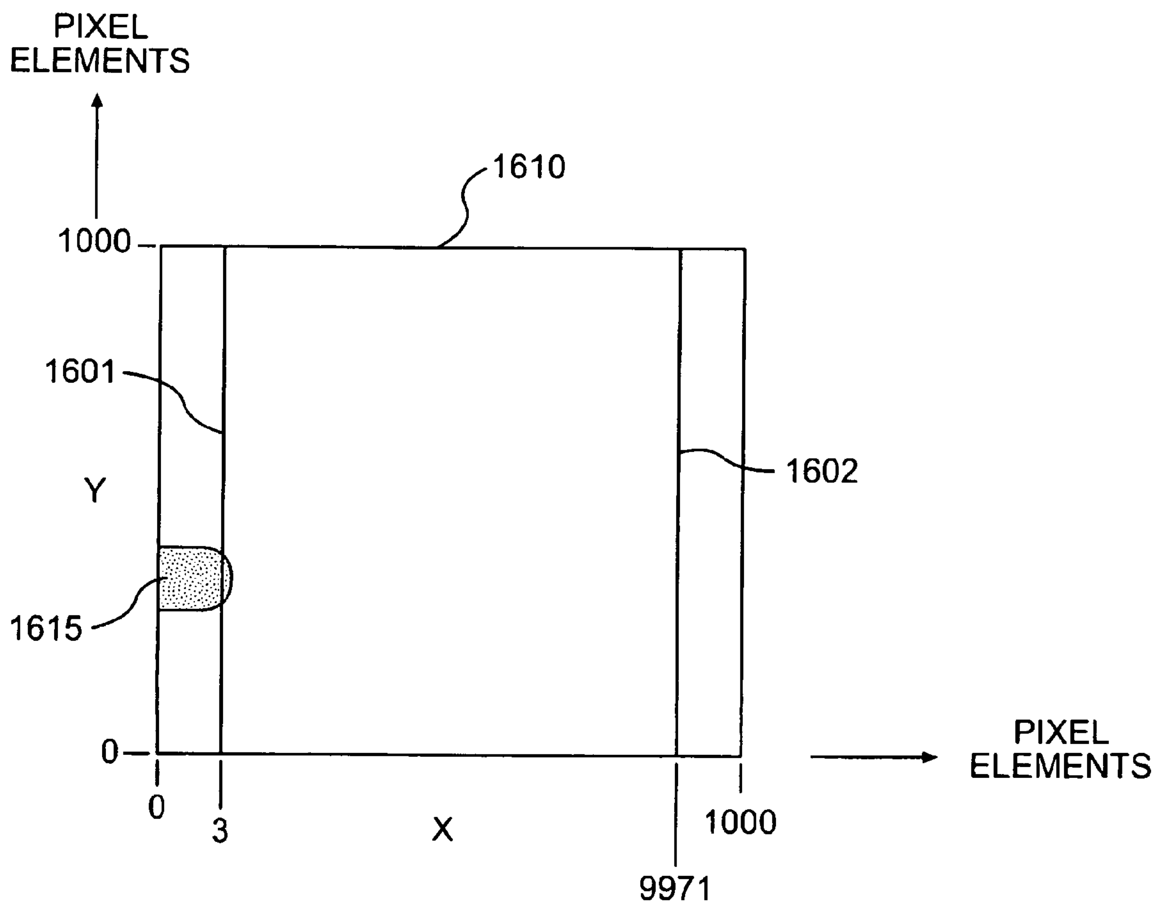

PIXEL
ELEMENTS
1610
1000
1601
Y
1602
1615
0
PIXEL
ELEMENTS
0
3
X
1000
9971

FIG. 16

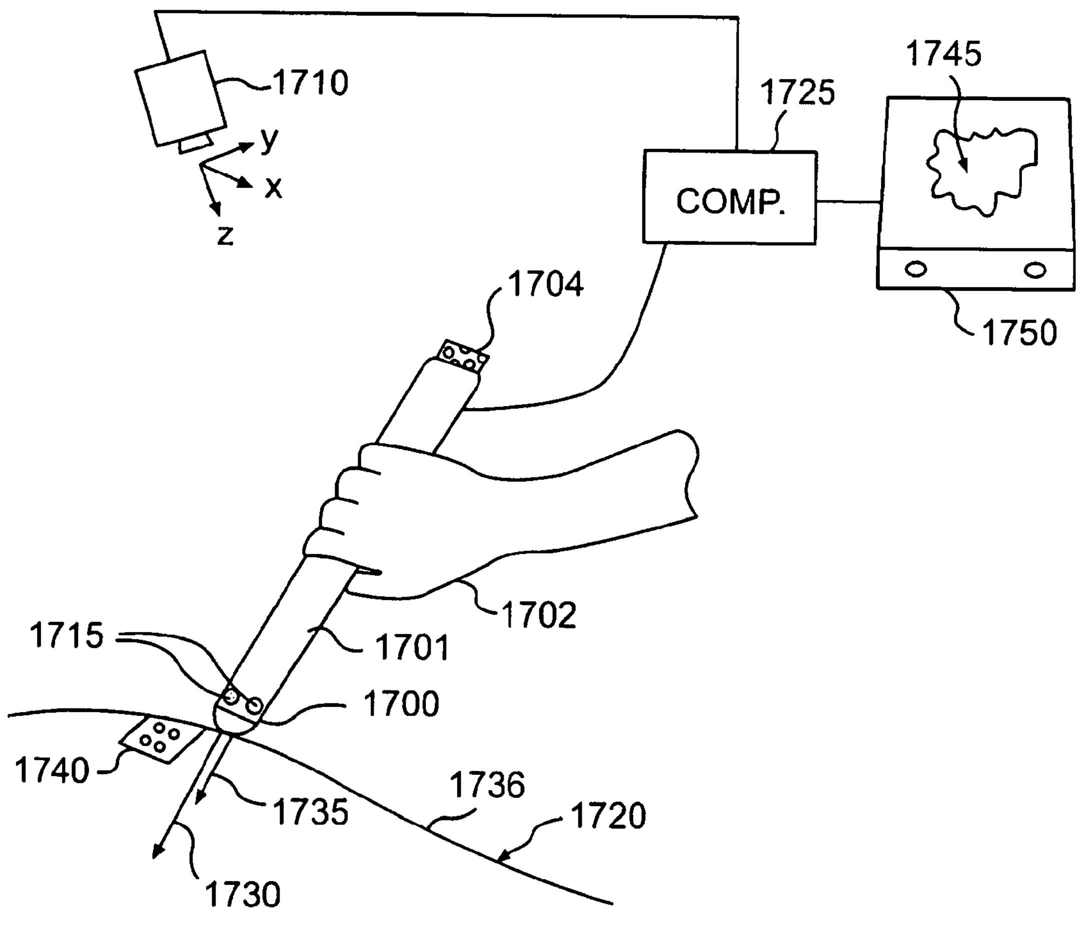


FIG. 17A

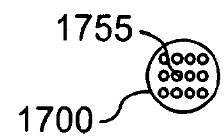


FIG. 17C

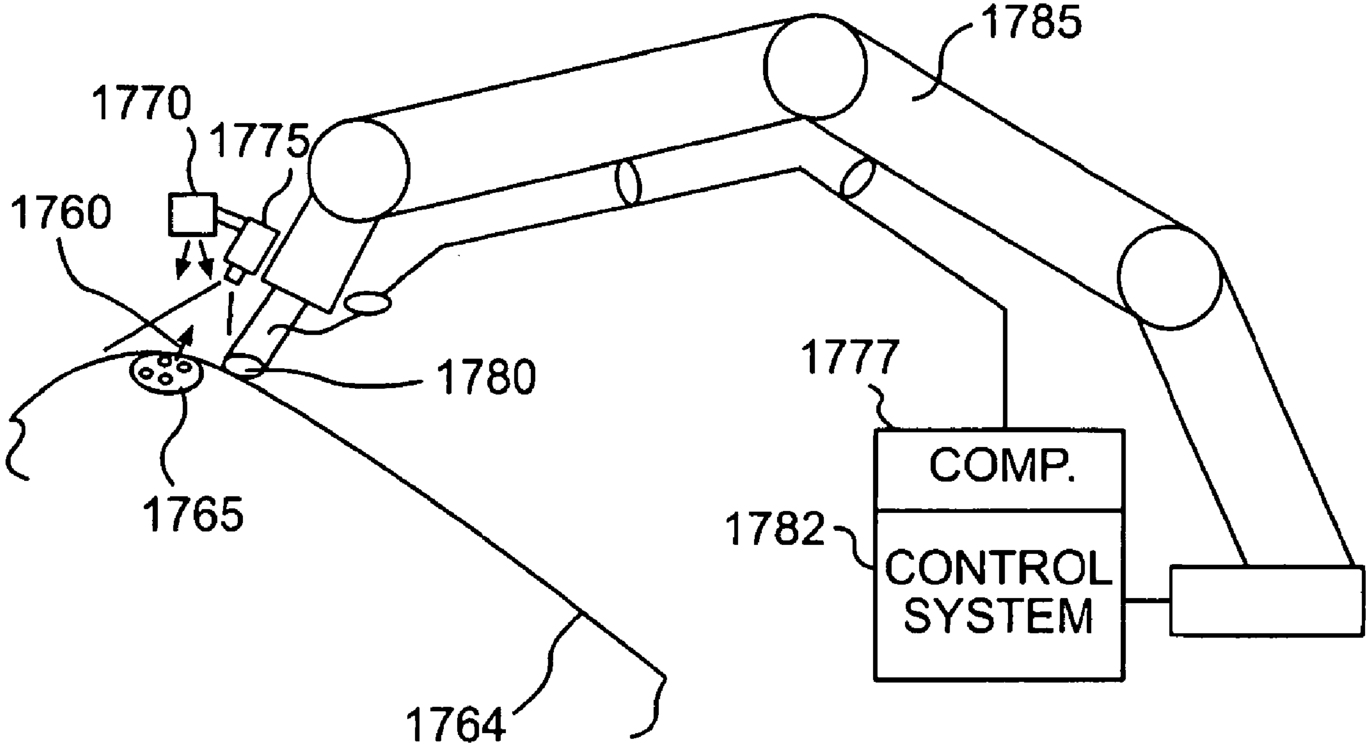

1785
1770
1775
1760
1780
1777
COMP.
1765
1782
CONTROL SYSTEM
1764

FIG. 17B

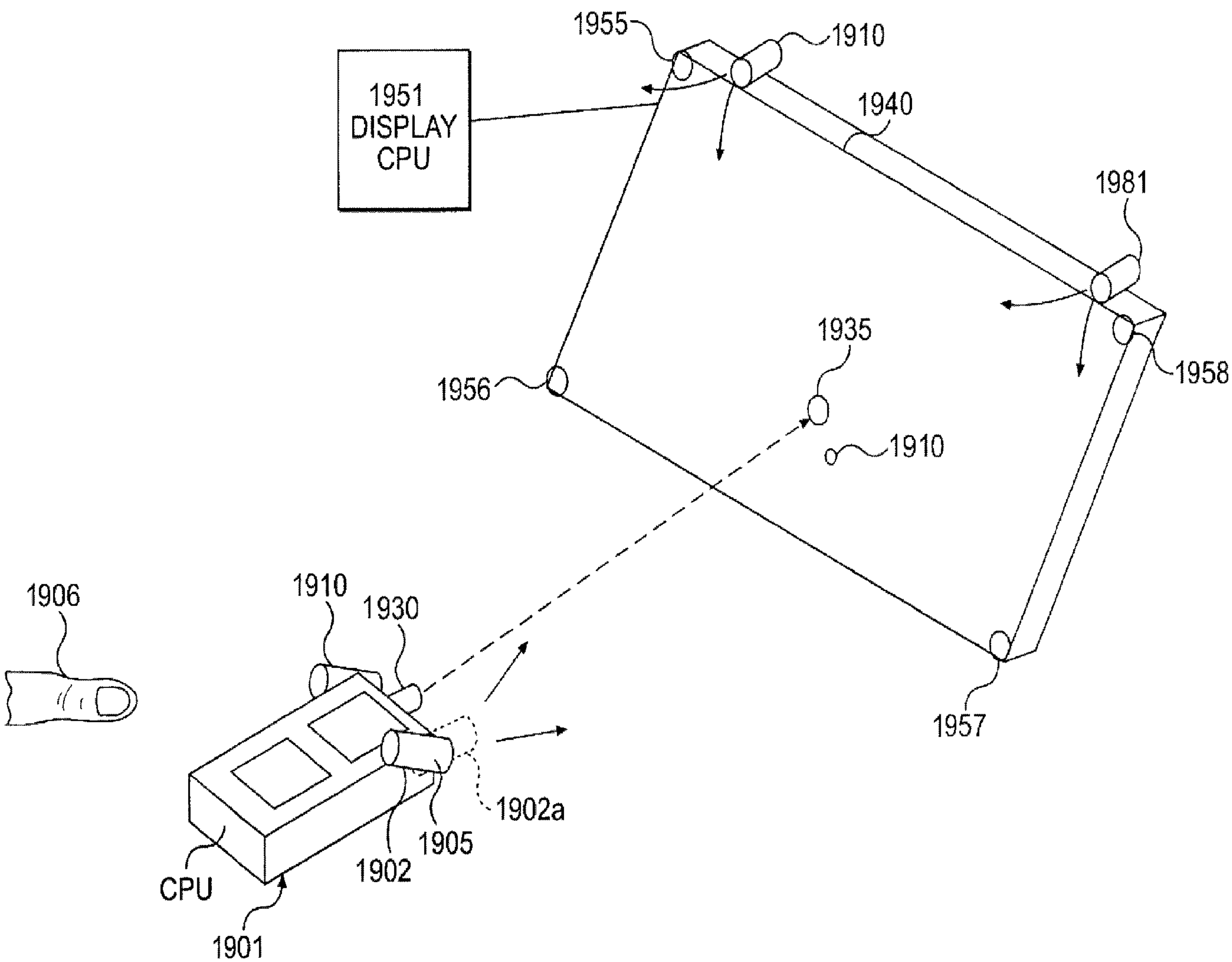

1955
1910
1951
DISPLAY
CPU
1940
1981
1935
1958
1956
1910
1906
1910
1930
1957
1902a
1905
1902
CPU
1901

Fig. 18

CAMERA BASED SENSING IN HANDHELD, MOBILE, GAMING OR OTHER DEVICES

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation of application Ser. No. 12/834,281, filed Jul. 12, 2010 (now U.S. Pat. No. 7,933,431), which is a continuation of application Ser. No. 11/980,710, filed Oct. 31, 2007 (now U.S. Pat. No. 7,756,297), which is a continuation of application Ser. No. 10/893,534, filed Jul. 19, 2004 (now U.S. Pat. No. 7,401,783), which is a continuation of application Ser. No. 09/612,225, filed Jul. 7, 2000 (now U.S. Pat. No. 6,766,036), which claims the benefit of U.S. Provisional Application No. 60/142,777, filed Jul. 8, 1999.

Cross references to related co-pending US applications by the inventor having similar subject matter.

1. Touch TV and other Man Machine Interfaces: Ser. No. 09/435,854 filed Nov. 8, 1999, now U.S. Pat. No. 7,098,891; which was a continuation of application Ser. No. 07/946,908, now U.S. Pat. No. 5,982,352;

2. More Useful Man Machine Interfaces and Applications: Ser. No. 09/433,297 filed Nov. 3, 1999, now U.S. Pat. No. 6,750,848;

3. Useful Man Machine interfaces and applications: Ser. No. 09/138,339, Pub. Appln. 2002-0036617, now abandoned;

4. Vision Target based assembly: Ser. No. 08/469,907 filed Jun. 6, 1995, now U.S. Pat. No. 6,301,783;

5. Picture Taking method and apparatus: provisional application 60/133,671, and regular application Ser. No. 09/568,552 filed May 11, 2000, now U.S. Pat. No. 7,015,950;

6. Methods and Apparatus for Man Machine Interfaces and Related Activity: Provisional Application: provisional application 60/133,673 filed May 11, 1999; and regular application Ser. No. 09/568,554 filed May 11, 2000, now U.S. Pat. No. 6,545,670;

7. Tactile Touch Screens for Automobile Dashboards, Interiors and Other Applications: provisional application Ser. No. 60/183,807; and regular application Ser. No. 09/789,538, now U.S. Pat. No. 7,084,859; and

8. Apparel Manufacture and Distance Fashion Shopping in Both Present and Future: provisional application 60/187,397 filed Mar. 7, 2000.

The disclosures of the following U.S. patents and co-pending patent applications by the inventor, or the inventor and his colleagues, are incorporated herein by reference:

1. "Man machine Interfaces": U.S. application Ser. No. 09/435,854 and U.S. Pat. No. 5,982,352, and U.S. application Ser. No. 08/290,516, filed Aug. 15, 1994, now U.S. Pat. No. 6,008,000, the disclosure of both of which is contained in that of Ser. No. 09/435,854;

2. "Useful Man Machine Interfaces and Applications": U.S. application Ser. No. 09/138,339, now Pub. Appln. 2002-0036617;

3. "More Useful Man Machine Interfaces and Applications": U.S. application Ser. No. 09/433,297, now U.S. Pat. No. 6,750,848;

4. "Methods and Apparatus for Man Machine Interfaces and Related Activity": U.S. Appln. Ser. No. 60/133,673 filed as regular application Ser. No. 09/568,554, now U.S. Pat. No. 6,545,670;

5. "Tactile Touch Screens for Automobile Dashboards, Interiors and Other Applications": U.S. provisional Appln. Ser. No. 60/183,807, filed Feb. 22, 2000, now filed as reg. application Ser. No. 09/789,538; and

6. "Apparel Manufacture and Distance Fashion Shopping in Both Present and Future": U.S. Appln. Ser. No. 60/187,397, filed Mar. 7, 2000.

FIELD OF THE INVENTION

The invention relates to simple input devices for computers, particularly, but not necessarily, intended for use with 3-D graphically intensive activities, and operating by optically sensing a human input to a display screen or other object and/or the sensing of human positions or orientations. The invention herein is a continuation in part of several inventions of mine, listed above.

This continuation application seeks to provide further useful embodiments for improving the sensing of objects. Also disclosed are new applications in a variety of fields such as computing, gaming, medicine, and education. Further disclosed are improved systems for display and control purposes.

The invention uses single or multiple TV cameras whose output is analyzed and used as input to a computer, such as a home PC, to typically provide data concerning the location of parts of, or objects held by, a person or persons.

DESCRIPTION OF RELATED ART

The above mentioned co-pending applications incorporated by reference discuss many prior art references in various pertinent fields, which form a background for this invention. Some more specific U.S. Patent references are for example:

DeMenthon—U.S. Pat. Nos. 5,388,059; 5,297,061; 5,227,985

Cipolla—U.S. Pat. No. 5,581,276

Pugh—U.S. Pat. No. 4,631,676

Pinckney—U.S. Pat. No. 4,219,847

DESCRIPTION OF FIGURES

FIG. **1** illustrates a basic computer terminal embodiment of the invention, similar to that disclosed in copending applications.

FIG. **2** illustrates object tracking embodiments of the invention employing a pixel addressable camera.

FIG. **3** illustrates tracking embodiments of the invention using intensity variation to identify and/or track object target datums.

FIG. **4** illustrates tracking embodiments of the invention using variation in color to identify and/or track object target datums.

FIG. **5** illustrates special camera designs for determining target position in addition to providing normal color images.

FIG. **6** identification and tracking with stereo pairs.

FIG. **7** illustrates use of an indicator or co-target.

FIG. **8** illustrates control of functions with the invention, using a handheld device which itself has functions.

FIG. **9** illustrates pointing at an object represented on a screen using a finger or laser pointer, and then manipulating the represented object using the invention.

FIG. **10** illustrates control of automobile or other functions with the invention, using detected knob, switch or slider positions.

FIG. **11** illustrates a board game embodiment of the invention.

FIG. **12** illustrates a generic game embodiment of the invention.

FIG. **13** illustrates a game embodiment of the invention, such as might be played in a bar.

FIG. **14** illustrates a laser pointer or other spot designator embodiment of the invention.

FIG. **15** illustrates a gesture based flirting game embodiment of the invention.

FIG. **16** illustrates a version of the pixel addressing camera technique wherein two lines on either side of a 1000 element square array are designated as perimeter fence lines to initiate tracking or other action.

FIG. **17** illustrates a 3-D acoustic imaging embodiment of the invention.

FIG. **18** illustrates an improved handheld computer embodiment of the invention, in which the camera or cameras may be used to look at objects, screens and the like as well as look at the user.

THE INVENTION EMBODIMENTS

FIG. **1**

The invention herein and disclosed in portions of other copending applications noted above, comprehends a combination of one or more TV cameras (or other suitable electro-optical sensors) and a computer to provide various position and orientation related functions of use. It also comprehends the combination of these functions with the basic task of generating, storing and/or transmitting a TV image of the scene acquired—either in two or three dimensions.

The embodiment depicted in FIG. **1**A illustrates the basic embodiments of many of my co-pending applications above. A stereo pair of cameras **100** and **101** located on each side of the upper surface of monitor **102** (for example a rear projection TV of 60 inch diagonal screen size) with display screen **103** facing the user, are connected to PC computer **106** (integrated in this case into the monitor housing), for example a 400 Mhz Pentium II. For appearances and protection a single extensive cover window may be used to cover both cameras and their associated light sources **110** and **111**, typically LEDs.

The LEDs in this application are typically used to illuminate targets associated with any of the fingers, hand, feet and head of the user, or objects such as **131** held by a user, **135** with hands **136** and **137**, and head **138**. These targets, such as circular target **140** and band target **141** on object **131** are desirably, but not necessarily, retro-reflective, and may be constituted by the object features themselves (e.g., a finger tip, such as **145**), or by features provided on clothing worn by the user (e.g., a shirt button **147** or polka dot **148**, or by artificial targets other than retroreflectors.

Alternatively, a three camera arrangement can be used, for example using additional camera **144**, to provide added sensitivity in certain angular and positional relationships. Still more cameras can be used to further improve matters, as desired. Alternatively, and or in addition, camera **144** can be used for other purposes, such as acquire images of objects such as persons, for transmission, storage or retrieval independent of the cameras used for datum and feature location determination.

For many applications, a single camera can suffice for measurement purposes as well, such as **160** shown in FIG. **1**B for example, used for simple 2 dimensional (2D) measurements in the xy plane perpendicular to the camera axis (z axis), or 3D (xyz, roll pitch yaw) where a target grouping, for example of three targets is used such as the natural features formed by the two eyes **164**, **165** and nose **166** of a human **167**. These features are roughly at known distances from each other, the data from which can be used to calculate the approximate position and orientation of the human face. Using for example the photogrammetric technique of Pinkney described below, the full 6 degree of freedom solution of the human face location and orientation can be achieved to an accuracy limited by the ability of the camera image processing software utilized to determine the centroids or other delineating geometric indicators of the position of the eyes and nose, (or some other facial feature such as the mouth), and the accuracy of the initial imputing of the spacing of the eyes and their respective spacing to the nose. Clearly if a standard human value is used (say for adult, or for a child or even by age) some lessening of precision results, since these spacings are used in the calculation of distance and orientation of the face of human **167** from the camera **160**.

In another generally more photogrammetrically accurate case, one might choose to use four special targets (e.g., glass bead retro-reflectors, or orange dots) **180-183** on the object **185** having known positional relationships relative to each other on the object surface, such as one inch centers. This is shown in FIG. **1**C, and may be used in conjunction with a pixel addressable camera such as described in FIG. **2** below, which allows one to rapidly determine the object position and orientation and track its movements in up to 6 degrees of freedom as disclosed by Pinkney U.S. Pat. No. 4,219,847 and technical papers referenced therein. For example, the system described above for FIGS. **1** and **2** involving the photogrammetric resolution of the relative position of three or more known target points as viewed by a camera is known and is described in a paper entitled "A Single Camera Method for the 6-Degree of Freedom Sprung Mass Response of Vehicles Redirected by Cable Barriers" presented by M. C. van Wijk and H. F. L. Pinkney to The Society of Photo-optical Instrumentation Engineers.

The stereo pair of cameras can also acquire a two view stereo image of the scene as well, which can be displayed in 3D using stereoscopic or auto-stereoscopic means, as well as transmitted or recorded as desired.

In many applications of the foregoing invention it is desirable not just to use a large screen but in fact one capable of displaying life size images. This particularly relates to human scaled images, giving a life-like presence to the data on the screen. In this way the natural response of the user with motions of hands, head, arms, etc., is scaled in "real" proportion to the data being presented.

FIG. **2**

This embodiment and others discloses special types of cameras useful with the invention. In the first case, that of FIG. **2**A, a pixel addressable camera such as the MAPP2200 made by IVP corporation of Sweden is used, which allows one to do many things useful for rapidly determining location of objects, their orientation and their motion.

For example, as shown in FIG. **2**A, an approximately circular image **201** of a target datum such as **180** on object **185** of FIG. **1**C may be acquired by scanning the pixel elements on a matrix array **205** on which the image is formed. Such an array in the future will have for example 1000.times.1000 pixels, or more (today the largest IVP makes is 512.times.512. The IVP also is not believed to be completely randomly addressable, which some future arrays will be).

As an illustration, computer **220** determines, after the array **205** has been interrogated, that the centroid "x, y" of the pixel elements on which the target image lies is at pixel x=500, y=300 (including a sub-fraction thereof in many cases). The centroid location can be determined for example by the moment method disclosed in the Pinkney patent, referenced above.

The target in this case is defined as a contrasting point on the object, and such contrast can be in color as well as, or instead of, intensity. Or with some added preprocessing, it can be a distinctive pattern on the object, such as a checkerboard or herringbone.

Subsequent Tracking

To subsequently track the movement of this target image, it is now only necessary to look in a small pixel window composed of a small number of pixels around the target. For example the square **230** shown, as the new position x'y' of the target image cannot be further distant within a short period of time elapsed from the first scan, and in consideration of the small required time to scan the window.

For example, if the window is 100.times.100 pixels, this can be scanned in 1 millisecond or less with such a pixel addressing camera, by interrogating only those pixels in the window, while still communicating with the camera over a relatively slow USB serial link of 12 mb transmission rate (representing 12,000 pixel gray level values in one millisecond).

One thus avoids the necessity to scan the whole field, once the starting target image position is identified. This can be known by an initial scan as mentioned, or can be known by having the user move an object with a target against a known location with respect to the camera such as a mechanical stop, and then indicate that tracking should start either by verbally saying so with voice recognition, or by actuating a control key such as **238** or whatever.

It is noted that if the tracking window is made large enough, then it can encompass a whole group of datums, such as **180-183** on an object.

FIG. **2**B Reduction in Acquisition Time

Another application of such a pixel addressing camera is shown in FIG. **2**B. One can look at the whole field, x y of the camera, **240**, but only address say every 10.sup.th pixel such as **250**, **251** and **252**, in each direction, i.e., for a total 10,000 pixels in a field of 1 million (1000.times.1000, say).

In this case computer **220** simply queries this fraction of the pixels in the image, knowing apriori that the target image such as **260** will have an image size larger than 10.times.10pixels, and must be detectable, if of sufficient contrast, by one of the queried pixels. (For smaller or larger target images, the number and spacing of queried pixels can be adjusted accordingly). This for example, allows one to find approximate location of targets with only 1/100 the pixel interrogation time otherwise needed, for example, plus any gain obtained as disclosed above, by knowing in what region of the image to look (for example during tracking, or given some apriori knowledge of approximate location due to a particular aspect of the physical arrangement or the program in question).

Once a target has been approximately found as just described, the addressing can be optimized for that region of the image only, as disclosed in subsequent tracking section above.

Given the invention, the potential for target acquisition in a millisecond or two thus is achievable with simple pixel addressable CMOS cameras coming on stream now (today costing under $50), assuming the target points are easily identifiable from at least one of brightness (over a value), contrast (with respect to surroundings), color, color contrast, and more difficult, shape or pattern (e.g., a plaid, or herringbone portion of a shirt). This has major ramifications for the robustness of control systems built on such camera based acquisition, be they for controlling displays, or machines or whatever.

It's noted that with new 2000.times.2000 cameras coming on stream, it may only be necessary to look at every 15.sup.th or 20.sup.th pixel in each direction to get an adequate feel for target location. This means every 200.sup.th to 400.sup.th pixel, not enough to cause image rendition difficulties even if totally dark grey (as it might be in a normal white light image if set up for IR wavelengths only).

FIG. **2**C

Another method for finding the target in the first place with limited pixel interrogation is to look at pixels near a home point where a person for example indicates that the target is. This could be for example, placing ones fingernail such as **270**, whose natural or artificial (e.g., reflective nail polish) features are readily seen by the camera **275** and determined to be in the right corner of a pad **271** in FIG. **2**C which approximately covers the field of view **274** of the camera **275**. The computer **220** analyzes the pixels in the right corner **278** of the image field **279** representing the pad portion **271** with the camera **275**, either continuously, or only when the finger for example hits a switch such as **280** at the edge of the pad, or on command (e.g., by the user pushing a button or key, or a voice message inputted via microphone **285** for example). After such acquisition, the target is then tracked to other locations in xy space of the pad, for example as described above. Its noted that it helps to provide a beep or other sound or indication when acquisition has been made.

Pick Windows in Real Time

Another aspect of the invention is that one can also pick the area of the image to interrogate at any desired moment. This can be done by creating a window of pixels with in the field to generate information, for example as discussed relative to a specific car dashboard application of FIG. **10**.

FIG. **2**D—Scan Pattern

A pixel addressing camera also allows a computer such as **220** to cause scans to be generated which are not typical raster scans. For example circular or radial, or even odd shapes as desired. This can be done by providing from the computer the sequential addresses of the successive pixels on the camera chip whose detected voltages are to be queried.

A circular scan of pixels addressed at high speed can be used to identify when and where a target enters a field enclosed by the circular pixel scan. This is highly useful, and after that, the approximate location of the target can be determined by further scans of pixels in the target region.

For example consider addressing the pixels c**1** c**2** c**3** . . . cn representing a circle **282** at the outer perimeter of the array, **285**, of 1000.times.1000 elements such as discussed above. The number of pixels in a full circle is approximately 1000 pi, which can be scanned even with USB (universal serial bus) limits at 300 times per second or better. For targets of 1/100 field in width, this means that a target image entering the field such as circular target image **289** (which is shown intersecting element cm and its neighbors) would have to travel 1/100 the field width in 0.0033 seconds to be totally missed in a worst case. If the image field corresponds to 20 inches in object field width this is 0.2 inches.times.300/sec or 60 inches/second, very fast for human movement, and not likely to be exceeded even where smaller targets are used.

Alternative shapes to circular "trip wire" perimeters may be used, such as squares, zig-zag, or other layouts of pixels to determine target presence. Once determined, a group of pixels such as group **292** can be interrogated to get a better determination of target location.

FIG. **3**

Since many applications of the invention concern, or at least have present a human caused motion, or motion of a part of a human, or an object moved by a human, the identification and tracking problem can be simplified if the features of interest, either natural or artificial of the object provide some kind of change in appearance during such motion.

FIG. **3** illustrates tracking embodiments of the invention using intensity variation to identify and/or track object target datums. In a simple case, a subtraction of successive images can aid in identifying zones in an image having movement of features as is well known. It is also useful to add pixel intensities of successive images in computer **220** for example. This is particular true with bright targets (with respect to their usual surroundings) such as LEDs or retro-reflectors. If the pixels in use by the camera are able to gather light preferentially at the same time a special illumination light is on, this will accentuate the target with respect to background. And if successive frames are taken in this way, not only will a stationary image of the special target build up, but if movement takes place the target image then will blur in a particular direction which itself can become identify-able. And the blur direction indicates direction of motion as well, at least in the 2-D plane of the pixel array used.

Another form of movement can take place artificially, where the target is purposely moved to provide an indication of its presence. This movement can be done by a human easily by just dithering ones finger for example (if a portion of the finger such as the tip is the target in question), or by vibrating an object having target features of interest on it, for example by moving the object up and down with ones hand.

For example consider FIG. **3**A, where a human **301** moves his finger **302** in a rapid up and down motion, creating different image positions sequentially in time of bright target ring **320**, **320'** on his finger, as seen by camera **325**. If the camera can read quickly enough each of these positions such as **326** and **327** in image field **328** can be resolved, other wise a blur image such as **330** is registered on the camera and recorded in the computer **335**.

Instead of using ones finger, it is also possible to create movement of a target for example with a tuning fork or other mechanism mechanically energizing the target movement, on what otherwise might be a static object say. And it is possible for the human, or a computer controlling the movement in question to create it in such a manner that it aids identification. For example, a certain number of moves of ones finger (e.g., 4), or 2 moves/sec of ones finger, or horizontal moves of ones finger etc., any or all of these could indicate to the computer upon analysis of the camera image, that a target was present.

The invention comprehends this as a method for acquiring the datum to be tracked in the first place, and has provided a camera mechanism for tracking fast enough not to lose the data, assuming a sufficiently distinct feature. For example, it is desirable to not require sophisticated image processing routines and the like if possible, to avoid the time it takes to execute same with affordable equipment. And yet in many scenes, finding a target cant be done easily today without some aid, either a high contrast target (contrasting brightness or color or both, for example). Or the aid can be movement as noted, which allows the search for the target to be at least localized to a small region of the field of view, and thence take much less time to run, even if a sophisticated algorithm is employed.

FIG. **3**B illustrates an embodiment wherein a target which blinks optically is used. The simplest case is a modulated LED target such **340** on object **341** shown. Successive frames taken with camera **345** looking at pixel window **346** at **300** scans of the pixels within the window per second where the image **347** of the LED target is located, can determine, using computer **349** (which may be separate from, or incorporated with the image sensor), 5 complete blinks of target **340**, if blinked at a 60 hz rate. Both blink frequency, blink spacing, blink pulse length can all be determined if the scan rate is sufficiently faster than the blink rate, or pulse time.

It should be noted that if the target **340** is a retro-reflector as in FIG. **1**, with an illumination source such as **355** near the axis of the camera, then the LEDs (or other sources) of the illuminator can be modulated, causing the same effect on the target.

Somewhat more sophisticated is the situation shown in FIG. **3**C where a target **380** (on object **360**) illuminated by a light source **365** provides a time variant intensity change in the camera image **368** obtained by camera **370** as the target moves its position and that of the image. This can be achieved naturally by certain patterns of material such as herringbone, or by multifaceted reflectors such as cut diamonds (genuine or glass), which “twinkle” as the object moves. A relative high frequency “twinkle” in the image indicates then the presence of the target in that area of the image in which it is found.

When analog sensors such as PSD (position sensing diode) sensor **369** described in a copending application is used in addition to, or instead of a matrix array in camera **370**, the variation in light intensity or twinkle can be obtained directly from the detected output voltage from the signal conditioning of the sensor as shown in trace **375** corresponding to the movement of diamond target **380** a distance in the camera field. From the PSD one can also determine the position of the detected target image, theoretically at least independent of the intensity fluctuation.

For digital array detectors, the intensity variation can also be detected by subtracting images and observing the difference due to such variation. Such images need to be taken frequently if the twinkle frequency is high, and this can cause problems unless high speed camera scanning is possible. For example, in a twinkle mode, a pixel addressable camera using the invention herein could scan every 5.sup.th pixel in both x and y. This would allow a 1000 frame per second operation of a camera which would normally go 40 frames per second. Such a rate should be able to capture most twinkle effects with the assumption that the light field changes on more than 25 pixels. If less, then scan density would need to be increased to every 3.sup.rd pixel say, with a corresponding reduction in twinkle frequency detection obtainable.

FIG. **4**

FIG. **4**A illustrates identification and tracking embodiments of the invention using color and color change in a manner similar in some aspects to the intensity variation from object datums described above.

Color can be used as has been noted previously to identify a target, as can a change in color with time. For example, a target can change its color in order to identify itself to successive interrogations of pixels on a color TV camera. This can be accomplished by having a retro-reflector which is illuminated in succession by light from different colored LEDs for example, in the arrangement of FIG. **1**. For example red led **401** illuminates retro reflector target **405** on object **406** during frame **1** (or partial frame, if not all pixels addressed) taken by camera **410**. Then yellow led **402** illuminates target **405** on the next frame, and so forth. For any reading of successive frames, one point in the image will appear to distinctly change color, while all other points will be more or less the same due to the room lighting overwhelming the led source illumination and the natural color rendition of the objects themselves.

To return color variation when moved, one can employ a target which changes color naturally as it moves, even with illumination of constant color. Such a target can contain a diffractive, refractive, or interference based element, for example, a reflective diffraction grating for example, which splits white light illumination into colors, which are seen differently as the target moves and changes angle with respect to the observer and/or illumination source.

For example, consider FIG. **4**B showing reflective grating **440** on object **445** at initial position P. When illuminated by white light for example from lamp **450**, it reflects the spectrum such that when the object has moved to a new position P' the color (or colors, depending on the grating type, and angles involved) returning to camera **460** is changed. Such gratings can be purchased from Edmund Scientific company, and are typically made as replicas of ruled or holographic gratings.

Some types of natural features which change color are forms of jewelry which have different colored facets pointing in different directions. Also some clothes look different under illumination from different angles. This could be called then "color twinkle".

FIG. **5**

FIG. **5** illustrates special camera designs for determining target position in addition to providing normal color images. As was pointed out in a co-pending application, it may be desirable to have two cameras looking at an object or area one for producing images of a person or scene, the other for feature location and tracking. These may be bore-sighted together using beam splitters or the like to look at the same field, or they may just have largely overlapping image fields. The reason this is desirable is to allow one to obtain images of activity in the field of view (e.g., a human playing a game) while at the same time ideally determine information concerning position or other aspects of features on the human or objects associated with him.

It is now of interest to consider a matrix array chip equipped with a special color filter on its face which passes a special wavelength in certain pixel regions, in addition to providing normal color rendition via RGB or other filtering techniques in the remaining regions. The chip could be pixel addressable, but does not have to be.

Version FIG. **5**A

One version would have one special pixel filter such as **505**, for each square group of 4 pixels in an array **500** (one special pixel filter **505**, and 3 pixels, **510-512** filtered for RGB (red green blue) or similar, as is commonly used now for example. In one functional example, the special pixel **505** is purposely not read during creation of the normal image of a scene, but rather read only on alternate frames (or as desired) to determine target locations. If the array can be addressed pixel wise, the actual time lost doing this can be low. Since 25% of the pixels are effectively dead in forming the image in this example, and assuming all pixels are of equal area (not necessarily required), then 25% of the image needs to be filled in. This can be done advantageously in the image displayed, by making the color and intensity of this pixel the same as the resultant color and average intensity value of the other 3 in the cluster.

Version FIG. **5**B

In this version, related to FIG. **2** above, and shown in FIG. **5***b*, isolated pixels such as **530** (exaggerated in size for clarity) on array **531** or clusters of pixels such as **540-543**, are used to rapidly find a target with low resolution, such as round dot target image **550**. These pixels can ideally have special filters on their face, for example having near IR bandpass filters (of a wavelength which can still be seen by the camera, typically up to 1 um wavelength max). If takes only a few pixels to see the rough presence of a target, then in an image field of 1000.times.1000 pixels there could be one or more target images occupying 10.times.10 pixels or more. Thus in any group of 10.times.10, you could have 5 near IR filtered receptive pixels say, i.e., only 5% of the total pixel count but sufficient to see the IR targets location to a modest accuracy. Once found, one can also use the "normal" pixels on which the target image also falls to aid in more precise determination of its location, for example using pixel group **555** composed of numerous pixels.

In short by having a camera with certain pixels responsive to selected wavelengths and/or scanned separately one can very rapidly scan for target features, then when found, take a regular picture if desired. Or just take regular pictures, until the necessity arises to determine target location.

Similarly the special filtered pixels such as **505** or **530** could be laser wavelength bandpass filtered for this purpose, used by the array for preferentially detecting laser light projected on an object (while ignoring other wavelengths). In a normal image, such a pixel would be nearly black as little white light passes (except that centered on the laser wavelength). To provide a normal picture using such a camera, the special IR or laser wavelengths pixels readings would be filed in with values and colors of light from the surrounding regions.

Such a laser wavelength filter can be extremely effective, even if a relatively weak laser is used to illuminate a large area, especially where retro-reflectors are used, and the light returned is concentrated by 1000 times or more.

FIG. **6**

The embodiments above have dealt with finding just one target, and generally with just one camera, even though two or more cameras may be used for stereo imaging. Where stereo pairs of cameras are used, clearly each camera must see the target, if range via image disparity (the shift in location of a feature in the image in two camera views separated by a baseline) is to be determined.

Using the invention, one camera can be considered a master, the other a slave. The master camera determines target location by any of the means described above. Then the slave need only look at the expected pixel location of the target assuming some a priori knowledge of range which can come from previous target readings, or known range zones where the target has to lie in a given application.

Consider cameras **600** (master) with lens **603** and **601** (slave) having lens **604**, the axes of the two cameras separated by baseline **602** and with interfaced to computer **605**. The image of target **610** on object **615** is formed at position **620** on array **630** of camera **600**, and at position **621** on array **631** of camera **601**. The difference in position x in the direction of the baseline, in this simple situation is directly proportional to range z. The knowledge then of target image position **620** found by interrogating some or all of the pixels of camera **600** can as mentioned be used to more rapidly find image **621** in the image field of the "slave" camera **601**, and thus the z location of the target **610**.

For example if range is known to be an approximate value of z, one can look in the image field of the camera **601** along a line of points at a calculated value x away from the edge of the field, assuming **620** has been found to lie as shown near the corresponding edge of the field of camera **600**.

Two or more cameras may be used for stereo image analysis including object range and orientation data as discussed in FIGS. **1** and **6**. Range can also be determined via triangulation with a single camera and one target if projected on to the object in question at an angle to the camera axis from a laser say, or by using a single camera and 3 or more points on an object whose relative relationship is known (including the case of a line of points and an external point).

FIG. **7**

As stated above, the TV camera of the invention can be used to see either natural or artificial features of objects. The former are just the object features, not those provided on the object especially for the purpose of enhancing the ability to determine the object location or other variable using computer analysis of TV camera images. Such natural features, as has been pointed out in many of the co-pending referenced applications, can be holes, corners, edges, indentations, protrusions, and the like of fingers, heads, objects held in the hand, or whatever.

But using simple inexpensive equipment it is often hard to determine the presence or location of such features in a rapid reliable enough manner to insure function of the application in question. In this case, one can employ one or more artificial features, provided on the object by attaching a artificial target onto the object, or manufacturing the object with such a target.

At least three types of artificial features can be employed.

1. The first is to provide special features required for object location, or orientation determination. Such a special feature can be of an optically contrasting material at the wavelength used to that of the object, for example a bright color, or a retroreflector;

2. The second is to provide one artificial feature (typically capable of more easily being found in an image than natural features of the object), and by finding it, localize to the region of that target environs the problem of finding any other features needed nearby; and

3. The third is to find an artificial feature on an object that actually by its shape, location, or coded features, provides a guide to the location of natural or other artificial features which are to be sensed in order to determine position or orientation of the same or related objects. This has been dubbed by me a co-target in co-pending applications incorporated by reference.

As shown in FIG. **7**, object **700** has co-target **701** at one end, visible to camera **705**. The co-target in this particular instance is a diamond shape, and is of high contrast for easy acquisition. For example it could be a yellow plastic retroreflector formed of molded corner cubes similar to those used on cars for taillights and other safety purposes.

The diamond shape in this case is significant for two reasons. First it is unusual relative to the object or background when used in the context intended, and makes the target still more identifiable (that is novel color, shape and brightness are all present). In addition, in this particular instance it has been chosen that a diamond shape, should indicate that the corners of the object are to be used for 6 axis position and orientation determination and that the choice of color for example, signifies that the object corners are within some predetermined distance from the target. If desired the target location on the object can also point to the corners. For example, in the drawing, the four corners of the diamond, **720-723**, point in the general direction of the four corners **730-733** of the rectangular object **700**.

FIG. **8**

The invention herein and disclosed in portions of other copending applications noted above, comprehends a combination of one or more TV cameras (or other suitable electro-optical sensors) and a computer to provide various position and orientation related functions of use. It also comprehends the combination of these functions with the basic task of generating, storing and/or transmitting a TV image of the scene acquired either in two or three dimensions.

FIG. **8**A illustrates control of functions with the invention, using a handheld device which itself has functions (for example, a cell phone). The purpose is to add functionality to the device, without complicating its base function, and/or alternatively add a method to interact with the device to achieve other purposes.

The basic idea here is that a device which one holds in ones hand for use in its own right, can also be used with the invention herein to perform a control function by determining its position, orientation, pointing direction or other variable with respect to one or more external objects, using an optical sensing apparatus such as a TV camera located externally to sense the handheld device, or with a camera located in the handheld device, to sense datums or other information external for example to the device.

This can have important safety and convenience aspects to it, particularly when the device is used while driving a car or operating other machinery. To date voice recognition has been the only alternative to keying data in to small handheld devices, and voice is limited in many cases very limited if some physical movement is desired of the thing being communicated with.

A cellular phone **800** held in the hand of a user can be used to also signal functions in a car using a projected laser spot from built in laser spot projector **801** as in FIG. **14**, in this case detected by detector **802** on the dashboard **803**. Alternatively and or in conjunction, one may use features such as round dot targets **805-807** on the cell phone which are sensed, for example, by a TV camera **815** located in the car headliner **816** or alternatively for example in the dashboard (in this case the targets would be on the opposite end of the cell phone). More than one set of targets can be used, indeed for most generality, they would be an all sides which point in any direction where a camera could be located to look at them.

Remote control units and dictating units are also everyday examples of some devices of this type which can serve control purposes according to the invention. One of the advantages here is that it keeps the number of switches etc on the device proper to a minimum, while allowing a multitude of added functions, also in noisy environments where voice recognition could be difficult or undesirable for other reasons.

Use of specialized target datums or natural features of devices held in the hand, or used with cameras on such devices, allows photogrammetric techniques such as described in FIG. **1** to be used to determine the location in 6 degrees of freedom of the device with respect to external objects.

As one illustrative example, to signal a fax unit **824** in the car to print data coming through on the phone, the user just points (as illustrated in position **2**) the cell phone toward the fax, and the TV camera **815** scans the images of targets **805-807** on the face toward the camera, and the computer **830** connected to the camera analyzes the target images (including successive images if motion in a direction for example is used as an indicator, rather than pointing angle for example), determines the cell phone position and/or orientation or motion and commands the fax to print if such is signaled by the cell phone position orientation or motion chosen. The knowledge in space of the cell phone location and its pointing direction (and motion as pointed out above) provides information as to the fact that the fax was the intended target of the effort. Such data can be taught to the system, after the fact even if the fax or any other item desired to be controlled is added later.

Another version has a camera and requisite computer (and or transmission capability to an external computer) in the handheld device, such as a cell phone or whatever. When pointed at an object, the camera can acquire the image of the object and/or any natural features or special datums on the object which are needed to perform the function desired.

One function is just to acquire an image for transmission via for example the cell phones own connection. This is illustrated in FIG. **8**B, where an image of object **849** acquired by camera **850** of cell phone **851** held by user **852** is transmitted over mobile phone link **853** to a remote location and displayed, for example. While this image can be of the user, or someone or something of interest, for example a house, if a real estate agent is making the call, it is also possible to acquire features of an object and use it to determine something.

For example, one purpose is recognition, for example one can point at the object, and let the computer recognize what it is from its TV image. Or point around in space taking multiple TV frames aiming in different directions, and when computer recognition of a desired object in one of the images takes place, transmit certain data to the object. Or it can be used to acquire and transmit to remote locations, only that data from recognized objects.

Thus the invention can provided on a hand held object for a variety of purposes,

To take images of things;

To determine datums on things; and

To automatically read things.

The combination of any or all of these functions in addition with other object functions such as hand held cell phones, dictation units, telephones, wearable computer devices and the like.

An alternative, shown with phantom lines in FIG. **8**A, to the some aspects of the above described operation of the embodiment is to use a laser pointer **801** in for example a cell phone to designate say the fax machine as shown. Then the TV camera **815** simply detects the presence of the laser pointer projected spot **820** on the fax, and via computer memory it is known that this is a device to be energized or connected in connection with the cell phone.

The camera located in a handheld device can also be used to point at a TV screen, such as that on the dashboard of a car, and to utilize data presented there for some purpose. For example, if pointed at a screen saying email message number **5**, the camera of the device can be used to obtain this image, recognize it through known character recognition techniques, and process it for transmission if desired. Or it might just say the message to the user of the phone through the speaker of the cell phone. Such a technique is not required if means exist to directly transmit the incoming information to the cell phone, but this may not be possible.

FIG. **9**

FIG. **9** illustrates pointing at a displayed image of an object represented on a screen using a finger or laser pointer, and then manipulating the represented object or a portion thereof using the invention. For example, consider user **901** pointing a laser pointer **905** at an image generated by computer **910** on display **912**, typically a large screen display (e.g., 5 feet diagonal or more) where control features here disclosed are of most value.

The user with the pointer, can point to an image or portion of the displayed image to be controlled, and then using the action of the pointer move the controlling portion of the image, for example a "virtual" slider control **930** projected on the screen whose lever **935** can be moved from left to right, to allow computer **910** sensing the image (for example by virtue of TV camera **940** looking at the screen as disclosed in copending applications) to make the appropriate change, for example in the heat in a room.

Alternatively one can also point at the object using ones fingers and using other aspects of the invention sense the motions of ones fingers with respect to the virtually displayed images on the screen, such as turning of a knob, moving of a slider, throwing a switch etc.

Such controls are not totally physical, as you don't feel the knob, so to speak. But they are not totally virtual either, as you turn it or other wise actuate the control just as if it was physical. For maximum effect, the computer should update the display as you make the move, so that you at least get visual feedback of the knob turning. You could also get an appropriate sound if desired, for example from speaker **950**, like an increase in pitch of the sound as the knob is "moved" clockwise.

FIG. **10**

The above control aspects can in some forms be used in a car as well even with a small display, or in some cases without the display.

Or it can be a real knob which is sensed, for example by determining position of a target on a steering wheel or the fingers turning it tracked (as disclosed in co-pending application references).

For example, consider car steering wheel rim **1000** in FIG. **10**A. In particular, consider hinged targeted switch, **1010** (likely in a cluster of several switches) on or near the top of the wheel, when the car is pointed straight ahead, and actuated by the thumb of the driver **1011**. A camera **1020** located in the headliner **1025**, and read out by microcomputer **1025** senses representative target **1030** on switch **1010**, when the switch is moved to an up position exposing the target to the camera (or one could cover the target with ones fingers, and when you take a finger off, it is exposed, or conversely one can cover the target to actuate the action).

The camera senses that target **1010** is desired to be signaled and accordingly computer **1025** assures this function, such as turning on the radio. As long as the switch stays in the position, the radio is on. However other forms of control can be used where the switch and target snap back to an original position, and the next actuation, turns the radio off. And too, the time the switch is actuated can indicate a function, such as increasing the volume of the radio until one lets off the switch, and the target is sensed to have swung back to its original position and the increase in volume thus terminated.

In operating the invention in this manner, one can see position, velocity, orientation, excursion, or any other attribute of actuation desired. Because of the very low cost involved in incremental additions of functions, all kinds of things not normally sensed can be economically provided. For example the position of a datum **1040** on manually or alternatively automatically movable plastic air outlet **1041** in the dashboard **1042** can be sensed, indicative of the direction of airflow. The computer **1025** can combine this with other data concerning driver or passenger wishes, other outlets, air temperature and the like, to perfect control of the ambiance of the car interior.

It is also noted that the same TV camera used to sense switch positions, wheel position, duct position, seat position (for example using datum **1045**), head rest position (for example using datum **1046**), and a variety of other aspects of physical positions or motions of both the car controls and the driver or passengers. And it can do this without wires or other complicating devices such as rotary encoders which otherwise add to the service complexity and cost.

When the camera is located as shown, it can also see other things of interest on the dashboard and indeed the human driver himself, for example his head **1048**. This latter aspect has significance in that it can be used to determine numerous aspects such as:

1. The identity of the driver. For example, if a certain band of height isn't reached, such as point P on the drivers head, the ignition can be interlocked. Much simpler than face recognition, but effective if properly interlocked to prevent repeated retries in a short time period.

2. The position of the head of the driver in case of an accident. As detailed in reference **4**, a camera or cameras can be used to determine head location, and indeed location of the upper torso if the field of view is large enough. This information can be used to control airbag deployment, or head rest position prior to or during an accident (noting too that headrest position can also be monitored without adding any hardware). Particularly of interest is that the pixel addressing camera of the invention can have the frequency response to be useful in a crash, sensing the movement of the person (particularly severe if unrestrained) within a millisecond or two, and providing a measure of the position for airbag deployment. Additional cameras may also be used to aid the determination, by providing other views or observing other features, for example.

Using a pixel addressing camera for camera **1020** confers additional advantages. For example consider the image of the car interior produced by the camera lens **1021**, on matrix of pixels **1061**, whose addressing and processing is controlled by computer **1025**. In the first instance one can confine the window of view of a certain group of pixels of the total matrix **1061** to be only in the region of the steering wheel, as in window **1065** shown. This allows much faster readout of the more limited number of pixels, and thus of the steering wheel switches, at the expense of not seeing anywhere else in that particular reading. But this may be desirable in some cases, since it may only be required to scan for heater controls or seat positions, every 10 seconds say, while scanning for other more immediate items a hundred times per second or more. A good example are safety related functions. 5 per second might suffice for seeing where the turn signal or windshield washer control was, as an example. Window **1066** dotted lines is illustrative of a window specialized for head, headrest and seat positions, say.

Scans in certain areas of the image can also depend on information obtained. For example one may initiate a scan of a control position, based on the increasing or decreasing frequency of an event occurrence. For example if the persons head is in a different location for a significant number of scans made at 15 second intervals for example, then in case of a crash, this data could be considered unreliable. Thus the camera window corresponding to pixels in the zone of the head location **1048** could be scanned more frequently henceforward, either until the car stopped, or until such action settled down for example. Such action is often the case of a person listening to rock music, for example.

Similarly, if someone is detected operating the heater controls, a scan of predominately heater function controls and related zones like air outlets can be initiated. Thus while normal polling of heater controls might be every 2 seconds say, once action is detected, polling can increase in the window(s) in question to 40 times per second for example. The detection of action can be made first via the camera, or via input from some other input device such as a convention heater knob and electric circuit operable therewith.

Scans in certain areas of the image can also depend on information obtained in other areas of scan, or be initiated by other control actions or by voice. For example, if hard deacceleration was detected by an accelerometer, but before a crash occurred, the camera could immediately be commanded to begin scanning as fast as possible in the region of the image occupied by the driver and/or any other humans in its field of view. This would be for the purpose of monitoring movements in a crash, if a crash came, in order to deploy an airbag for example.

One might utilize the invention to actuate a function, based on positions of people or other objects in the vehicle. As one example, suppose the drivers hand is resting on a console mounted gear lever. By scanning the image of this region, one can determine from the image the position of the console shift lever, and use the image thereof to control gear change via computer **1025**. However if the driver rests his hands on the windshield wiper stalk, it could in the same manner, become a column mounted gear lever so to speak. Or just be used for up down gear changes, like a paddle shifter on a racing car. In fact in the latter sense, the camera could be instructed to detect ones finger or hand movement to do this function for example, wherever one desired to rest ones hand (within the camera field of view at least). This function is also useful for physically disabled persons wishing to drive the car. And it can be different for different persons as well, via programming of the control functions associated with any given hand, switch or other position or movement.

FIG. **10**B illustrates alternative types of control mechanisms which can be used with the invention, in this case illustrated on the steering wheel of a car, although as can be appreciated, any suitable function or location may be used or created. And too, combinations of functions can be used. The invention is generic to car steering wheel controls, dishwashers, audio systems in ones home, heating and air conditioning elements and virtually all other forms of human related control functions. The key is that the camera computer combination makes a very inexpensive way to share a wide variety of functions with one or just a few basic systems and over a large population base.

As shown in FIG. **10**B, the steering wheel **1070** has two additional types of controls visible to camera **1020** and able to be sensed and generate the appropriate control function via computer. These are rotating device **1072** built to rotate around the steering wheel rim circular cross section, and expose a continuously variable, or digital or step wise increment component to the camera. For example, three bars are shown, short **1075**, medium **1076**, and long **1077**. The computer senses which of the three is visible by comparing the length to pre-stored values (or taught values, see below), and causes the desired action to occur.

The second control **1080** is a sliding device **1081** which can be slid clockwise, or counterclockwise along a circumferential section of the steering wheel at the top, sides or whereever. As before, Its position is determined by camera **1020** again providing more data than just a switch up or down as shown before.

While illustrated on the steering wheel where it is readily at hand, it can be appreciated that the position of either the slider **1081** or the rotary device **1072**, or other similar devices for the purpose at hand could be elsewhere than the wheel, for example on stalk or on a piece of the dash, or other interior component indeed wherever a camera of the invention can view them without excessive obscuration by persons or things in the car. It need not be on a car either, controls of this type can be in the home or elsewhere. Indeed a viewable control datum can even be on a portable component such as ones key chain, phone, or article of clothing apparel, or whatever. Similarly the camera **1020** can view these items for other purposes as well.

The teach-ability of the invention is achieved by showing the camera the code marker in question (e.g., a short bar located on the wheel), and in the computer recording this data along with what it is supposed to signify as a control function for example, turn rear wiper on to first setting. This added functionality of being easily changed after manufacture is an important advantage in some cases, as for example, today after-market addition of wired in accessories is difficult.

Games Using the Invention

The co-pending referenced applications have described games which can be played with target sensing and touch screen based devices, typically but not necessarily, electro-optically based (e.g., TV camera). The cameras of the invention can be used to, for example: Sense the player or players in the game or portions thereof; sense objects held or manipulated by the players (e.g., a ball, a pistol); sense physical tokens used in the game, such as monopoly game tokens; and sense game accessories such as checkerboards, croquet wickets; compare positions of objects with respect to other objects or players.

In addition, the cameras can be used to take images which can be displayed also a major feature given the ability to create life size displays. And the computer of the invention can be used to control the presentation of background image data from stored images, or even images downloaded from the internet for example.

Some or all of these aspects will now be illustrated in some representative game illustrations (again noting that some more are in the co-pending applications).

FIG. **11** Board Game

Even today, popular board games such as Monopoly and the like are being provided in computer playable form, with the "board" represented on the screen of the computer monitor. The invention here builds on this by providing various added features which allow a physical nature of the game just as the real game, but with new aspects and providing physical game play which can be transmitted over the internet to others. These features also can be turned off or on at as desired.

In one version shown in FIG. **11**A, the player tokens such as **1101** and **1102** are observed by camera of the invention **1110** placed directly overhead of the play board **1115**, which can for example be a traditional monopoly board (chess board, checker board, etc). points on the board such as corners **1130**, **1131**, **1132**, and **1133** can also be observed to establish a reference coordinate system for the computer **1140** to track the moves of the markers, either from their natural features, or from specialized datums thereon (e.g., retro-reflective hat top **1141** on marker **1101**). For example a train shape **1102** of a marker can be called from memory, or taught to the computer by showing it to the camera. Rotation invariant image analysis programs such as the PATMAX program from Cognex company can be used to identify the marker in any normal orientation, together with its location on the board (the board itself can be taught to the computer using the camera, but is preferably called up from memory).

The board position and relative scale in the field of view is determined easily by knowing the spacing of the corner points **1130-1133** and using this to calibrate the camera (to provide extra contrast, the corners can have retro-reflective glass bead edging or beading as shown). For example if the points are spaced 20 inch on corners of the board, and the camera is positioned so that 20 inches occupies 80% of its field of view, then the field of view is 25 inches square (for a square matrix of camera pixels), and each pixel of 1000 pixels square, occupies 0.025 inches in the object field.

The play of both players (and others as desired) can be displayed on the monitor **1150**, along with an image of the board (which also can be called from computer memory). But other displays can be provided as well. For example to lend more realism to the game, the display (and if desired sound from speaker **1155** connected to computer **1140**) can also be programmed to show an image or sound that corresponds to the game. For example, when the camera image has provided information that one player has landed on "Boardwalk" (the most valuable property) a big building could be caused to be shown on the screen, corresponding to it also suitable sounds like wow or something provided).

The camera can be used to see monopoly money (or other game accessories) as well, and to provide input so the computer can count it or do whatever.

A large, wall sized for example, screen can add added realism, by allowing one to actually get the feeling of being inside the property purchased, for example.

One of the exciting aspects of this game is that it can be used to turn an existing board game into something different. For example, in the original monopoly the streets are named after those in Atlantic City. By using the computer, and say a DVD disc such as **1160** stored images of any city desired can be displayed, together with sounds. For example, one could land on the Gritti Palace Hotel in Venice, instead of Boardwalk. As shown in FIG. **11**B, the TV camera senses the image of train marker **1101**, and conveys this information to computer **1140**, which causes the display **1150** and speaker of the invention to display the information desired by the program in use.

Making the game in software in this way, allows one to bring it home to any city desired. This is true of a pure (virtual) computer game as well, where the board only exists on the computer screen.

For added fun, for example in a small town context, local stores and properties could be used, together with local images, local personages appearing on the screen hawking them, and the like. A local bank could be displayed to take your money, (even with sounds of the local banker, or their jingle from the radio) etc. This makes the game much more local and interesting for many people. Given the ease of creating such local imagery and sounds with cameras such as digital camcorder **1151** used as an input of display imagery (e.g., from local celebrity **1158**) to the game program, one can make any monopoly experience more interesting and fun at low cost.

The same holds true with other well known games, such as Clue, where local homes could be the mystery solving location, for example. One can also create games to order, by laying out ones own board. If one of the persons is remote, their move can be displayed on the screen **1150**.

In the above, the display has been treated as sort of backdrop or illustration related. However, one can also create a whole new class of games in which the display and/or computer and the board are intertwined. For example as one takes a trip around the monopoly board, several chance related drawings opportunities occur during play. In this new game, such could be internet addresses one draws, which, via modem **1152**, send the board game computer **1140** to any of a large number of potential internet sites where new experiences await, and are displayed in sight and sound on the display.

It should also be noted that the board can be displayed on the screen as well, or alternatively projected on a wall or table (from overhead). A particularly neat mixture of new and old is shown in FIG. **11**B, where the board is displayed on a screen pointed vertically upward just as it would be on a table, and indeed in this case physically resident on a table **1165**. The board is displayed (from software images or cad models of the board in computer **1166**) on a high resolution table top HDTV LCD screen **1167** with a suitable protective plastic shield (not shown for clarity). Play can proceed just as before using physical tokens such as **1101** and **1102**. In this case the display used to augment the game can actually be shown on the same screen as the board, if desired.

The TV camera **1110** in this context is used to see the tokens and any other objects of the game, the people as desired, and the play, as desired. The camera can be used to see the display screen, but the data concerning the board configuration displayed may be best imputed to the computer program from direct data used to create the display.

A beauty of the invention is that it allows the interaction of both computer generated images and simulations, with the play using normal objects, such as one might be accustomed to for example, or which give a "real" feel, or experience to the game.

FIG. **12** Sports Game

FIG. **12** illustrates a generic physical game of the invention using points such as **1201-1205** on the human (or humans) **1210** sensed by a TV camera such as stereo camera pair **1215** and transmitted to the computer of the invention **1220**. While points can be sensed in 2D, this illustration uses as stereo camera pair located on large screen display **1225** as shown to provide a unitary package built into the screen display (pointed out in other co-pending applications). In this particular instance a 3D display is illustrated, though this isn't necessary to obtain value and a good gaming experience. The human optionally wears red and green filter glasses **1235** such that red images on the screen are transmitted to one eye, green to another, so as to provide a 3D effect. Similarly crossed polarized filter glasses (with appropriate display), and any other sort of stereoscopic, or autosteroscopic method can also be used, but the one illustrated is simple, requires no connecting wires to the human, and can be viewed by multiple uses, say in a gym aerobics room.

The game is generic, in that it totally depends on the program of the computer. For example, it can be an exercise game, in which one walks on a treadmill **1250**, but the image displayed on screen **1225** and sound from speakers **1255** and **1266** carry one through a Bavarian forest or the streets of New York as one walks, for example.

Or it can be a parasail game in which one flies over the water near Wakiki beach, with suitable images and sounds. In any case action determined by sensing position, velocity acceleration, or orientation of points **1201-1206** on the player, **1210** is converted by computer **1220** into commands for the display and sound system. Note in the figure this player is shown viewing the same screen as the treadmill walker. This has been shown for illustration purposes, and it is unlikely the same game could be applied to both, but it is possible.

It is noted that fast sensing, such as provided by the pixel addressing camera method disclosed above is highly desirable to allow realistic responses to be generated. This is especially true where velocities or accelerations need to be calculated from the point position data present in the image (and in comparison to previous images).

For example, consider points **1201** and **1202** on player **1210**. If point **1201** moves to **1201***a*, and **1202** moves to **1202***a* indicative of a quick jerk movement to turn the displayed parasail, this movement could occur in a 0.1 second. But the individual point movements to trace the action would have to be sensed in 0.01 second or quicker for example to even approximately determine the acceleration and thus force exerted on the glider, to cause it to move.

It is important to note that the invention is not only generic in so far as the variety of these games are concerned, but it also achieves the above with virtually no mechanical devices requiring maintenance and creating reliability problems which can eliminate profits from arcade type businesses especially with ever more sophistication required of the games themselves.

FIG. **13** Bar Game

FIG. **13** illustrates a game which is in a class of gesture based games, in which the flirting game of FIG. **15** is also an example. In such games one senses the position, velocity or acceleration of a part of a person, or an object associated with the person. This can also include a sequence of positions, itself constituting the gesture. The detected data is then related to some goal of the contest. Consider FIG. **13**, wherein the object in ones hand is monitored using the invention, and a score or other result is determined based on the position, velocity, orientation or other variable of the object determined. For example, in a bar one can monitor the position, orientation, and rate of change thereof of drinking glasses.

A two person game is illustrated, but any reasonable number can play as long as the targets can all be tracked sufficiently for the game (in one test over 200 targets were acquired, but as can be appreciated this uses most of the field of view of the camera, and thus speed improvements made possible by pixel addressing become more difficult.

As shown, a single camera **1301** observes one or more targets such as **1305** on glass **1310** held by contestant **1315**, and target **1320** on glass **1325** of contestant **1330**. On a signal, each drinks, and a score is calculated by program resident in computer **1350** based on the time taken to raise the glass, and place it back empty on table **1355**. A display of the score, and an image desired, for example of the winner (taken with camera **1301** or another camera), or a funny image called from computer memory, is displayed on monitor display **1370**.

If the glass features are sufficiently distinct for reliable and rapid acquisition and tracking, for example as might be provided by an orange color, or a distinct shape, then specialized target features are not required.

Alternatively the velocity, path of movement of the glass (or other object), acceleration, or any other variable from which target data is sufficient to calculate, can be used to determine a score or other information to be presented or used.

FIG. **14**

The referenced co-pending applications have described a game where by laser pointers can be used to designate images on a TV screen. In this case of FIG. **14**A, the TV camera of the invention such as **1410** is used in a two player game to see laser pointer spots such as **1420** and **1421** projected by players **1430** and **1431** respectively, using laser pointers **1440** and **1441** respectively. When one player's spot hits the other, the event is recorded in memory of computer **1450** for further analysis and display.

In a somewhat different context, a person can use a laser pointer to point at an object to designate it for some purpose, for example for action. For example consider FIG. **14**B, in which housewife **1460** who points with laser pointer **1462** so as to provide a laser spot **1465** on dishwasher **1470**. TV camera of the invention **1475** in corner of the kitchen **1480** picks up all laser spots in an image of the room (made easier to process in terms of signal to background imagery if one locates a laser wavelength band-pass interference filter **1481** in front of the TV camera as shown) and compares via computer **1483**, the location of the spot detected in the image to stored memory locations of objects such as the dishwasher **1470** or fridge **1485** in the camera field of view, so as to identify the object needing action. In this case too, housewife may signal via a spatially variant laser pointer projection image (see copending referenced applications for further examples in other applications), or a series of spots in time, what action is desired, for example to turn the washer on. In this case the computer **1483** can cause a command to do so to be sent to the washer.

Any one with a simple laser pointer can make these commands effective. No learning is needed just point at the item desired, with the TV camera and computer of the invention acquiring the data and interpreting it. This is much simpler than remote controls of today, and a major advantage for those who have difficulty or inclination to learn complex electronic devices and procedures. It should be noted that these pointing procedures can easily be combined with voice recognition to further define the desired control activity for example inputting the housewife's voice in this example by virtue of microphone **1476**.

The stored locations can be taught. For example in a setup mode, one can point a laser pointer at the dishwasher, and indicate to the computer that that spot is the dishwasher. The indication can be provided by keyboard, voice recognition or any other means that is satisfactory.

Clearly other items can be monitored or controlled in this manner. The camera can also detect optical indications provided by other means, for example lights in the appliance itself. And one can detect whether light have been left on at night (or not left on) and cause them to be turned off or on as desired.

Such a camera if it is responsive to normal illumination as well as that of the laser wavelength, can also be used to see movements and locations of people. For example, it can look at the top of the stove, and assure that no movement is near the stove **1486**, or objects on it if programmed to do so, thus sounding an alarm if an infant should get near the stove, for example.

The housewife in the kitchen can also point at a board on which preprogrammed actions are represented. For example consider board **1490**, shown in greater detail in FIG. **14**C, in which 3 squares **1491**-**1493** are to represent different functions. Thus if **1491** is programmed (via keyboard, voice or whatever) to represent turning on the clothes dryer in the laundry, when the TV camera sees, and via the computer, identifies spot **1496** projected by the user on square **1491**, it causes the dryer to turn on. Operated in this manner, the board **1490**, in combination with a TV camera of the invention (such as **1475** or a more dedicated one for the board alone) and computer such as **1483** can be considered a form of touch screen, where the user, in this case in the kitchen can point at a portion of the board with a finger, or a laser pointer, and register a choice, much like touching an icon on a conventional computer touch screen.

Similarly, squares or other zones representing choices or the like can be on the item itself. For example, a stove can have four areas on its front, which can be pointed at individually for control purposes, what ever they are (e.g., representing heat settings, burner locations or the like). For security, it could be that only a coded sequence of laser pulses would be seen, or as pointed out in co-pending reference Ser. No. 60/133,673, a spatial code, for example representing the user such as an initial could be projected, and sensed on the object by the TV camera.

The laser pointer can be held in the hand of the user, or, like **1497** attached for example to a finger, such as forefinger **1498**. Or it can be on or in another object, desirably one which is often hand held in the normal course of work, such as a TV remote control, a large spoon, or the like. Or using other aspects of the invention, the finger of the user can be observed to point directly, and the object being pointed at determined. For example if finger **1498** is moved 4 times, it could indicate to the TV camera and thence computer that channel four was desired on a TV display not shown.

If a special pointer is used, it can be any workable optical device, not necessarily a laser. The camera and computer of the invention can also be used to observe the user pointing directly, and compute the pointing vector, as has been described in my co-pending applications.

FIG. **15** A "Flirting" Game

Another game type is where the camera looks at the human, and the humans expressions are used in the game. In this case it is facial expressions, hand or body gestures that are the thing most used.

For example, one idea is to have a scene in a restaurant displayed on a display screen **1500**, preferably a large HDTV screen or wall projection to be as lifelike as possible, and preferably life size as well which lends extra realism to some games, such as this one due to the human element involved.

Let us consider that seated at the table in the restaurant displayed on the screen is a handsome man **1501** whose picture (likely a 3D rendered animation, or alternatively photo-imagery called from memory), and the goal for the girl **1510** playing the game is to flirt with this man until he gets up and comes over to say hello, ask her out or what ever (what he does, could be a function of the score obtained, even!).

Player **1510** seated at table **1511** (for authenticity, for example) is observed by TV camera **1515** (or stereo pair as desired, depending whether 3D information is thought required) and computer of the invention **1520**, which through software determines the position of eyebrows, lips, hands, fingers and any other features needed for the game. If necessary, specialized targets can be used as disclosed herein and elsewhere to augment this discrimination, for example such as optically contrasting nail polish, lipstick, eyeliner or other. Contrast can be in a color sense, or in a reflectivity sense such as even retro-reflective materials such as Scotchlite 7615 by 3M company. Even special targets can be used to enhance expressions if desired.

This can be a fun type game, as the response of the displayed person can be all kinds of things even contrary to the actual gestures if desired. Sounds, such as from speaker **1530** can also be added. And voice recognition of players words sensed by microphone **1550** can also be used, if verbal as well as expressive flirting is used.

While the game here has been illustrated in a popular flirting context, it is more generally described as a gesture based game. It can also be done with another contestant acting as the other player. And For example, the contestants can be spaced by the communication medium of the internet. The displayed characters on the screen (of the other player) can be real, or representations whose expressions and movements change due to sensed data from the player, transmitted in vector or other form to minimize communication bandwidth if desired.

Other games of interest might be:

"Down on the Farm" in which a farmer with live animals is displayed on a life size screen, and the children playing the game are to help the farmer by calling the animals to come over to them. This would use recognition of voice and gesture to make the animal images move and make sounds.

A player can find someone in a display and point at him, like the "Whereas Waldo" puzzle game. Then the subject moves, child runs to peek at him, and to find him, say running down a street whose image is displayed on the screen.

One can also use the camera of the invention to monitor the progress made by a child building blocks, and show an Video displayed image of a real skyscraper progressing as he builds his little version. Note the benefit of group activity like a board game and children's play with each other.

FIG. **16**

FIG. **16** illustrates a version of the pixel addressing camera technique wherein two lines on either side of a 1000 element square array are designated as perimeter fence lines to initiate tracking or other action.

Some "pixel addressing" cameras such as the IVP MAPP 2500 512.times.512 element camera, are smart, that is can process on the same chip. However, in some cases the control of such a camera may not allow one to actually read just one pixel, say, but rather one must read the whole line on which the pixel rests. Now some processing can be in parallel such that no speed is lost, at least in many instances.

If however, one does have to read a whole line serially into a computer portion, then to fully see a 10.times.10 pixel round target say, one would have to read at least 10 lines.

If two targets both were located on the same lines, the time involved to read would be the same.

In the same vein, if lines of data must be scanned, then the approach of **2***b* wherein every 20.sup.th pixel say is interrogated can be specialized to having such pixels fall on scan lines wherever possible. And where one is restricted to reading all pixels on a scan line and where a target entry zone is anticipated, one can have a scan line oriented to be crossed by such entry. For example in FIG. **16**, the two lines **1601** (line of pixels **3**) and **1602** (line of pixels **997**) of a 1000.times.1000 element pixel array **1610** are designated as perimeter fence lines, to trigger a target tracking or other function on the entry of a target image on to the array, such as **1615** from either the right or left side in the drawing. This is often the case where entry from top or bottom is precluded by constraints of the application, such as a table top at the bottom, or the height of a person at the top. Or in a stereo example such as FIG. **6**, the baseline defines the direction of excursion of a target as z is varied again calling for crossing of scan lines out of the plane of the paper at some point.

The invention herein has provided an exciting method by which common board games can become more fun. The invention provides a link with that past, as well as all of the benefits of the video and computer revolution, also via the internet.

It is envisioned that the same approach may be applied to many card games as well. It is also thought that the invention will find use in creating ones own games, or in downloading from the internet others creations. For example, common everyday objects can become the tokens of the games, and taught to the game computer by presenting them to the video camera. Similarly, the people playing the game can be taught, including their names and interests.

FIG. **17**

FIG. **17** illustrates a 3D acoustic imaging embodiment of the invention which at low cost may generate accurate 3D images of the insides of objects, when used in conjunction with ultrasonic transducers and particularly a matrix array of ultrasonic transducers.

As shown in FIG. **17**A, the position in xyz of the ultrasonic imaging head **1700** on wand **1701** held in a users hand **1702** is monitored electro-optically as taught in FIG. **1**, using a single camera **1710** and a simple four dot target set **1715** on the head **1700** at the end of the transducer wand **1701** in contact with the object to be examined **1720**. Alternatively, as also taught in FIG. **1**, a stereo pair for example providing higher resolution in angle can be employed.

Computer **1725** combines ultrasonic ranging data from the ultrasound transducer head **1700** and from the sensor of transducer location (in this case performed optically by camera **1710** using the optically visible targets on the transducer head) in order to create a range image of the internal body of the object **1720** which is thus referenced accurately in space to the external coordinate system in the is case represented by the camera co-ordinates xy in the plane of the TV camera scan, and z in the optical axis of the camera.

In many cases it is also desirable to know the pointing angles of the transducer. One instance is where it is not possible to see the transducer itself due to obscuration, in which case the target may alternately be located at the end **1704** of the wand for example. Here the position and orientation of the wand is determined from the target data, and the known length of the wand to the tip is used, with the determined pointing angle in pitch and yaw (obtained from the foreshortening of the target spacings in the camera image field) to calculate the tip position in space.

This pitch and yaw determination also has another use however, and that is to determine any adjustments that need to be made in the ultrasonic transduction parameters or to the data obtained, realizing that the direction of ultrasound propagation from the transducer is also in the pointing direction. And that the variation in ultrasound response may be very dependent on the relation of this direction **1730** with respect to the normal **1735** of the surface **1736** of the object (the normal vector is shown for clarity pointing inward to the object).

The difference in direction can be calculated by using the TV camera (which could be a stereo pair for greater angular resolution) as well to determine the surface normal direction. This can, for example, be done by placing a target set such as **1740** on the surface in the field of the camera as shown. This can be dynamically or statically accomplished using the photogrammetric method described in the Pinkney references.

Differences in direction between the surface normal and the transducer pointing direction are then utilized by software in the computer **1725** of the invention in analysis of the ultrasound signals detected. The pointing angle and the position of the transducer on the surface of the object are used by the computer in predicting the location of various returns from internal points within the object, using a suitable coordinate transformation to relate them to the external coordinate reference of the TV camera.

All data, including transducer signals and wand location is fed to computer **1725** which then allows the 3D image of the inside of the body to be determined as the wand is moved around, by a human, or by a robot. This is really neat as all the images sequentially obtained in this manner can be combined in the computer to give an accurate 3D picture **1745** displayed on monitor **1750**.

In one preferred embodiment as shown in FIG. **17**C, the transducer head **1700** is comprised of a matrix **1755** of 72 individual transducer elements which send and receive ultrasound data at for example, 5 MHZ. This allows an expanded scan capability, since the sensor can be held steady at each discrete location xyz on the object surface, and a 3D image obtained with out movement of the transducer head, by analyzing the outputs of each of the transducers. Some earlier examples are described in articles such as: Richard E. Davidsen, 1996 IEEE Ultrasonics Symposium, A Multiplexed Two-Dimensional Array For Real Time Volumetric and B-Mode; Stephen W. Smith, 1995 IEEE Ultrasonics Symposium, Update On 2-D Array Transducers For Medical Ultrasound, 1995.

If the wand is now moved in space, fine scan resolution is obtained, due to the operation of the individual elements so positioned with out the need to move the wand in a fine pitch manner to all points needed for spatial resolution of this order.

This eases the operators task, if manually performed, and makes robotization of such examination much easier from a control point of view.

Consider FIG. **17**B which illustrates a transducer as just described, also with automatic compensation at each point for pointing angle, robotically positioned by robot, **1785** with respect to object **1764**. In this case a projection technique such as described in U.S. Pat. No. 5,854,491 is used to optically determine the attitude of the object surface, and the surface normal direction **1760** from the position of target set **1765** projected on the surface by diode laser set **1770**, and observed by TV Camera **1775** located typically near the working end of the robot. Differences between the normal direction and the transducer propagation direction (typically parallel to the housing of the transducer) is then used by computer **1777** to correct the data of the ultrasonic sensor **1780** whose pointing direction in space is known through the joint angle encoders and associated control system **1782** of robot **1785** holding the sensor. Alternatively the pointing direction of this sensor can be monitored by an external camera such as **1710** of FIG. **17**A.

It should be noted that the data obtained by TV camera **1775** concerning the normal to the surface and the surface range from the robot/ultrasonic sensor, can be used advantageously by the control system **1782** to position the robot and sensor with respect to the surface, in order to provide a fully automatic inspection of object **1764**. Indeed the camera sensor operating in triangulation can be used to establish the coordinates of the exterior surface of object **1764** as taught for example in U.S. Pat. No. 5,854,491, while at the same time, the acoustic sensor can determine the range to interior points which can be differentiated by their return signal time or other means. In this manner, a complete 3D map of the total object, interior and exterior, can be obtained relative to the coordinate system of the Robot, which can then be transformed to any coordinate system desired.

The invention has a myriad of applications beyond those specifically described herein. The games possible with the invention in particular are limited only by the imagination.

Consider hand held computer **1901** of FIG. **18**, incorporating a camera **1902** which can optionally be rotated about axis **1905** so as to look at the user or a portion thereof such as finger **1906**, or at objects at which it is pointed. Optionally, and often desirably, a stereo pair of cameras to further include camera **1910** can also be used. It too may rotate, as desired. Alternatively fixed cameras can be used when physical rotation is not desired, for ruggedness, ease of use, or other reasons (noting that fixed cameras have fixed fields of view, which limit versatility in some cases).

When aimed at the user, as shown, it can be used, for example, to view and obtain images of:

Ones self—facial expression etc, also for image reasons—id etc. combined effect.

Ones fingers (any or all), one finger to other and the like. This in turn allows conversing with the computer in a form of sign language which can replace the keyboard of a conventional computer.

One or more objects in ones hand. Includes a pencil or pen—and thus can be used rather than having a special touch screen and pencil if the pencil itself is tracked as disclosed in the above figure. It also allows small children to use the device, and those who cannot hold an ordinary stylus.

Ones gestures.

The camera **1902** (and **1910** if used, and if desired), can also be optionally rotated and used to view points in space ahead of the device, as shown in dotted lines **1902***a*. In this position for example it can be used for the purposes described in the previous application. It can also be used to observe or point at (using optional laser pointer **1930**) Points such as **1935** on a wall or a mounted

LCD or projection display such as **1940** on a wall or elsewhere such as on the back of an airline seat.

With this feature of the invention, there is no requirement to carry a computer display with you as with an infrared connection (not shown) such as known in the art one can also transmit all normal control information to the display control computer **1951**. As displays become ubiquitous, this makes increasing sense—other wise the displays get bigger the computers smaller trend doesn't make sense if they need to be dragged around together. As one walks into a room, one uses the display or displays in that room (which might themselves be interconnected).

The camera unit **1902** can sense the location of the display in space relative to the handheld computer, using for example the four points **1955**-**1958** on the corners of the display as references. This allows the handheld device to become an accurate pointer for objects displayed on the screen, including control icons. And it allows the objects on the screen to be sensed directly by the camera—if one does not have the capability to spatially synchronize and coordinate the display driver with the handheld computer.

The camera can also be used to see gestures of others, as well as the user, and to acquire raw video images of objects in its field.

A reverse situation also exists where the cameras can be on the wall mounted display, such as cameras **1980** and **1981** can be used to look at the handheld computer module **1901** and determine its position and orientation relative to the display.

Note that a camera such as **1902**, looking at you the user, if attached to hand held unit, always has reference frame of that unit. If one works with a screen on a wall, one can aim the handheld unit with camera at it, and determine its reference frame to the handheld unit. Also can have two cameras operating together, one looking at wall thing, other at you (as **1902** and **1902**a) in this manner, one can dynamically compare ref frames of the display to the human input means in determining display parameters. This can be done in real time, and if so one can actually wave the handheld unit around while still imputing accurate data to the display using ones fingers, objects or whatever.

Use of a laser pointer such as **1930** incorporated into the handheld unit has also been disclosed in the referenced copending applications. For example, a camera on the hand held computer unit such as **1902** viewing in direction **1902***a* would look at laser spot such as **1990** (which might or might not have come from the computers own laser pointer **1930**) on the wall display say, and recognized by color and size/shape reference to edge of screen, and to projected spots on screen.

The invention claimed is:

1. A handheld device comprising:
- a housing;
- a computer within the housing;
- a first camera oriented to view a user of the handheld device and having a first camera output; and
- a second camera oriented to view an object other than the user of the device and having a second camera output, wherein the first and second cameras include non-overlapping fields of view, and wherein the computer is adapted to perform a control function of the handheld device based on at least one of the first camera output and the second camera output.

2. The handheld device of claim **1** wherein the handheld device comprises a mobile phone.

3. The handheld device of claim **1** wherein the first camera is adapted to acquire an image of at least a portion of the user.

4. The handheld device of claim **1** wherein the second camera is adapted to acquire an image of the object.

5. The handheld device of claim **1** wherein the second camera is adapted to acquire a video of the object.

6. The handheld device of claim **1** wherein the computer is operable to determine a gesture based on at least one of the first camera output and the second camera output.

7. The handheld device of claim **1** wherein the computer is operable to determine a facial expression based on at least one of the first camera output and the second camera output.

8. The handheld device of claim **1** wherein the computer is adapted to determine at least one of the position and the orientation of the object based on the second camera output.

9. The handheld device of claim **6** wherein the gesture is performed by a person other than the user of the handheld device.

10. The handheld device of claim **1** wherein the computer is adapted to recognize the object based on the second camera output.

11. The handheld device of claim **1** wherein the computer is adapted to generate control instructions for a display that is separate from the handheld device.

12. The handheld device of claim **1** wherein the computer is adapted to determine a reference frame of the object.

13. The handheld device of claim **1** wherein the computer is adapted to perform a control function based on the first camera output and based on the second camera output.

14. The handheld device of claim **1** wherein the computer is adapted to transmit information over an internet connection.

* * * * *

EXHIBIT B

US007933431B2

(12) **United States Patent**
Pryor

(10) **Patent No.: US 7,933,431 B2**
(45) **Date of Patent: Apr. 26, 2011**

(54) **CAMERA BASED SENSING IN HANDHELD, MOBILE, GAMING, OR OTHER DEVICES**

(76) Inventor: **Timothy R. Pryor**, Tecumseh (CA)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/834,281**

(22) Filed: **Jul. 12, 2010**

(65) **Prior Publication Data**

US 2010/0277412 A1 Nov. 4, 2010

Related U.S. Application Data

(63) Continuation of application No. 11/980,710, filed on Oct. 31, 2007, now Pat. No. 7,756,297, which is a continuation of application No. 10/893,534, filed on Jul. 19, 2004, now Pat. No. 7,401,783, which is a continuation of application No. 09/612,225, filed on Jul. 7, 2000, now Pat. No. 6,766,036.

(60) Provisional application No. 60/142,777, filed on Jul. 8, 1999.

(51) **Int. Cl.**
G06K 9/00 (2006.01)

(52) **U.S. Cl.** **382/103**; 382/104; 348/207.1

(58) **Field of Classification Search** 382/103, 382/104; 348/207.1
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,878,174	A *	3/1999	Stewart et al.	382/293
6,342,917	B1 *	1/2002	Amenta	348/207.1
6,453,180	B1 *	9/2002	Endoh et al.	455/567
6,597,817	B1 *	7/2003	Silverbrook	382/289

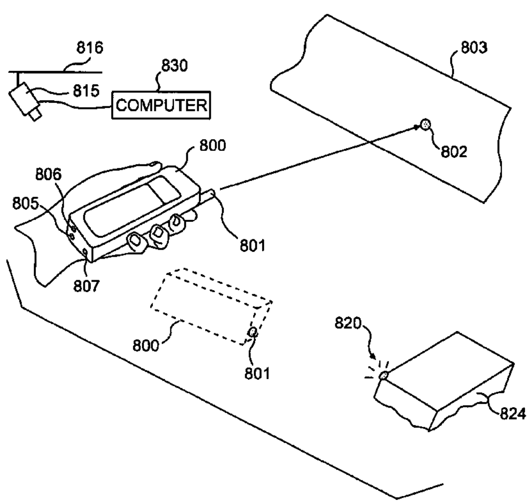
* cited by examiner

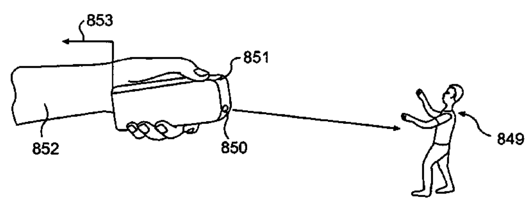
Primary Examiner — Tom Y Lu
(74) *Attorney, Agent, or Firm* — Warner Norcross & Judd LLP

(57) **ABSTRACT**

Method and apparatus are disclosed to enable rapid TV camera and computer based sensing in many practical applications, including, but not limited to, handheld devices, cars, and video games. Several unique forms of social video games are disclosed.

31 Claims, 22 Drawing Sheets





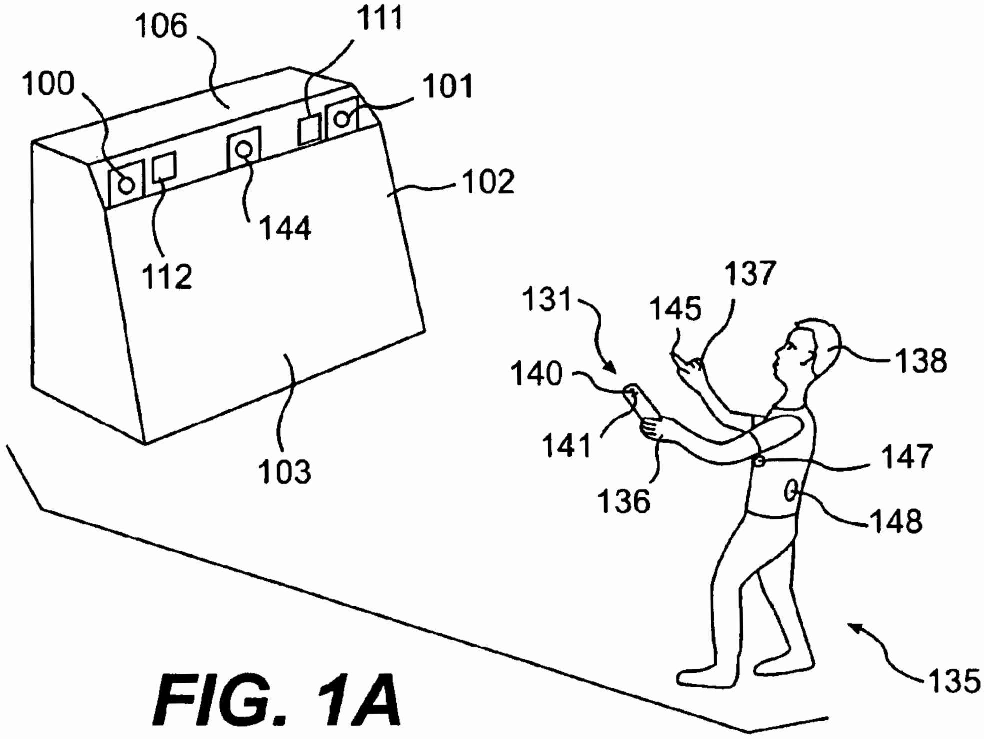


FIG. 1A

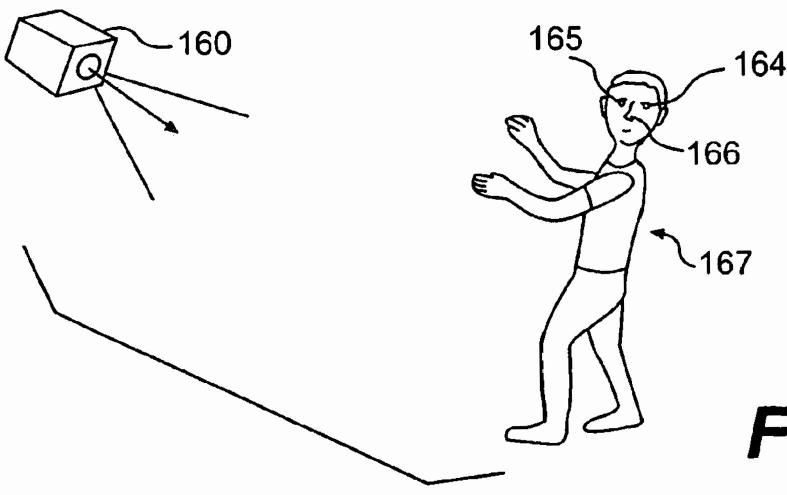


FIG. 1B

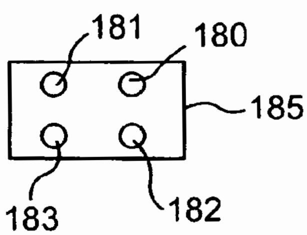


FIG. 1C

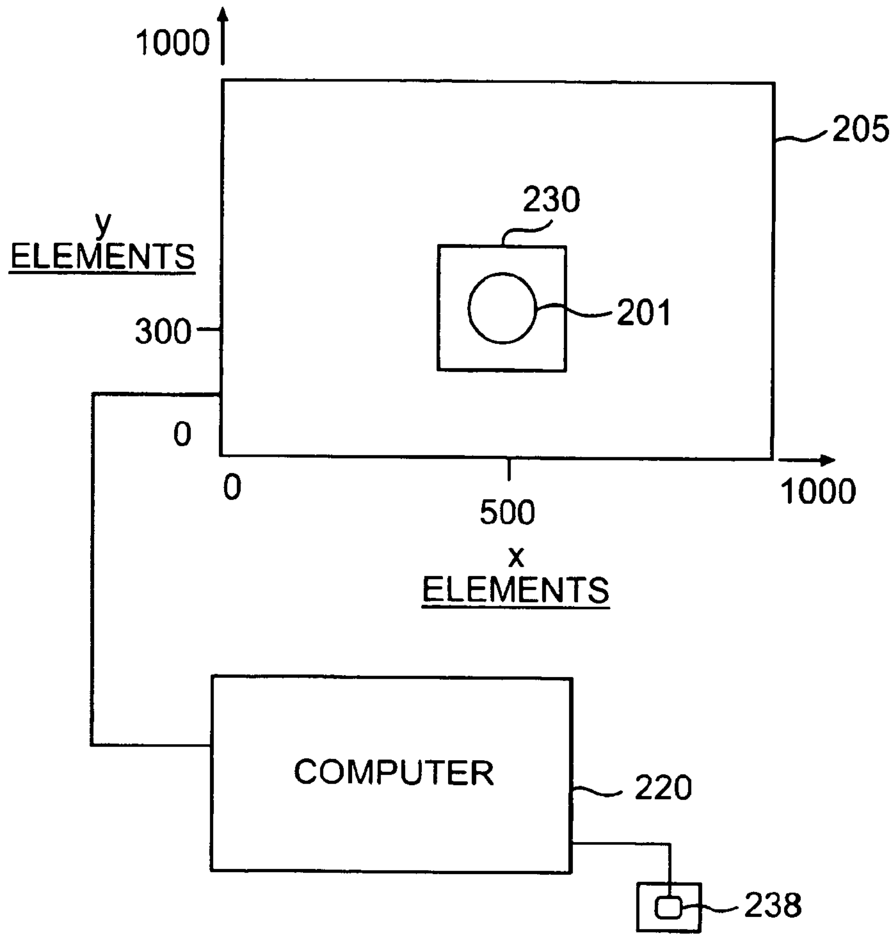

1000
205
230
y
ELEMENTS
201
300
0
0
500
1000
x
ELEMENTS
COMPUTER
220
238

FIG. 2A

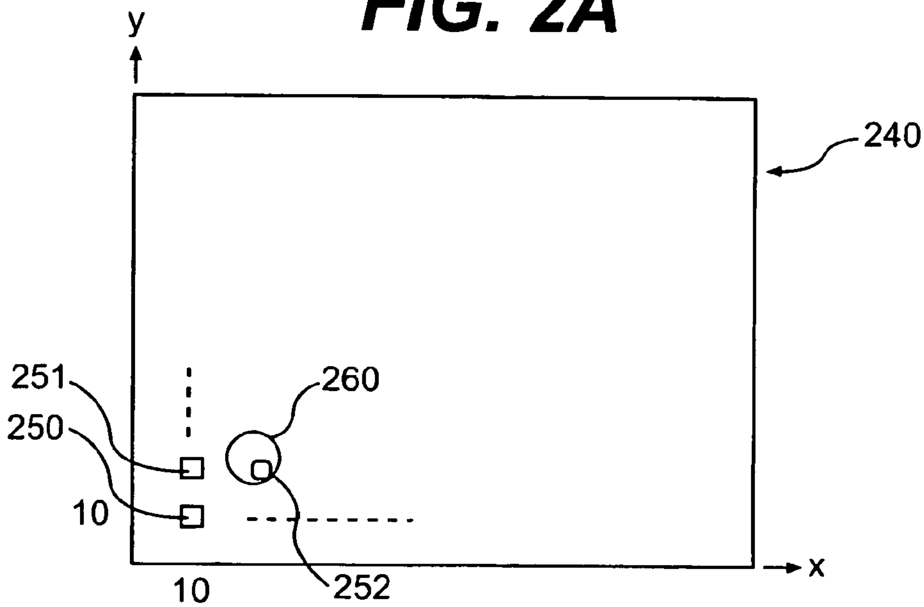

y
240
251
260
250
10
10
252
x

FIG. 2B

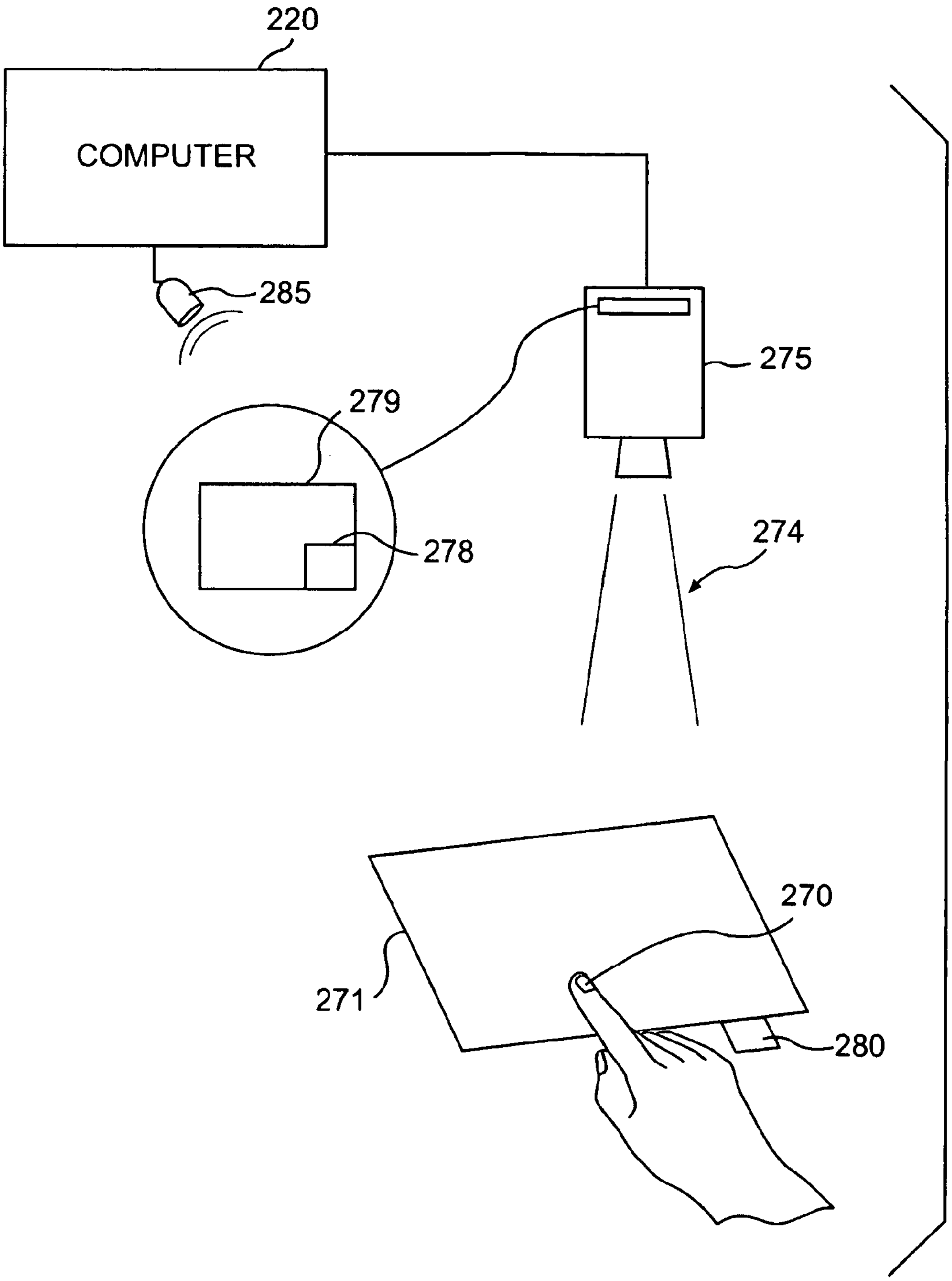


FIG. 2C

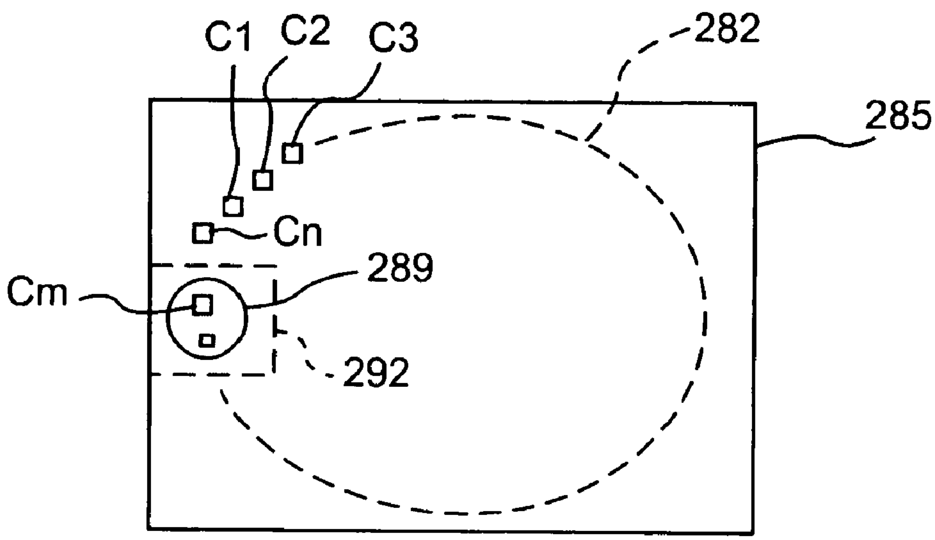


FIG. 2D

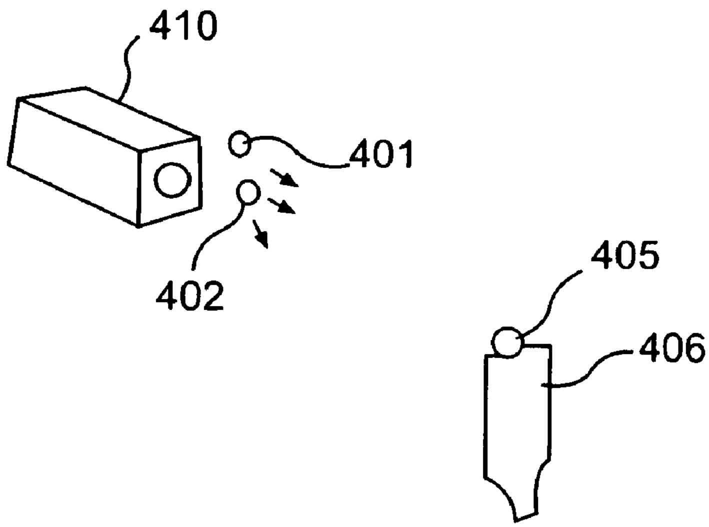


FIG. 4A

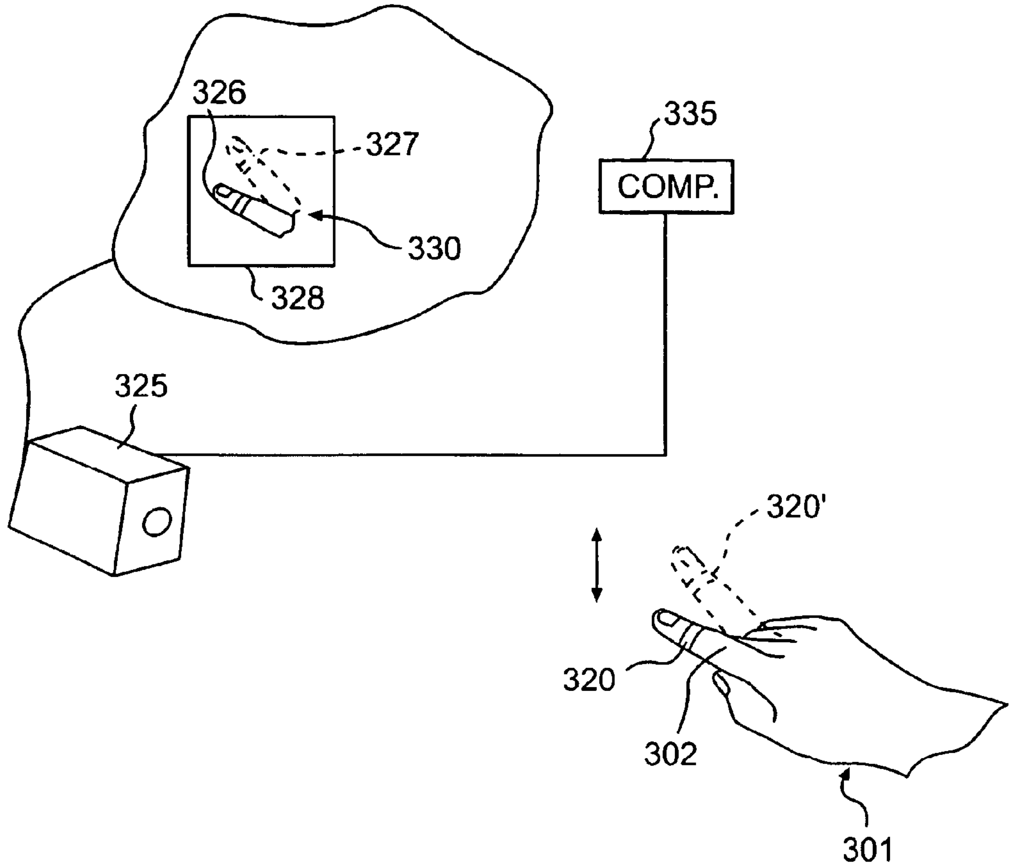


FIG. 3A

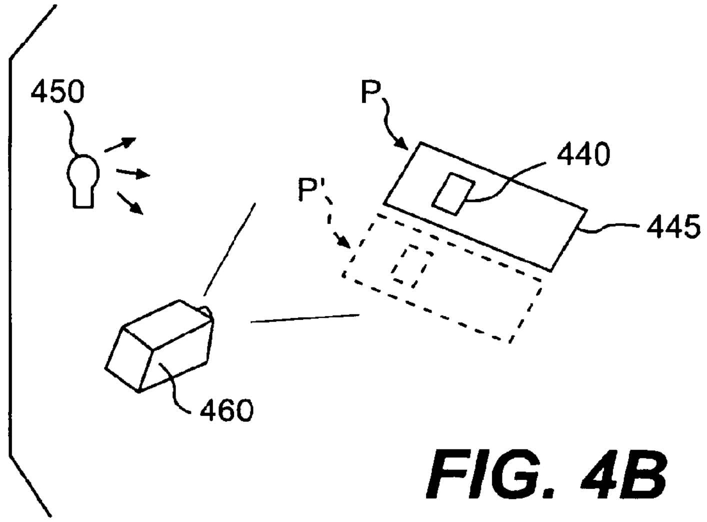


FIG. 4B

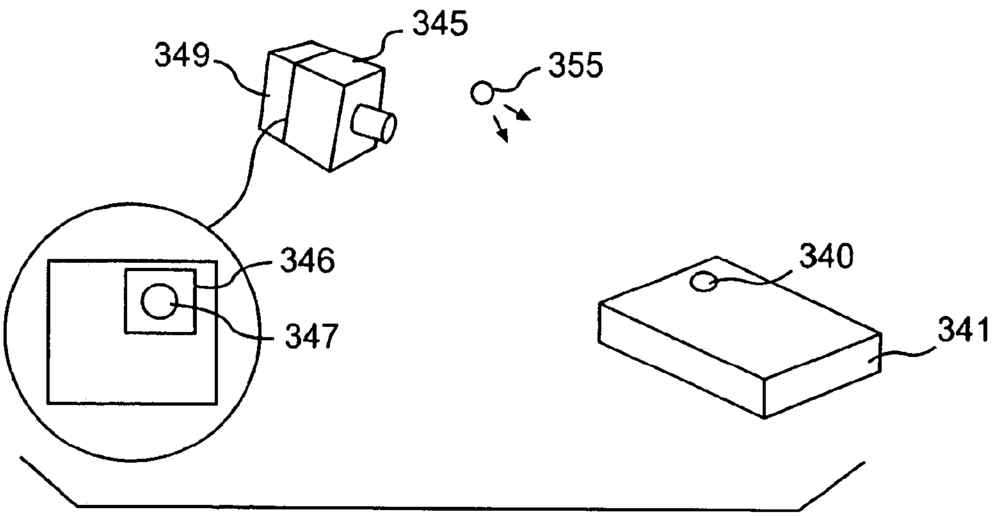


FIG. 3B

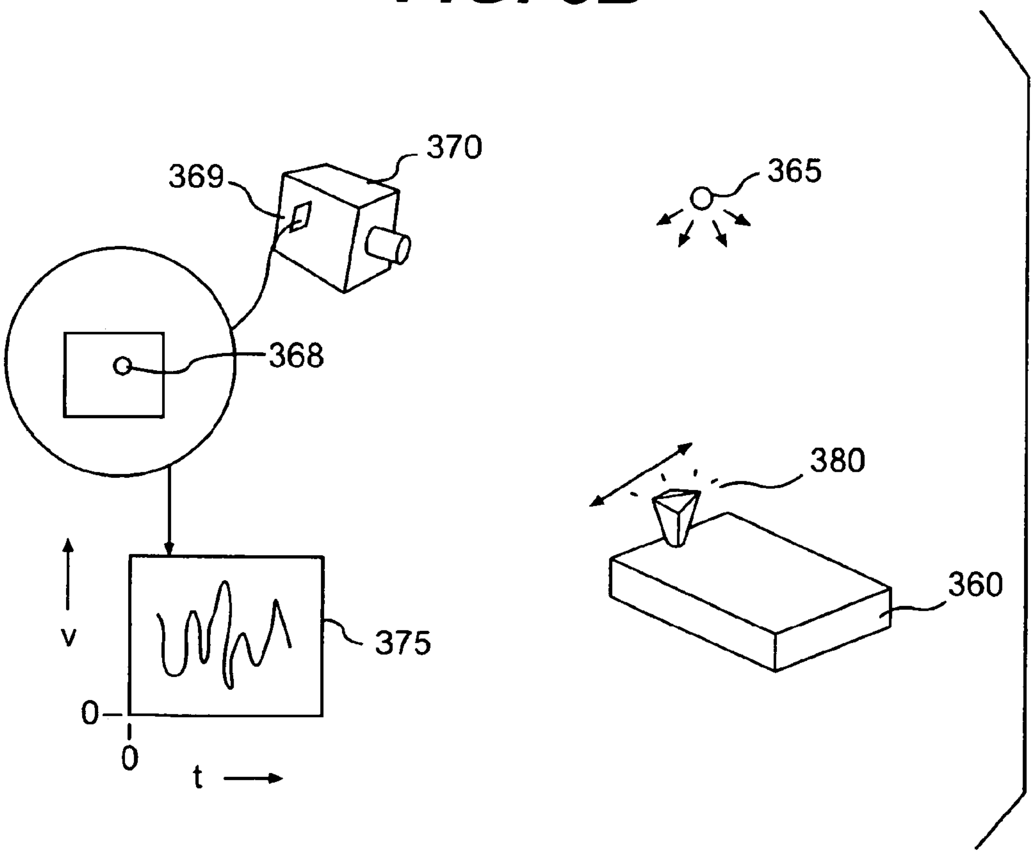


FIG. 3C

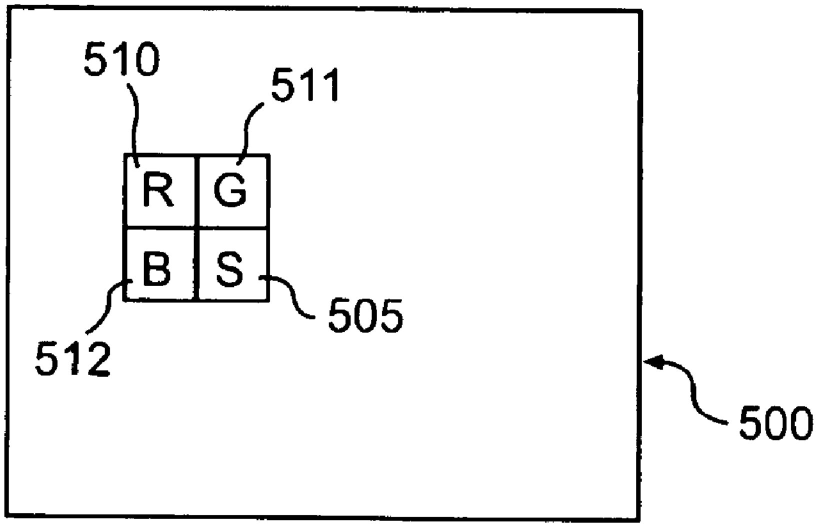


FIG. 5A

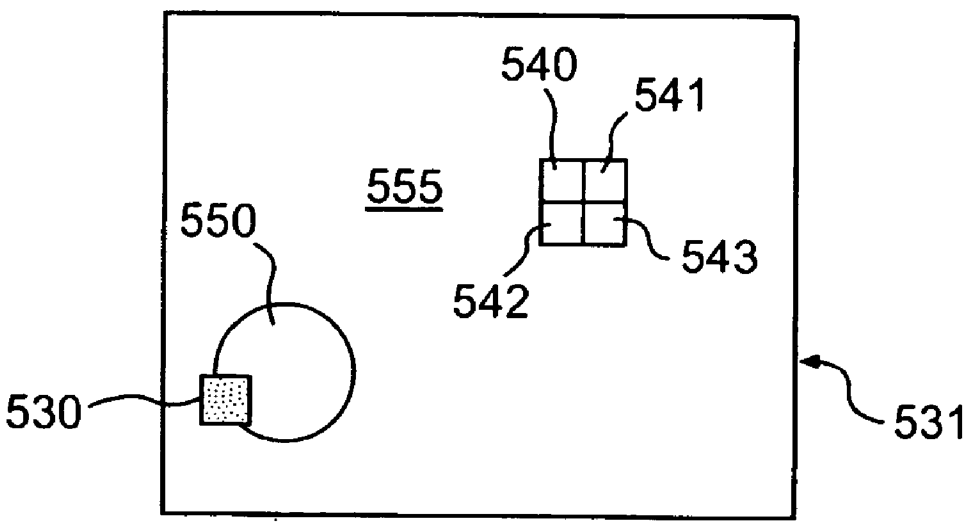


FIG. 5B

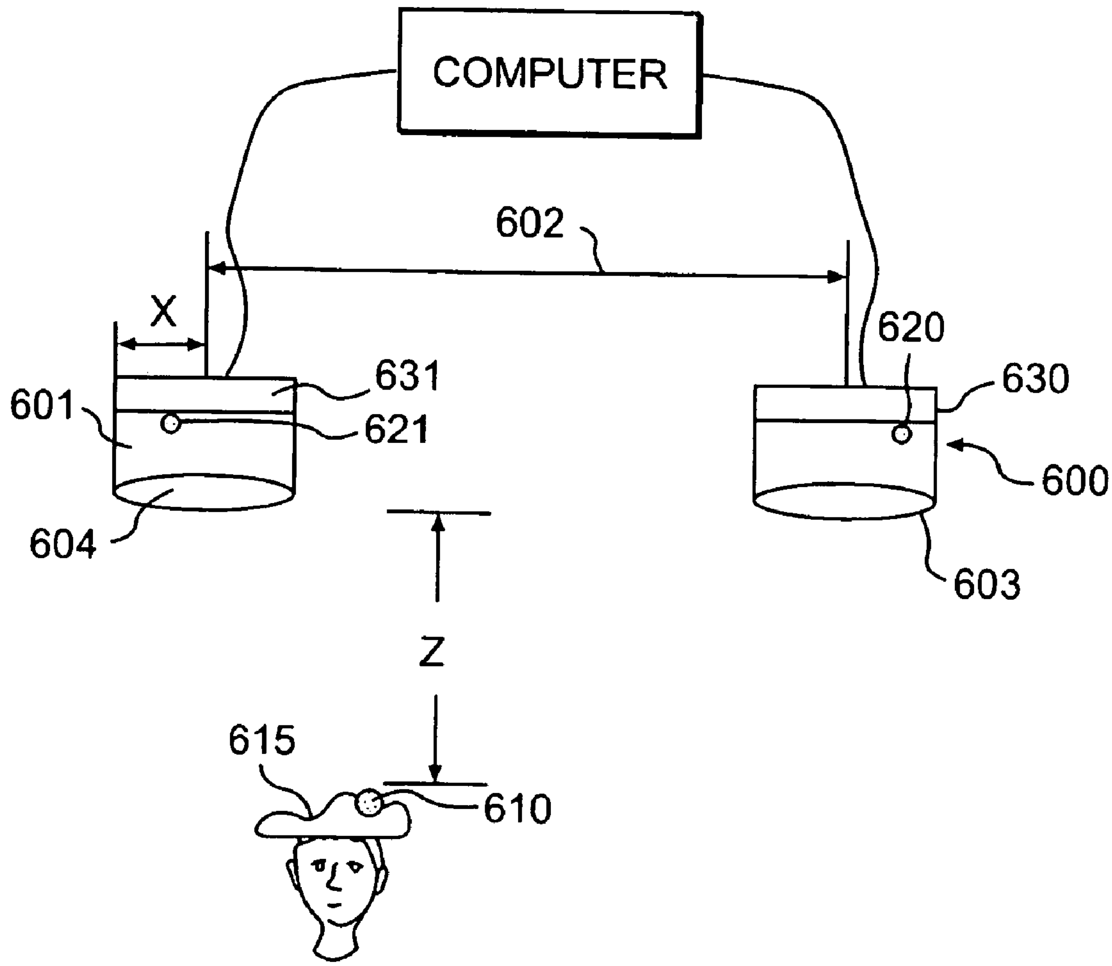
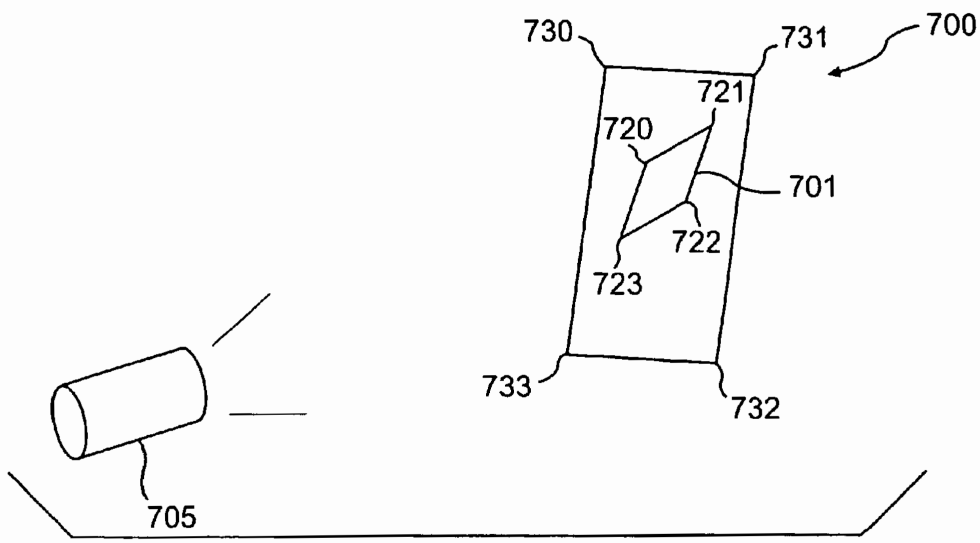
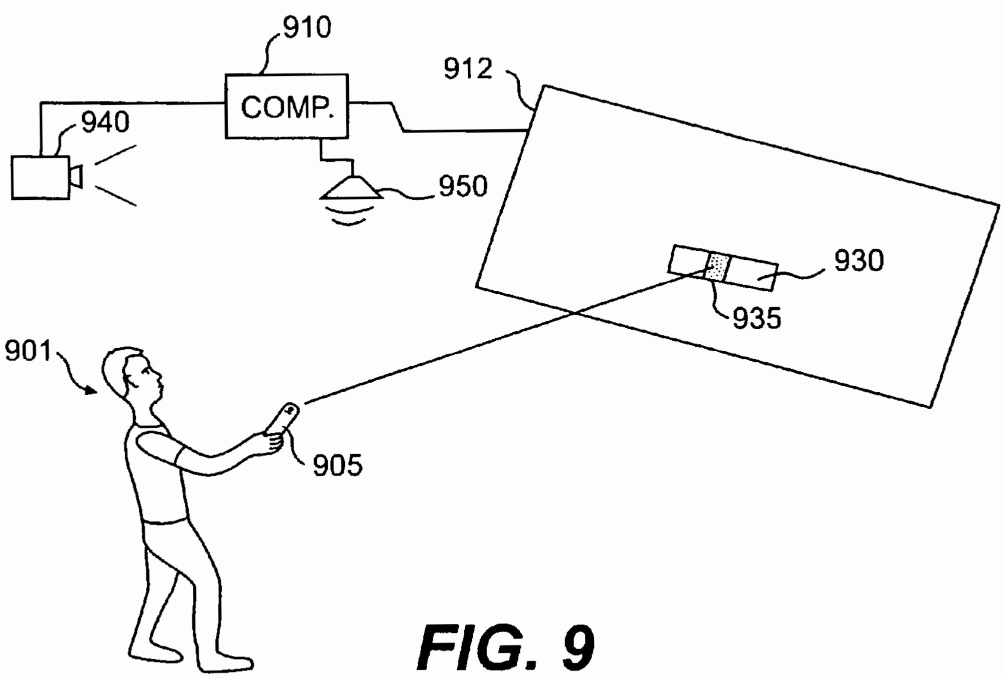


FIG. 6



730
731
700
721
720
701
722
723
733
732
705

FIG. 7



910
912
COMP.
940
950
930
935
901
905

FIG. 9

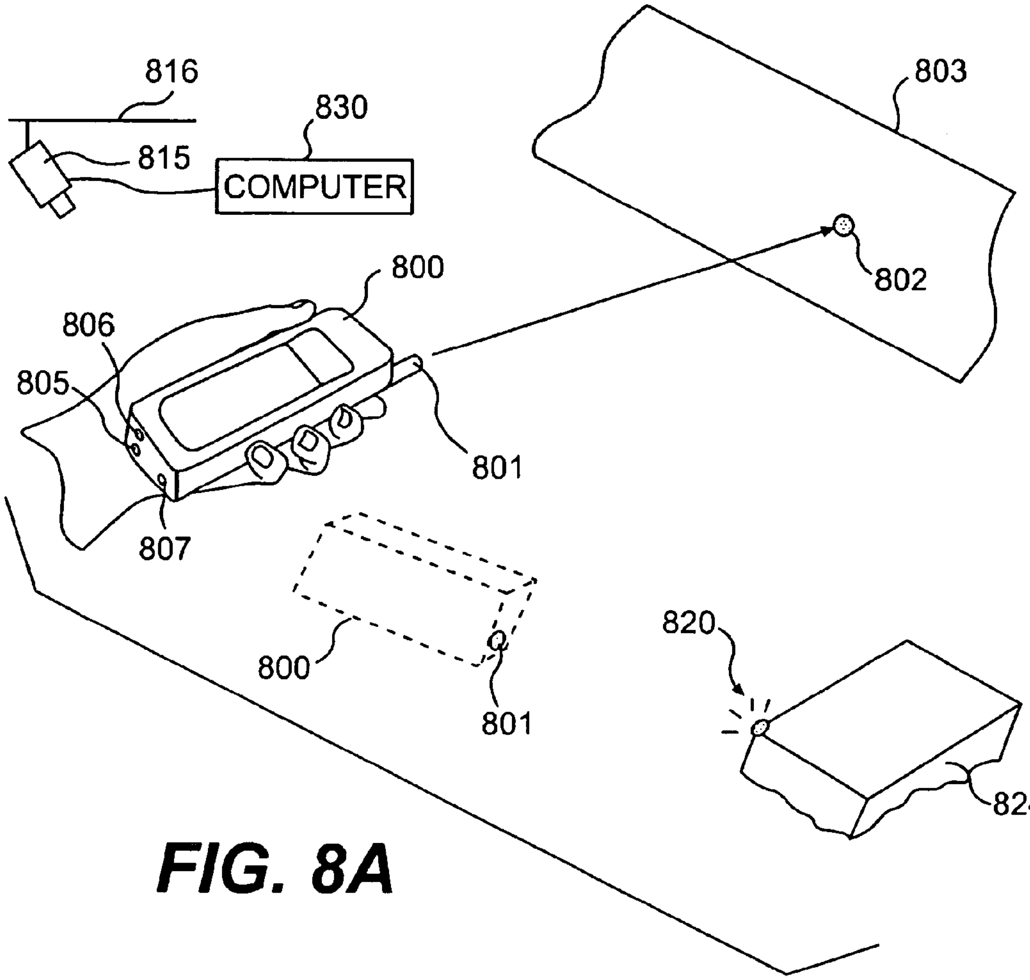

816
830
815
COMPUTER
803
800
802
806
805
801
807
800
801
820
824

FIG. 8A

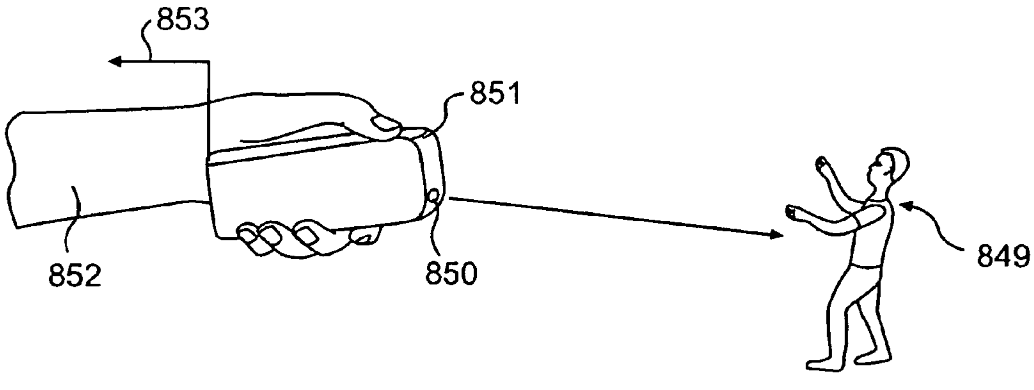

853
851
852
850
849

FIG. 8B

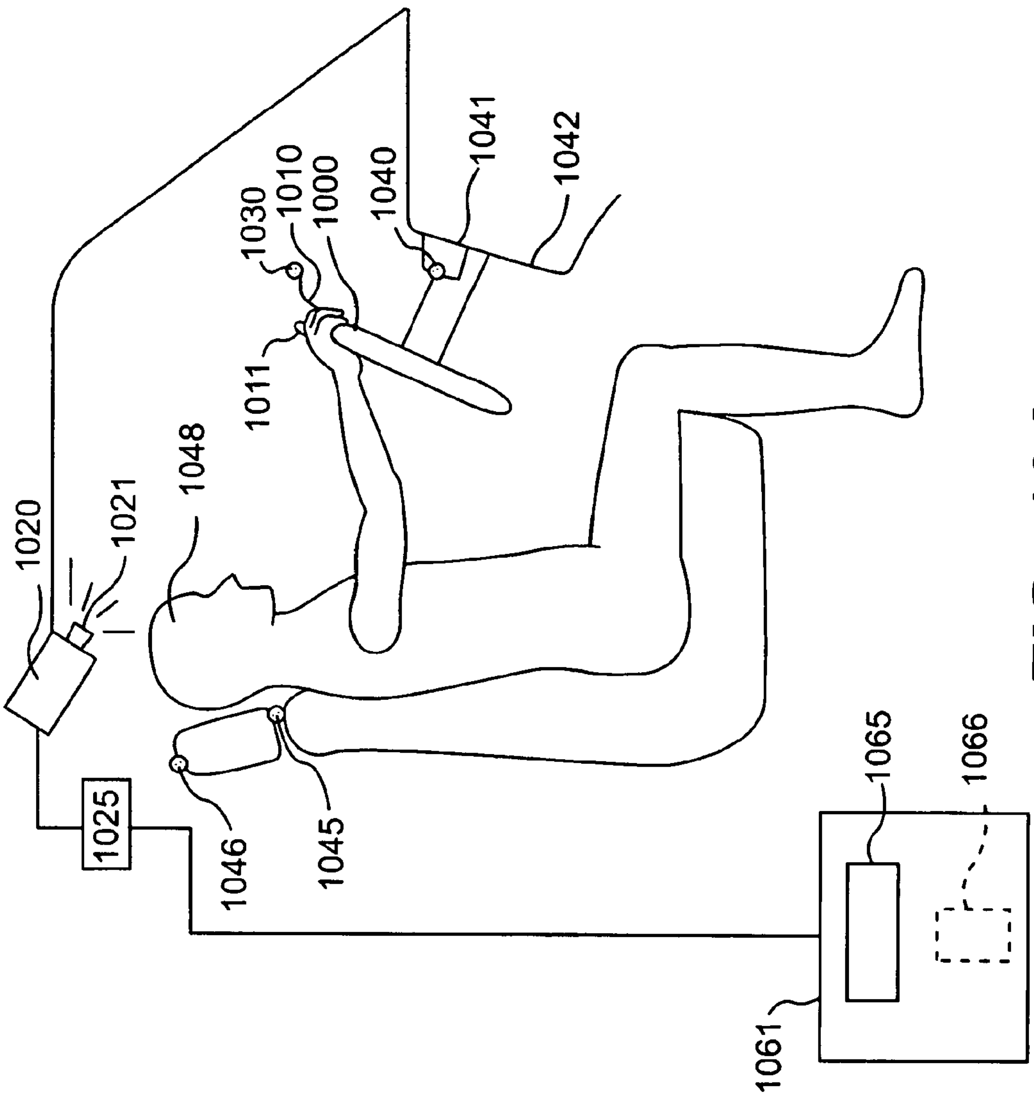

1000
1010
1011
1020
1021
1025
1030
1040
1041
1042
1045
1046
1048
1061
1065
1066

FIG. 10A

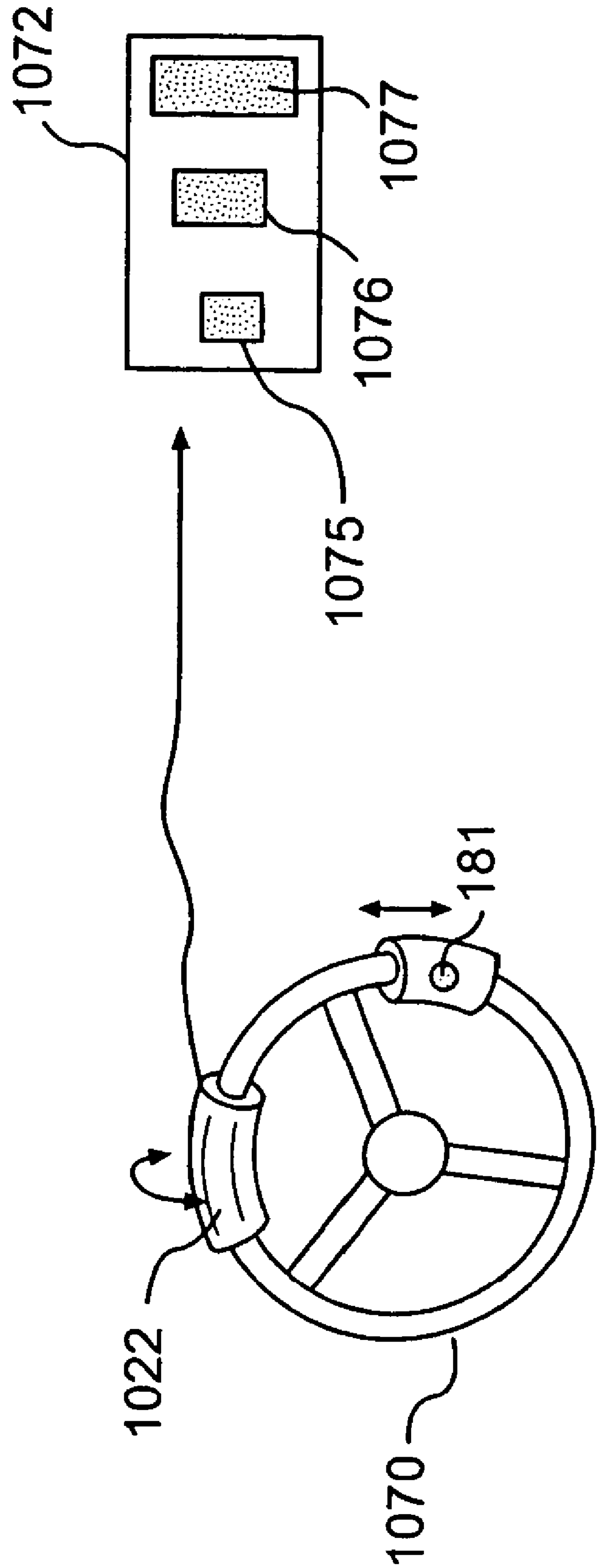


FIG. 10B

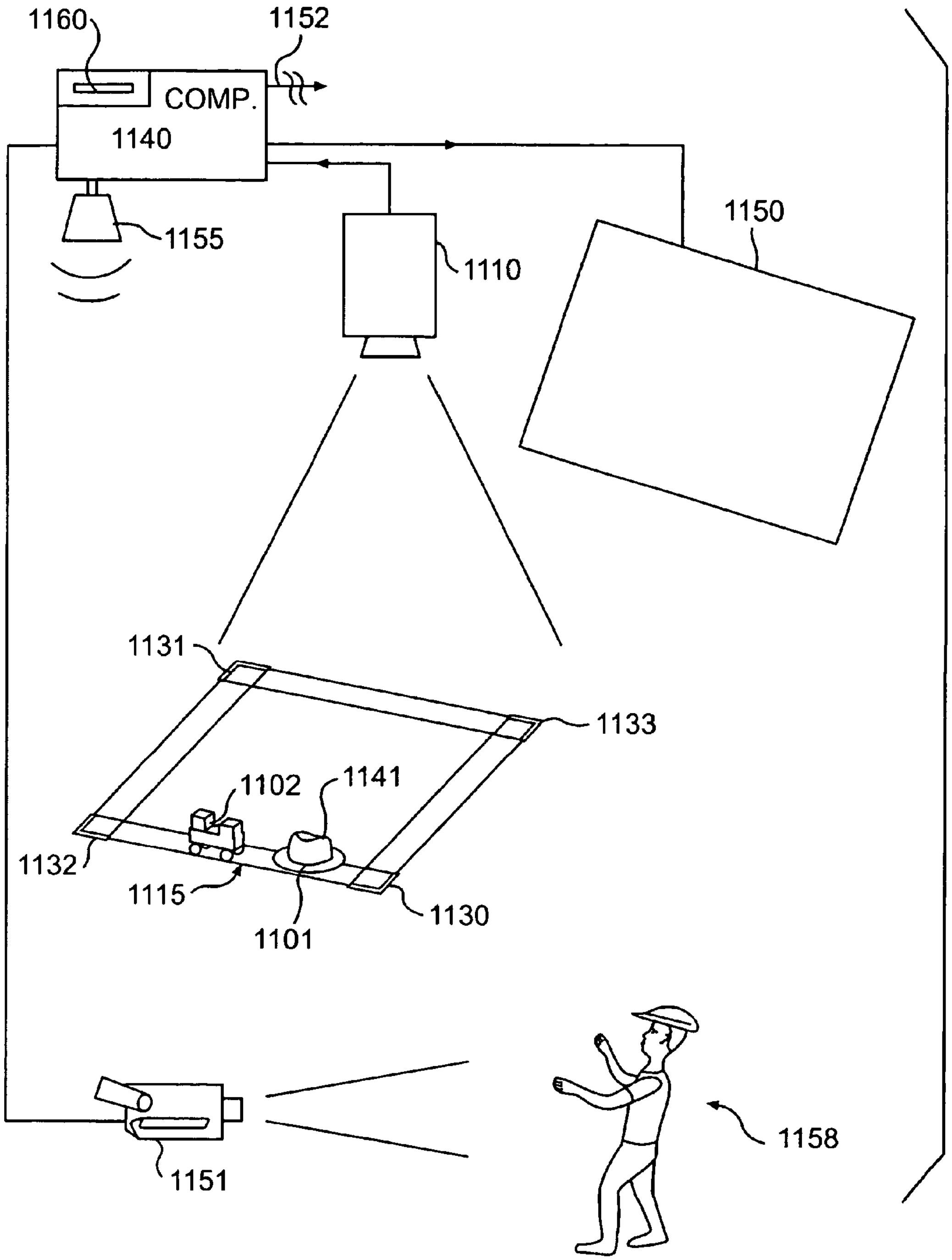

1160
1152
COMP.
1140
1155
1110
1150
1131
1133
1102
1141
1132
1115
1101
1130
1158
1151

FIG. 11A

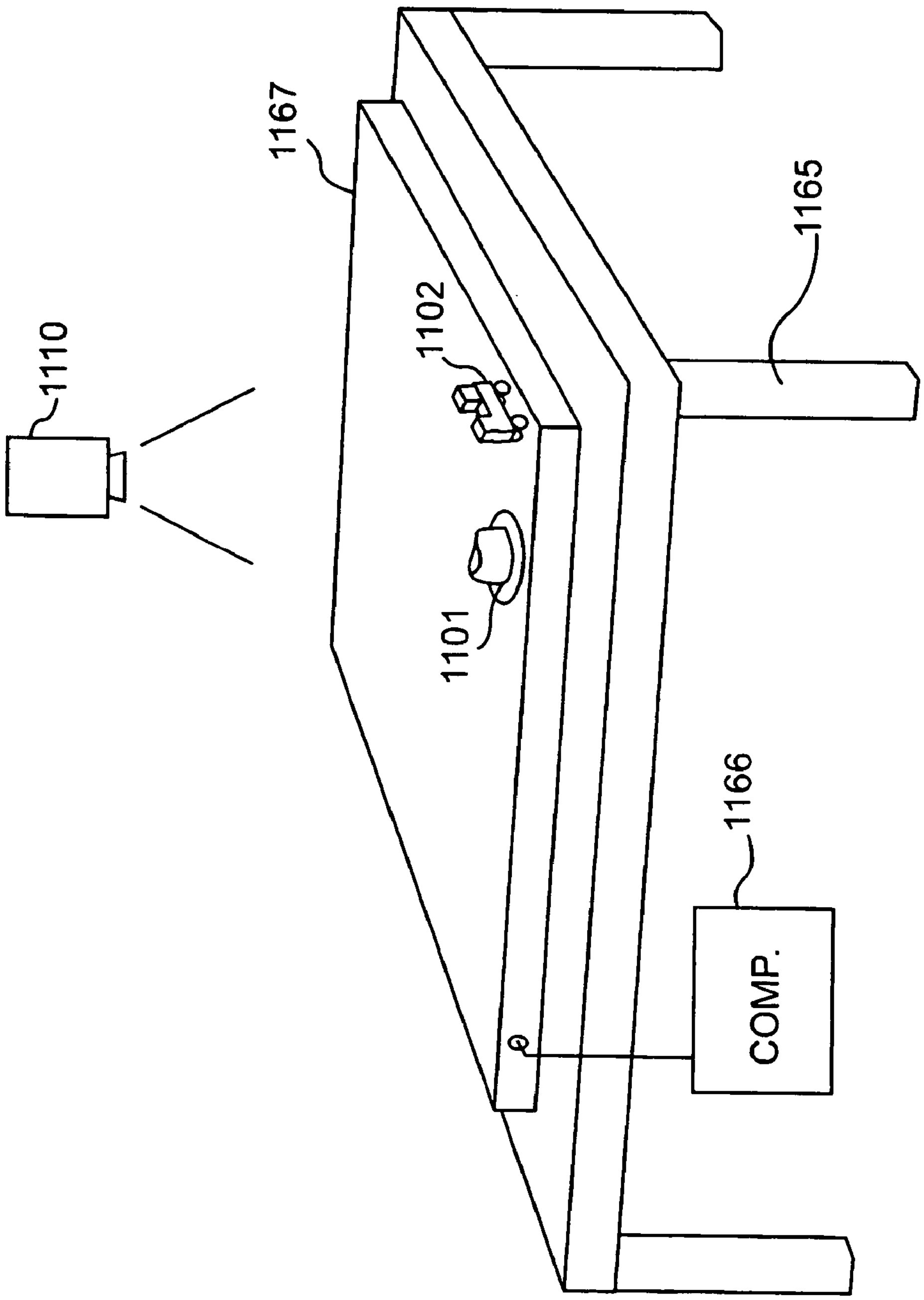


FIG. 11B

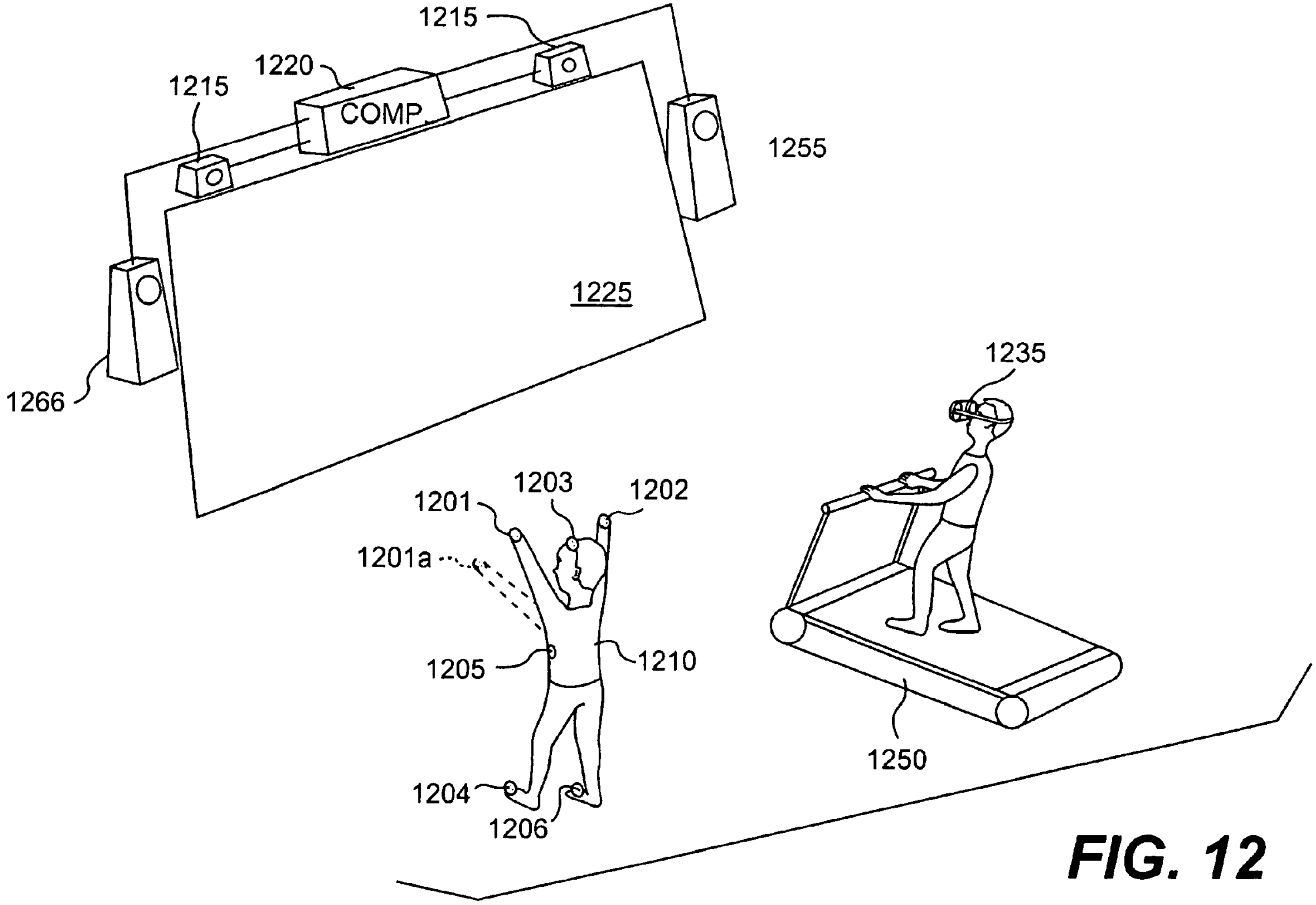

1215
1220
1215
COMP
1255
1225
1266
1235
1203
1201
1202
1201a
1205
1210
1250
1204
1206

FIG. 12

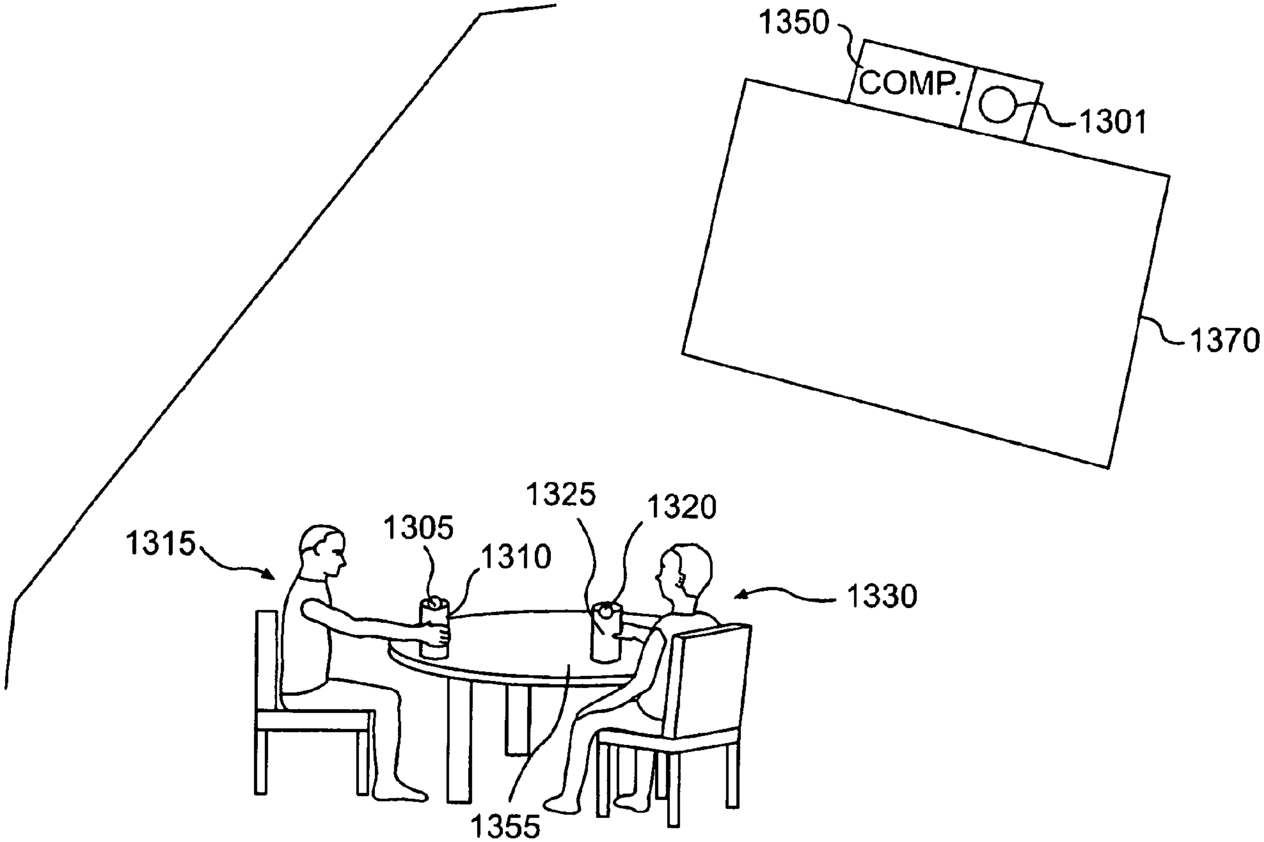


FIG. 13

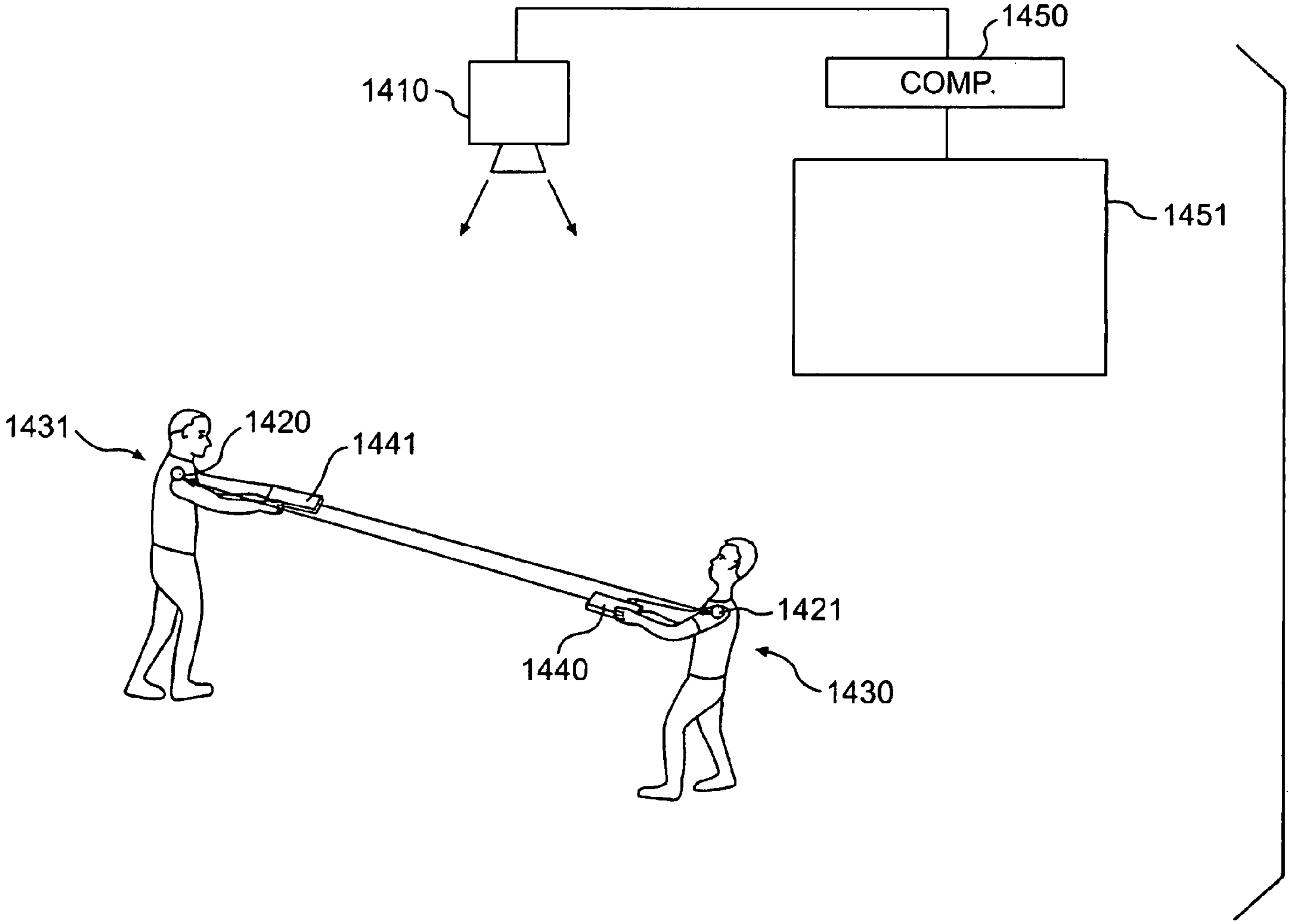


FIG. 14A

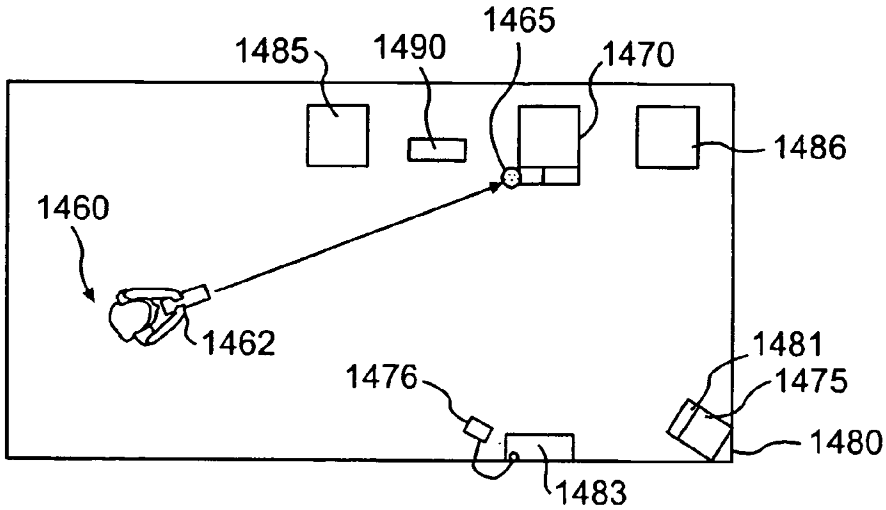


FIG. 14B

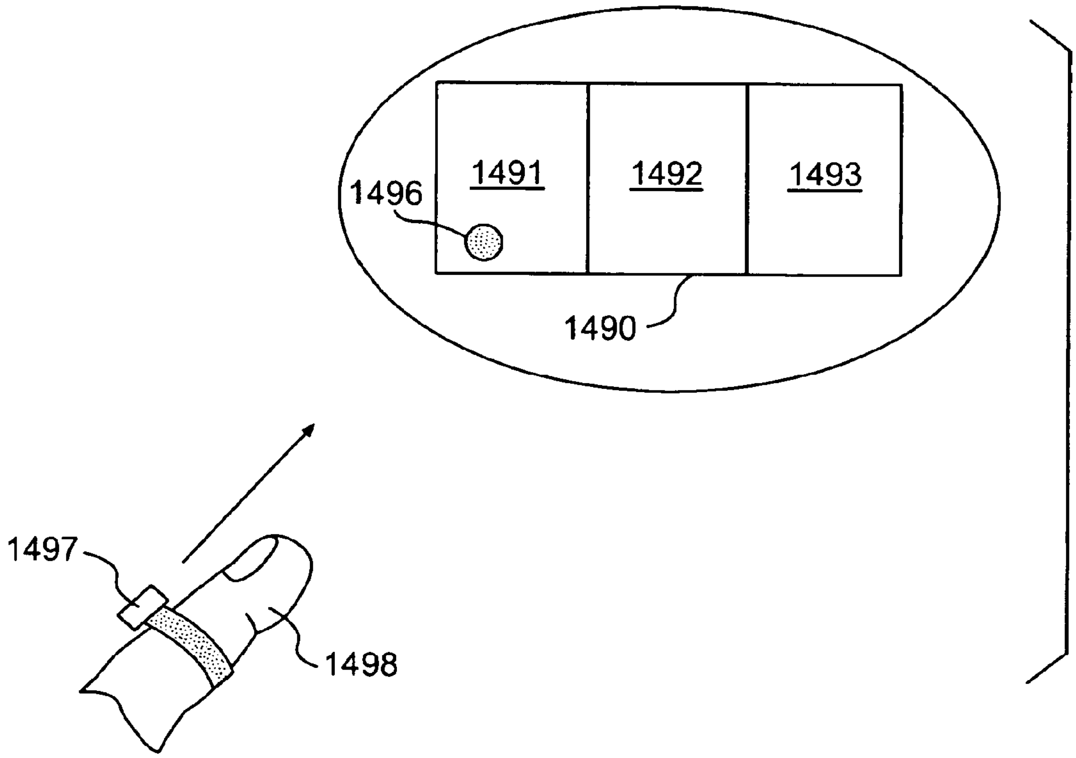


FIG. 14C

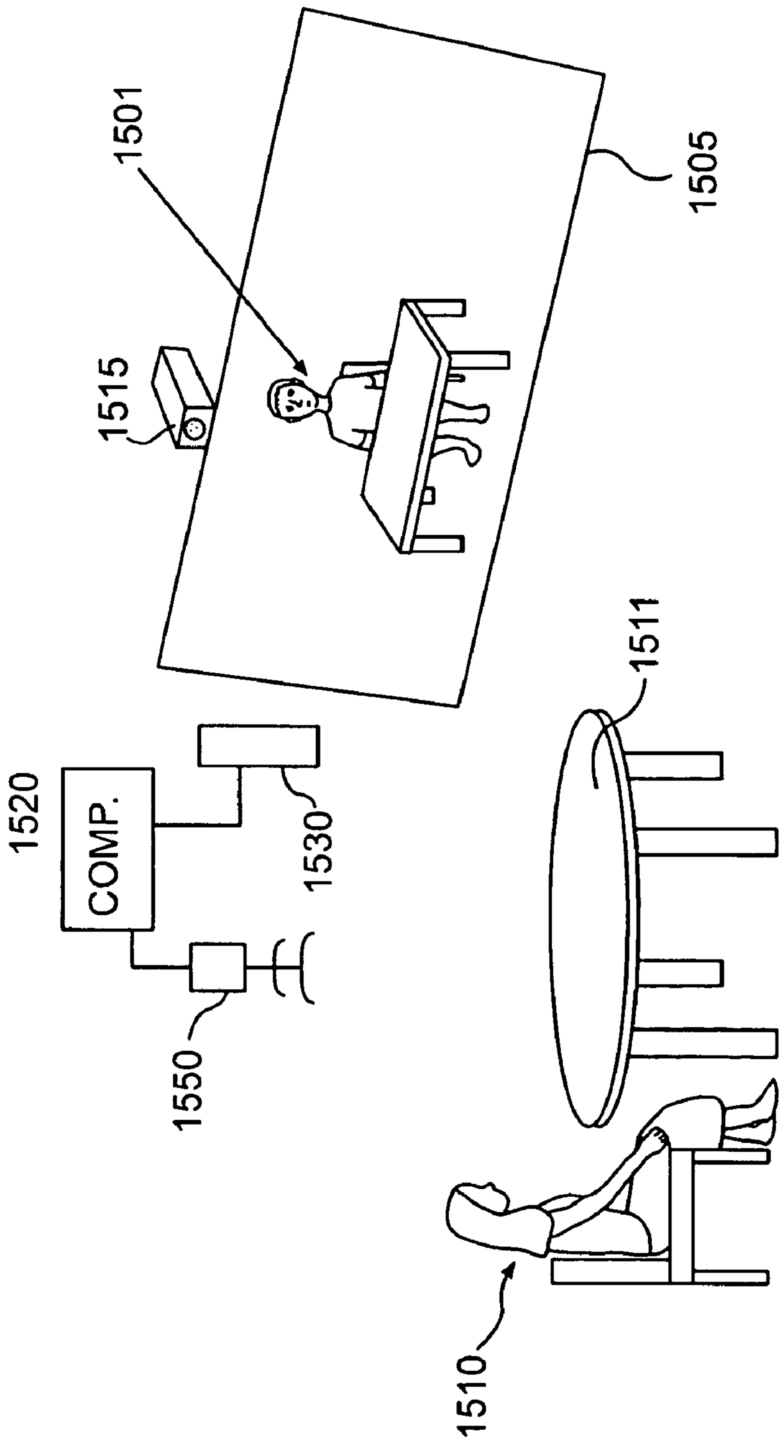


FIG. 15

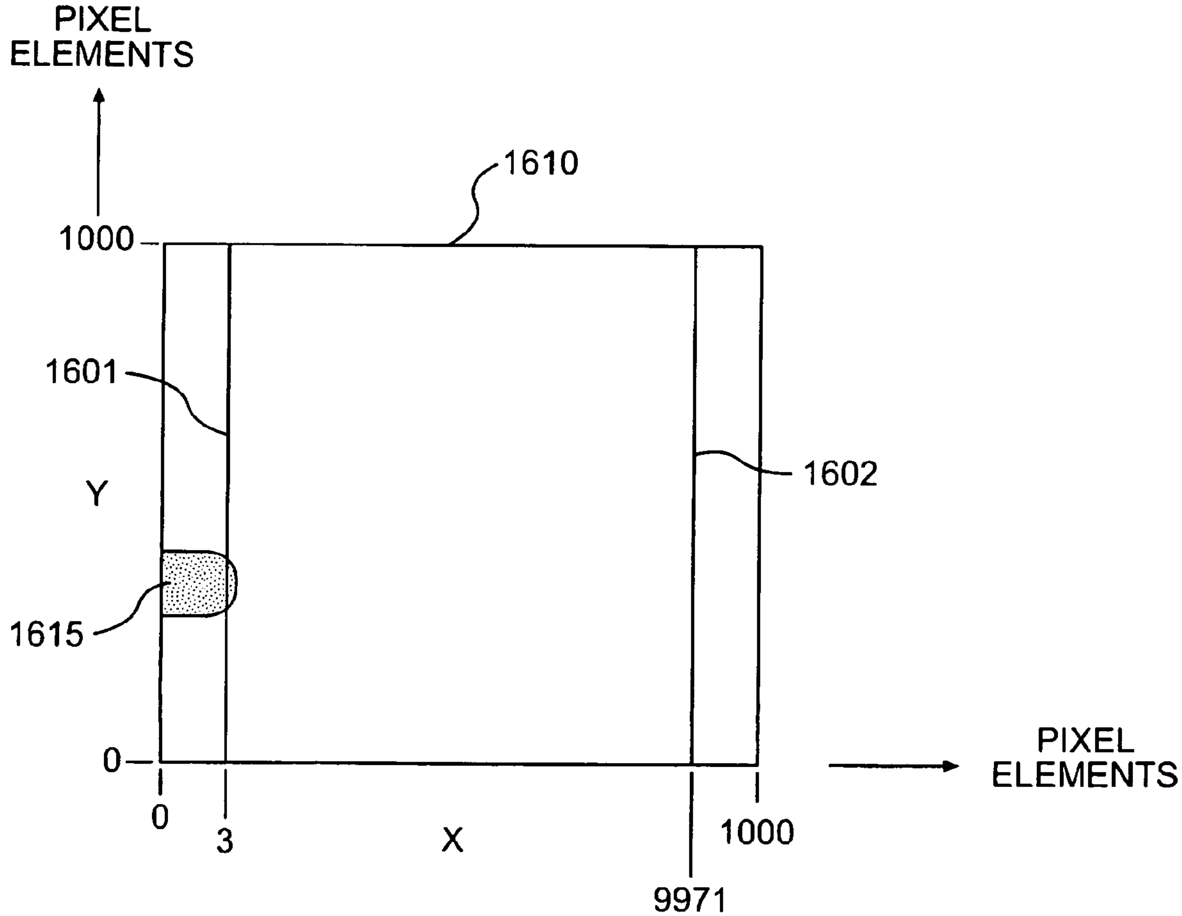

PIXEL
ELEMENTS
1610
1000
1601
1602
Y
1615
0
PIXEL
ELEMENTS
0
3
X
1000
9971

FIG. 16

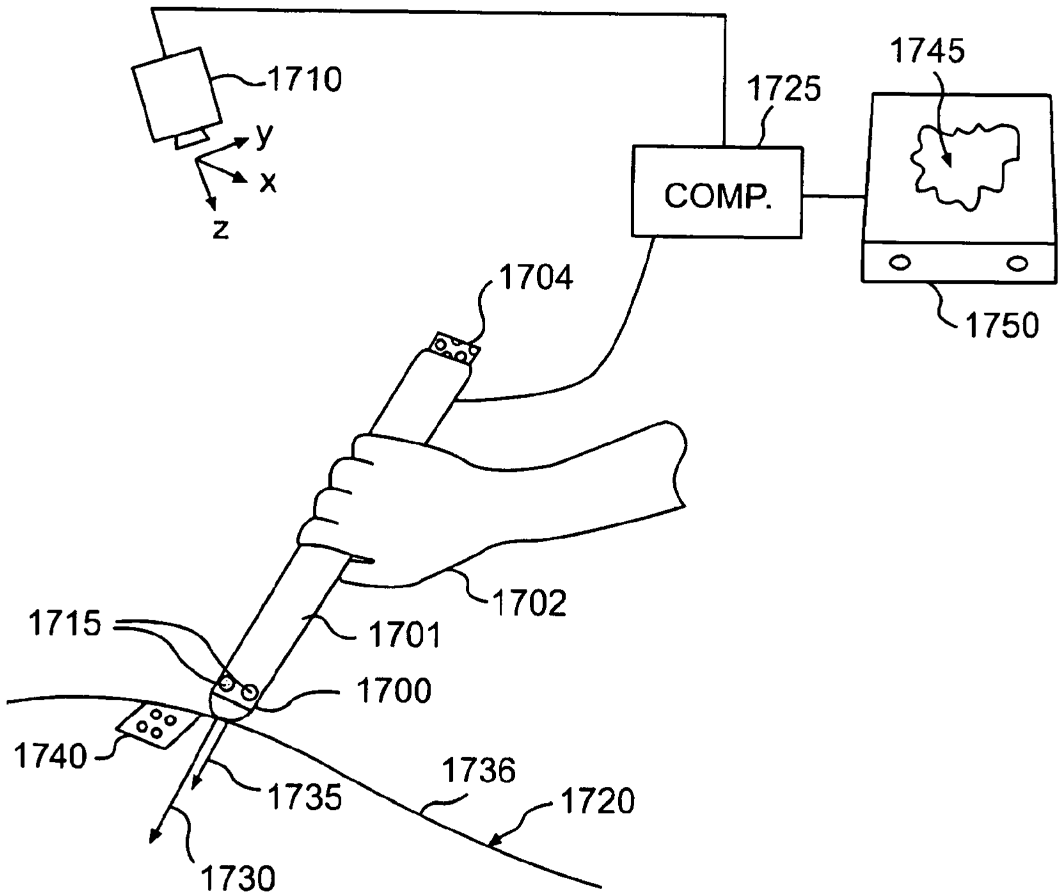


FIG. 17A

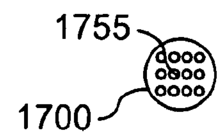


FIG. 17C

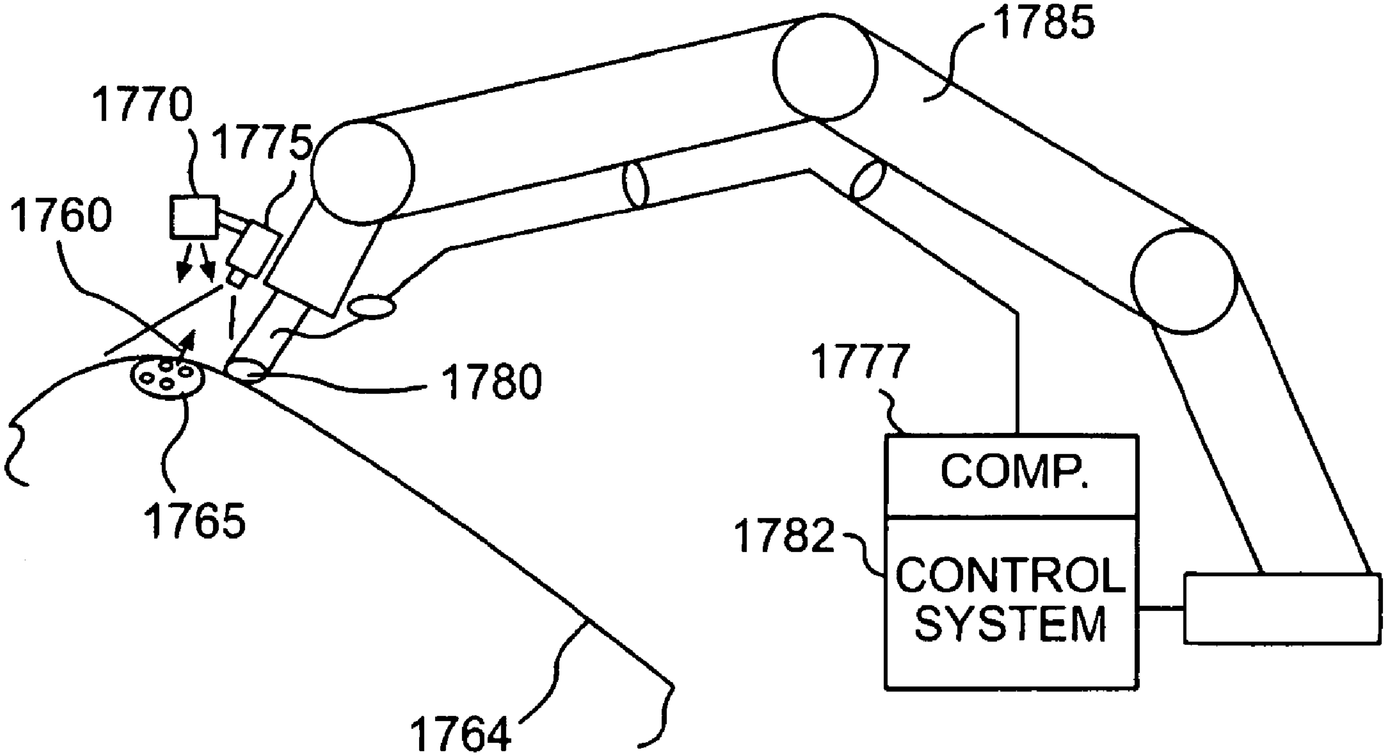

1785
1770
1775
1760
1780
1777
COMP.
1782
CONTROL SYSTEM
1765
1764

FIG. 17B

CAMERA BASED SENSING IN HANDHELD, MOBILE, GAMING, OR OTHER DEVICES

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation of application Ser. No. 11/980,710 filed Oct. 31, 2007, now U.S. Pat. No. 7,756,297; which is a continuation of application Ser. No. 10/893,534 filed Jul. 19, 2004, now U.S. Pat. No. 7,401,783; which is a continuation of application Ser. No. 09/612,225 filed Jul. 7, 2000, now U.S. Pat. No. 6,766,036; which claims the benefit of U.S. Provisional Application No. 60/142,777, filed Jul. 8, 1999.

Cross references to related co-pending US applications by the inventor having similar subject matter.

1. Touch TV and other Man Machine Interfaces: Ser. No. 09/435,854 filed Nov. 8, 1999, now U.S. Pat. No. 7,098,891; which was a continuation of application Ser. No. 07/946,908, now U.S. Pat. No. 5,982,352;
2. More Useful Man Machine Interfaces and Applications: Ser. No. 09/433,297 filed Nov. 3, 1999, now U.S. Pat. No. 6,750,848;
3. Useful Man Machine interfaces and applications: Ser. No. 09/138,339, Pub. Appln. 2002-0036617, now abandoned;
4. Vision Target based assembly: Ser. No. 08/469,907 filed Jun. 6, 1995, now U.S. Pat. No. 6,301,783;
5. Picture Taking method and apparatus: provisional application 60/133,671, and regular application Ser. No. 09/568,552 filed May 11, 2000, now U.S. Pat. No. 7,015,950;
6. Methods and Apparatus for Man Machine Interfaces and Related Activity: Provisional Application: provisional application 60/133,673 filed May 11, 1999; and regular application Ser. No. 09/568,554 filed May 11, 2000, now U.S. Pat. No. 6,545,670;
7. Tactile Touch Screens for Automobile Dashboards, Interiors and Other Applications: provisional application Ser. No. 60/183,807; and regular application Ser. No. 09/789,538, now U.S. Pat. No. 7,084,859; and
8. Apparel Manufacture and Distance Fashion Shopping in Both Present and Future: provisional application 60/187,397 filed Mar. 7, 2000.

The disclosures of the following U.S. patents and co-pending patent applications by the inventor, or the inventor and his colleagues, are incorporated herein by reference:

1. "Man machine Interfaces": U.S. application Ser. No. 09/435,854 and U.S. Pat. No. 5,982,352, and U.S. application Ser. No. 08/290,516, filed Aug. 15, 1994, now U.S. Pat. No. 6,008,000, the disclosure of both of which is contained in that of Ser. No. 09/435,854;
2. "Useful Man Machine Interfaces and Applications": U.S. application Ser. No. 09/138,339, now Pub. Appln. 2002-0036617;
3. "More Useful Man Machine Interfaces and Applications": U.S. application Ser. No. 09/433,297;
4. "Methods and Apparatus for Man Machine Interfaces and Related Activity": U.S. Appln. Ser. No. 60/133,673 filed as regular application Ser. No. 09/568,554, now U.S. Pat. No. 6,545,670;
5. "Tactile Touch Screens for Automobile Dashboards, Interiors and Other Applications": U.S. provisional Appln. Ser. No. 60/183,807, filed Feb. 22, 2000, now filed as reg. application Ser. No. 09/789,538; and

6. "Apparel Manufacture and Distance Fashion Shopping in Both Present and Future": U.S. Appln. Ser. No. 60/187,397, filed Mar. 7, 2000.

FIELD OF THE INVENTION

The invention relates to simple input devices for computers, particularly, but not necessarily, intended for use with 3-D graphically intensive activities, and operating by optically sensing a human input to a display screen or other object and/or the sensing of human positions or orientations. The invention herein is a continuation in part of several inventions of mine, listed above.

This continuation application seeks to provide further useful embodiments for improving the sensing of objects. Also disclosed are new applications in a variety of fields such as computing, gaming, medicine, and education. Further disclosed are improved systems for display and control purposes.

The invention uses single or multiple TV cameras whose output is analyzed and used as input to a computer, such as a home PC, to typically provide data concerning the location of parts of, or objects held by, a person or persons.

DESCRIPTION OF RELATED ART

The above mentioned co-pending applications incorporated by reference discuss many prior art references in various pertinent fields, which form a background for this invention. Some more specific U.S. Patent references are for example:

DeMenthon—U.S. Pat. Nos. 5,388,059; 5,297,061; 5,227,985

Cipolla—U.S. Pat. No. 5,581,276

Pugh—U.S. Pat. No. 4,631,676

Pinckney—U.S. Pat. No. 4,219,847

DESCRIPTION OF FIGURES

FIG. **1** illustrates a basic computer terminal embodiment of the invention, similar to that disclosed in copending applications.

FIG. **2** illustrates object tracking embodiments of the invention employing a pixel addressable camera.

FIG. **3** illustrates tracking embodiments of the invention using intensity variation to identify and/or track object target datums.

FIG. **4** illustrates tracking embodiments of the invention using variation in color to identify and/or track object target datums.

FIG. **5** illustrates special camera designs for determining target position in addition to providing normal color images.

FIG. **6** identification and tracking with stereo pairs.

FIG. **7** illustrates use of an indicator or co-target.

FIG. **8** illustrates control of functions with the invention, using a handheld device which itself has functions.

FIG. **9** illustrates pointing at an object represented on a screen using a finger or laser pointer, and then manipulating the represented object using the invention.

FIG. **10** illustrates control of automobile or other functions with the invention, using detected knob, switch or slider positions.

FIG. **11** illustrates a board game embodiment of the invention.

FIG. **12** illustrates a generic game embodiment of the invention.

FIG. **13** illustrates a game embodiment of the invention, such as might be played in a bar.

FIG. **14** illustrates a laser pointer or other spot designator embodiment of the invention.

FIG. **15** illustrates a gesture based flirting game embodiment of the invention.

FIG. **16** illustrates a version of the pixel addressing camera technique wherein two lines on either side of a 1000 element square array are designated as perimeter fence lines to initiate tracking or other action.

FIG. **17** illustrates a 3-D acoustic imaging embodiment of the invention.

THE INVENTION EMBODIMENTS

FIG. **1**

The invention herein and disclosed in portions of other copending applications noted above, comprehends a combination of one or more TV cameras (or other suitable electro-optical sensors) and a computer to provide various position and orientation related functions of use. It also comprehends the combination of these functions with the basic task of generating, storing and/or transmitting a TV image of the scene acquired—either in two or three dimensions.

The embodiment depicted in FIG. **1**A illustrates the basic embodiments of many of my co-pending applications above. A stereo pair of cameras **100** and **101** located on each side of the upper surface of monitor **102** (for example a rear projection TV of 60 inch diagonal screen size) with display screen **103** facing the user, are connected to PC computer **106** (integrated in this case into the monitor housing), for example a 400 Mhz Pentium II. For appearances and protection a single extensive cover window may be used to cover both cameras and their associated light sources **110** and **111**, typically LEDs.

The LEDs in this application are typically used to illuminate targets associated with any of the fingers, hand, feet and head of the user, or objects such as **131** held by a user, **135** with hands **136** and **137**, and head **138**. These targets, such as circular target **140** and band target **141** on object **131** are desirably, but not necessarily, retro-reflective, and may be constituted by the object features themselves (e.g., a finger tip, such as **145**), or by features provided on clothing worn by the user (e.g., a shirt button **147** or polka dot **148**, or by artificial targets other than retroreflectors.

Alternatively, a three camera arrangement can be used, for example using additional camera **144**, to provide added sensitivity in certain angular and positional relationships. Still more cameras can be used to further improve matters, as desired. Alternatively, and or in addition, camera **144** can be used for other purposes, such as acquire images of objects such as persons, for transmission, storage or retrieval independent of the cameras used for datum and feature location determination.

For many applications, a single camera can suffice for measurement purposes as well, such as **160** shown in FIG. **1**B for example, used for simple 2 dimensional (2D) measurements in the xy plane perpendicular to the camera axis (z axis), or 3D (xyz, roll pitch yaw) where a target grouping, for example of three targets is used such as the natural features formed by the two eyes **164**, **165** and nose **166** of a human **167**. These features are roughly at known distances from each other, the data from which can be used to calculate the approximate position and orientation of the human face. Using for example the photogrammetric technique of Pinkney described below, the full 6 degree of freedom solution of the human face location and orientation can be achieved to an accuracy limited by the ability of the camera image processing software utilized to determine the centroids or other delineating geometric indicators of the position of the eyes and nose, (or some other facial feature such as the mouth), and the accuracy of the initial imputing of the spacing of the eyes and their respective spacing to the nose. Clearly if a standard human value is used (say for adult, or for a child or even by age) some lessening of precision results, since these spacings are used in the calculation of distance and orientation of the face of human **167** from the camera **160**.

In another generally more photogrammetrically accurate case, one might choose to use four special targets (e.g., glass bead retro-reflectors, or orange dots) **180**-**183** on the object **185** having known positional relationships relative to each other on the object surface, such as one inch centers. This is shown in FIG. **1**C, and may be used in conjunction with a pixel addressable camera such as described in FIG. **2** below, which allows one to rapidly determine the object position and orientation and track its movements in up to 6 degrees of freedom as disclosed by Pinkney U.S. Pat. No. 4,219,847 and technical papers referenced therein. For example, the system described above for FIGS. **1** and **2** involving the photogrammetric resolution of the relative position of three or more known target points as viewed by a camera is known and is described in a paper entitled "A Single Camera Method for the 6-Degree of Freedom Sprung Mass Response of Vehicles Redirected by Cable Barriers" presented by M. C. van Wijk and H. F. L. Pinkney to The Society of Photo-optical Instrumentation Engineers.

The stereo pair of cameras can also acquire a two view stereo image of the scene as well, which can be displayed in 3D using stereoscopic or auto-stereoscopic means, as well as transmitted or recorded as desired.

In many applications of the foregoing invention it is desirable not just to use a large screen but in fact one capable of displaying life size images. This particularly relates to human scaled images, giving a life-like presence to the data on the screen. In this way the natural response of the user with motions of hands, head, arms, etc., is scaled in "real" proportion to the data being presented.

FIG. **2**

This embodiment and others discloses special types of cameras useful with the invention. In the first case, that of FIG. **2**A, a pixel addressable camera such as the MAPP2200 made by IVP corporation of Sweden is used, which allows one to do many things useful for rapidly determining location of objects, their orientation and their motion.

For example, as shown in FIG. **2**A, an approximately circular image **201** of a target datum such as **180** on object **185** of FIG. **1**C may be acquired by scanning the pixel elements on a matrix array **205** on which the image is formed. Such an array in the future will have for example 1000×1000 pixels, or more (today the largest IVP makes is 512×512. The IVP also is not believed to be completely randomly addressable, which some future arrays will be).

As an illustration, computer **220** determines, after the array **205** has been interrogated, that the centroid "x, y" of the pixel elements on which the target image lies is at pixel x=500, y=300 (including a sub-fraction thereof in many cases). The centroid location can be determined for example by the moment method disclosed in the Pinkney patent, referenced above.

The target in this case is defined as a contrasting point on the object, and such contrast can be in color as well as, or instead of, intensity. Or with some added preprocessing, it can be a distinctive pattern on the object, such as a checkerboard or herringbone.

Subsequent Tracking

To subsequently track the movement of this target image, it is now only necessary to look in a small pixel window composed of a small number of pixels around the target. For example the square **230** shown, as the new position x'y' of the target image cannot be further distant within a short period of time elapsed from the first scan, and in consideration of the small required time to scan the window.

For example, if the window is 100×100 pixels, this can be scanned in 1 millisecond or less with such a pixel addressing camera, by interrogating only those pixels in the window, while still communicating with the camera over a relatively slow USB serial link of 12 mb transmission rate (representing 12,000 pixel gray level values in one millisecond).

One thus avoids the necessity to scan the whole field, once the starting target image position is identified. This can be known by an initial scan as mentioned, or can be known by having the user move an object with a target against a known location with respect to the camera such as a mechanical stop, and then indicate that tracking should start either by verbally saying so with voice recognition, or by actuating a control key such as **238** or whatever.

It is noted that if the tracking window is made large enough, then it can encompass a whole group of datums, such as **180-183** on an object.

FIG. **2**B Reduction in Acquisition Time

Another application of such a pixel addressing camera is shown in FIG. **2**B. One can look at the whole field, x y of the camera, **240**, but only address say every 10^{th} pixel such as **250**, **251** and **252**, in each direction, i.e., for a total 10,000 pixels in a field of 1 million (1000×1000, say).

In this case computer **220** simply queries this fraction of the pixels in the image, knowing apriori that the target image such as **260** will have an image size larger than 10×10 pixels, and must be detectable, if of sufficient contrast, by one of the queried pixels. (For smaller or larger target images, the number and spacing of queried pixels can be adjusted accordingly). This for example, allows one to find approximate location of targets with only 1/100 the pixel interrogation time otherwise needed, for example, plus any gain obtained as disclosed above, by knowing in what region of the image to look (for example during tracking, or given some apriori knowledge of approximate location due to a particular aspect of the physical arrangement or the program in question).

Once a target has been approximately found as just described, the addressing can be optimized for that region of the image only, as disclosed in subsequent tracking section above.

Given the invention, the potential for target acquisition in a millisecond or two thus is achievable with simple pixel addressable CMOS cameras coming on stream now (today costing under $50), assuming the target points are easily identifiable from at least one of brightness (over a value), contrast (with respect to surroundings), color, color contrast, and more difficult, shape or pattern (e.g., a plaid, or herringbone portion of a shirt). This has major ramifications for the robustness of control systems built on such camera based acquisition, be they for controlling displays, or machines or whatever.

It's noted that with new 2000×2000 cameras coming on stream, it may only be necessary to look at every 15^{th} or 20^{th} pixel in each direction to get an adequate feel for target location. This means every 200^{th} to 400^{th} pixel, not enough to cause image rendition difficulties even if totally dark grey (as it might be in a normal white light image if set up for IR wavelengths only).

FIG. **2**C

Another method for finding the target in the first place with limited pixel interrogation is to look at pixels near a home point where a person for example indicates that the target is. This could be for example, placing ones fingernail such as **270**, whose natural or artificial (e.g., reflective nail polish) features are readily seen by the camera **275** and determined to be in the right corner of a pad **271** in FIG. **2**C which approximately covers the field of view **274** of the camera **275**. The computer **220** analyzes the pixels in the right corner **278** of the image field **279** representing the pad portion **271** with the camera **275**, either continuously, or only when the finger for example hits a switch such as **280** at the edge of the pad, or on command (e.g., by the user pushing a button or key, or a voice message inputted via microphone **285** for example). After such acquisition, the target is then tracked to other locations in xy space of the pad, for example as described above. Its noted that it helps to provide a beep or other sound or indication when acquisition has been made.

Pick Windows in Real Time

Another aspect of the invention is that one can also pick the area of the image to interrogate at any desired moment. This can be done by creating a window of pixels with in the field to generate information, for example as discussed relative to a specific car dashboard application of FIG. **10**.

FIG. **2**D—Scan Pattern

A pixel addressing camera also allows a computer such as **220** to cause scans to be generated which are not typical raster scans. For example circular or radial, or even odd shapes as desired. This can be done by providing from the computer the sequential addresses of the successive pixels on the camera chip whose detected voltages are to be queried.

A circular scan of pixels addressed at high speed can be used to identify when and where a target enters a field enclosed by the circular pixel scan. This is highly useful, and after that, the approximate location of the target can be determined by further scans of pixels in the target region.

For example consider addressing the pixels **c1 c2 c3** . . . cn representing a circle **282** at the outer perimeter of the array, **285**, of 1000×1000 elements such as discussed above. The number of pixels in a full circle is approximately 1000 pi, which can be scanned even with USB (universal serial bus) limits at 300 times per second or better. For targets of 1/100 field in width, this means that a target image entering the field such as circular target image **289** (which is shown intersecting element cm and its neighbors) would have to travel 1/100 the field width in 0.0033 seconds to be totally missed in a worst case. If the image field corresponds to 20 inches in object field width this is 0.2 inches×300/sec or 60 inches/second, very fast for human movement, and not likely to be exceeded even where smaller targets are used.

Alternative shapes to circular "trip wire" perimeters may be used, such as squares, zig-zag, or other layouts of pixels to determine target presence. Once determined, a group of pixels such as group **292** can be interrogated to get a better determination of target location.

FIG. **3**

Since many applications of the invention concern, or at least have present a human caused motion, or motion of a part of a human, or an object moved by a human, the identification and tracking problem can be simplified if the features of interest, either natural or artificial of the object provide some kind of change in appearance during such motion.

FIG. **3** illustrates tracking embodiments of the invention using intensity variation to identify and/or track object target datums. In a simple case, a subtraction of successive images can aid in identifying zones in an image having movement of features as is well known. It is also useful to add pixel intensities of successive images in computer **220** for example. This is particular true with bright targets (with respect to their usual surroundings) such as LEDs or retro-reflectors. If the pixels in use by the camera are able to gather light preferentially at the same time a special illumination light is on, this will accentuate the target with respect to background. And if successive frames are taken in this way, not only will a stationary image of the special target build up, but if movement takes place the target image then will blur in a particular direction which itself can become identify-able. And the blur direction indicates direction of motion as well, at least in the 2-D plane of the pixel array used.

Another form of movement can take place artificially, where the target is purposely moved to provide an indication of its presence. This movement can be done by a human easily by just dithering ones finger for example (if a portion of the finger such as the tip is the target in question), or by vibrating an object having target features of interest on it, for example by moving the object up and down with ones hand.

For example consider FIG. **3**A, where a human **301** moves his finger **302** in a rapid up and down motion, creating different image positions sequentially in time of bright target ring **320**, **320'** on his finger, as seen by camera **325**. If the camera can read quickly enough each of these positions such as **326** and **327** in image field **328** can be resolved, other wise a blur image such as **330** is registered on the camera and recorded in the computer **335**.

Instead of using ones finger, it is also possible to create movement of a target for example with a tuning fork or other mechanism mechanically energizing the target movement, on what otherwise might be a static object say. And it is possible for the human, or a computer controlling the movement in question to create it in such a manner that it aids identification. For example, a certain number of moves of ones finger (e.g., 4), or 2 moves/sec of ones finger, or horizontal moves of ones finger etc., any or all of these could indicate to the computer upon analysis of the camera image, that a target was present.

The invention comprehends this as a method for acquiring the datum to be tracked in the first place, and has provided a camera mechanism for tracking fast enough not to lose the data, assuming a sufficiently distinct feature. For example, it is desirable to not require sophisticated image processing routines and the like if possible, to avoid the time it takes to execute same with affordable equipment. And yet in many scenes, finding a target cant be done easily today without some aid, either a high contrast target (contrasting brightness or color or both, for example). Or the aid can be movement as noted, which allows the search for the target to be at least localized to a small region of the field of view, and thence take much less time to run, even if a sophisticated algorithm is employed.

FIG. **3**B illustrates an embodiment wherein a target which blinks optically is used. The simplest case is a modulated LED target such **340** on object **341** shown. Successive frames taken with camera **345** looking at pixel window **346** at **300** scans of the pixels within the window per second where the image **347** of the LED target is located, can determine, using computer **349** (which may be separate from, or incorporated with the image sensor), 5 complete blinks of target **340**, if blinked at a 60 hz rate. Both blink frequency, blink spacing, blink pulse length can all be determined if the scan rate is sufficiently faster than the blink rate, or pulse time.

It should be noted that if the target **340** is a retro-reflector as in FIG. **1**, with an illumination source such as **355** near the axis of the camera, then the LEDs (or other sources) of the illuminator can be modulated, causing the same effect on the target.

Somewhat more sophisticated is the situation shown in FIG. **3**C where a target **380** (on object **360**) illuminated by a light source **365** provides a time variant intensity change in the camera image **368** obtained by camera **370** as the target moves its position and that of the image. This can be achieved naturally by certain patterns of material such as herringbone, or by multifaceted reflectors such as cut diamonds (genuine or glass), which “twinkle” as the object moves. A relative high frequency “twinkle” in the image indicates then the presence of the target in that area of the image in which it is found.

When analog sensors such as PSD (position sensing diode) sensor **369** described in a copending application is used in addition to, or instead of a matrix array in camera **370**, the variation in light intensity or twinkle can be obtained directly from the detected output voltage from the signal conditioning of the sensor as shown in trace **375** corresponding to the movement of diamond target **380** a distance in the camera field. From the PSD one can also determine the position of the detected target image, theoretically at least independent of the intensity fluctuation.

For digital array detectors, the intensity variation can also be detected by subtracting images and observing the difference due to such variation. Such images need to be taken frequently if the twinkle frequency is high, and this can cause problems unless high speed camera scanning is possible. For example, in a twinkle mode, a pixel addressable camera using the invention herein could scan every 5^{th} pixel in both x and y. This would allow a 1000 frame per second operation of a camera which would normally go 40 frames per second. Such a rate should be able to capture most twinkle effects with the assumption that the light field changes on more than 25 pixels. If less, then scan density would need to be increased to every 3^{rd} pixel say, with a corresponding reduction in twinkle frequency detection obtainable.

FIG. **4**

FIG. **4**A illustrates identification and tracking embodiments of the invention using color and color change in a manner similar in some aspects to the intensity variation from object datums described above.

Color can be used as has been noted previously to identify a target, as can a change in color with time. For example, a target can change its color in order to identify itself to successive interrogations of pixels on a color TV camera. This can be accomplished by having a retro-reflector which is illuminated in succession by light from different colored LEDs for example, in the arrangement of FIG. **1**. For example red led **401** illuminates retro reflector target **405** on object **406** during frame **1** (or partial frame, if not all pixels addressed) taken by camera **410**. Then yellow led **402** illuminates target **405** on the next frame, and so forth. For any reading of successive frames, one point in the image will appear to distinctly change color, while all other points will be more or less the same due to the room lighting overwhelming the led source illumination and the natural color rendition of the objects themselves.

To return color variation when moved, one can employ a target which changes color naturally as it moves, even with illumination of constant color. Such a target can contain a diffractive, refractive, or interference based element, for example, a reflective diffraction grating for example, which splits white light illumination into colors, which are seen differently as the target moves and changes angle with respect to the observer and/or illumination source.

For example, consider FIG. **4**B showing reflective grating **440** on object **445** at initial position P. When illuminated by white light for example from lamp **450**, it reflects the spectrum such that when the object has moved to a new position P' the color (or colors, depending on the grating type, and angles involved) returning to camera **460** is changed. Such gratings can be purchased from Edmund Scientific company, and are typically made as replicas of ruled or holographic gratings.

Some types of natural features which change color are forms of jewelry which have different colored facets pointing in different directions. Also some clothes look different under illumination from different angles. This could be called then "color twinkle".

FIG. **5**

FIG. **5** illustrates special camera designs for determining target position in addition to providing normal color images. As was pointed out in a co-pending application, it may be desirable to have two cameras looking at an object or area one for producing images of a person or scene, the other for feature location and tracking. These may be bore-sighted together using beam splitters or the like to look at the same field, or they may just have largely overlapping image fields. The reason this is desirable is to allow one to obtain images of activity in the field of view (e.g., a human playing a game) while at the same time ideally determine information concerning position or other aspects of features on the human or objects associated with him.

It is now of interest to consider a matrix array chip equipped with a special color filter on its face which passes a special wavelength in certain pixel regions, in addition to providing normal color rendition via RGB or other filtering techniques in the remaining regions. The chip could be pixel addressable, but does not have to be.

Version FIG. **5**A

One version would have one special pixel filter such as **505**, for each square group of 4 pixels in an array **500** (one special pixel filter **505**, and 3 pixels, **510-512** filtered for RGB (red green blue) or similar, as is commonly used now for example. In one functional example, the special pixel **505** is purposely not read during creation of the normal image of a scene, but rather read only on alternate frames (or as desired) to determine target locations. If the array can be addressed pixel wise, the actual time lost doing this can be low. Since 25% of the pixels are effectively dead in forming the image in this example, and assuming all pixels are of equal area (not necessarily required), then 25% of the image needs to be filled in. This can be done advantageously in the image displayed, by making the color and intensity of this pixel the same as the resultant color and average intensity value of the other 3 in the cluster.

Version FIG. **5**B

In this version, related to FIG. **2** above, and shown in FIG. **5***b*, isolated pixels such as **530** (exaggerated in size for clarity) on array **531** or clusters of pixels such as **540-543**, are used to rapidly find a target with low resolution, such as round dot target image **550**. These pixels can ideally have special filters on their face, for example having near IR bandpass filters (of a wavelength which can still be seen by the camera, typically up to 1 um wavelength max). If takes only a few pixels to see the rough presence of a target, then in an image field of 1000×1000 pixels there could be one or more target images occupying 10×10 pixels or more. Thus in any group of 10×10, you could have 5 near IR filtered receptive pixels say, i.e., only 5% of the total pixel count but sufficient to see the IR targets location to a modest accuracy. Once found, one can also use the "normal" pixels on which the target image also falls to aid in more precise determination of its location, for example using pixel group **555** composed of numerous pixels.

In short by having a camera with certain pixels responsive to selected wavelengths and/or scanned separately one can very rapidly scan for target features, then when found, take a regular picture if desired. Or just take regular pictures, until the necessity arises to determine target location.

Similarly the special filtered pixels such as **505** or **530** could be laser wavelength bandpass filtered for this purpose, used by the array for preferentially detecting laser light projected on an object (while ignoring other wavelengths). In a normal image, such a pixel would be nearly black as little white light passes (except that centered on the laser wavelength). To provide a normal picture using such a camera, the special IR or laser wavelengths pixels readings would be filed in with values and colors of light from the surrounding regions.

Such a laser wavelength filter can be extremely effective, even if a relatively weak laser is used to illuminate a large area, especially where retro-reflectors are used, and the light returned is concentrated by 1000 times or more.

FIG. **6**

The embodiments above have dealt with finding just one target, and generally with just one camera, even though two or more cameras may be used for stereo imaging. Where stereo pairs of cameras are used, clearly each camera must see the target, if range via image disparity (the shift in location of a feature in the image in two camera views separated by a baseline) is to be determined.

Using the invention, one camera can be considered a master, the other a slave. The master camera determines target location by any of the means described above. Then the slave need only look at the expected pixel location of the target assuming some a priori knowledge of range which can come from previous target readings, or known range zones where the target has to lie in a given application.

Consider cameras **600** (master) with lens **603** and **601** (slave) having lens **604**, the axes of the two cameras separated by baseline **602** and with interfaced to computer **605**. The image of target **610** on object **615** is formed at position **620** on array **630** of camera **600**, and at position **621** on array **631** of camera **601**. The difference in position x in the direction of the baseline, in this simple situation is directly proportional to range z. The knowledge then of target image position **620** found by interrogating some or all of the pixels of camera **600** can as mentioned be used to more rapidly find image **621** in the image field of the "slave" camera **601**, and thus the z location of the target **610**.

For example if range is known to be an approximate value of z, one can look in the image field of the camera **601** along a line of points at a calculated value x away from the edge of the field, assuming **620** has been found to lie as shown near the corresponding edge of the field of camera **600**.

Two or more cameras may be used for stereo image analysis including object range and orientation data as discussed in FIGS. **1** and **6**. Range can also be determined via triangulation with a single camera and one target if projected on to the object in question at an angle to the camera axis from a laser say, or by using a single camera and 3 or more points on an object whose relative relationship is known (including the case of a line of points and an external point).

FIG. **7**

As stated above, the TV camera of the invention can be used to see either natural or artificial features of objects. The former are just the object features, not those provided on the object especially for the purpose of enhancing the ability to determine the object location or other variable using computer analysis of TV camera images. Such natural features, as has been pointed out in many of the co-pending referenced applications, can be holes, corners, edges, indentations, protrusions, and the like of fingers, heads, objects held in the hand, or whatever.

But using simple inexpensive equipment it is often hard to determine the presence or location of such features in a rapid reliable enough manner to insure function of the application in question. In this case, one can employ one or more artificial features, provided on the object by attaching a artificial target onto the object, or manufacturing the object with such a target.

At least three types of artificial features can be employed.

1. The first is to provide special features required for object location, or orientation determination. Such a special feature can be of an optically contrasting material at the wavelength used to that of the object, for example a bright color, or a retroreflector;
2. The second is to provide one artificial feature (typically capable of more easily being found in an image than natural features of the object), and by finding it, localize to the region of that target environs the problem of finding any other features needed nearby; and
3. The third is to find an artificial feature on an object that actually by its shape, location, or coded features, provides a guide to the location of natural or other artificial features which are to be sensed in order to determine position or orientation of the same or related objects. This has been dubbed by me a co-target in co-pending applications incorporated by reference.

As shown in FIG. **7**, object **700** has co-target **701** at one end, visible to camera **705**. The co-target in this particular instance is a diamond shape, and is of high contrast for easy acquisition. For example it could be a yellow plastic retroreflector formed of molded corner cubes similar to those used on cars for taillights and other safety purposes.

The diamond shape in this case is significant for two reasons. First it is unusual relative to the object or background when used in the context intended, and makes the target still more identifiable (that is novel color, shape and brightness are all present). In addition, in this particular instance it has been chosen that a diamond shape, should indicate that the corners of the object are to be used for 6 axis position and orientation determination and that the choice of color for example, signifies that the object corners are within some predetermined distance from the target. If desired the target location on the object can also point to the corners. For example, in the drawing, the four corners of the diamond, **720-723**, point in the general direction of the four corners **730-733** of the rectangular object **700**.

FIG. **8**

The invention herein and disclosed in portions of other copending applications noted above, comprehends a combination of one or more TV cameras (or other suitable electro-optical sensors) and a computer to provide various position and orientation related functions of use. It also comprehends the combination of these functions with the basic task of generating, storing and/or transmitting a TV image of the scene acquired either in two or three dimensions.

FIG. **8**A illustrates control of functions with the invention, using a handheld device which itself has functions (for example, a cell phone). The purpose is to add functionality to the device, without complicating its base function, and/or alternatively add a method to interact with the device to achieve other purposes.

The basic idea here is that a device which one holds in ones hand for use in its own right, can also be used with the invention herein to perform a control function by determining its position, orientation, pointing direction or other variable with respect to one or more external objects, using an optical sensing apparatus such as a TV camera located externally to sense the handheld device, or with a camera located in the handheld device, to sense datums or other information external for example to the device.

This can have important safety and convenience aspects to it, particularly when the device is used while driving a car or operating other machinery. To date voice recognition has been the only alternative to keying data in to small handheld devices, and voice is limited in many cases very limited if some physical movement is desired of the thing being communicated with.

A cellular phone **800** held in the hand of a user can be used to also signal functions in a car using a projected laser spot from built in laser spot projector **801** as in FIG. **14**, in this case detected by detector **802** on the dashboard **803**. Alternatively and or in conjunction, one may use features such as round dot targets **805-807** on the cell phone which are sensed, for example, by a TV camera **815** located in the car headliner **816** or alternatively for example in the dashboard (in this case the targets would be on the opposite end of the cell phone). More than one set of targets can be used, indeed for most generality, they would be an all sides which point in any direction where a camera could be located to look at them.

Remote control units and dictating units are also everyday examples of some devices of this type which can serve control purposes according to the invention. One of the advantages here is that it keeps the number of switches etc on the device proper to a minimum, while allowing a multitude of added functions, also in noisy environments where voice recognition could be difficult or undesirable for other reasons.

Use of specialized target datums or natural features of devices held in the hand, or used with cameras on such devices, allows photogrammetric techniques such as described in FIG. **1** to be used to determine the location in 6 degrees of freedom of the device with respect to external objects.

As one illustrative example, to signal a fax unit **824** in the car to print data coming through on the phone, the user just points (as illustrated in position **2**) the cell phone toward the fax, and the TV camera **815** scans the images of targets **805-807** on the face toward the camera, and the computer **830** connected to the camera analyzes the target images (including successive images if motion in a direction for example is used as an indicator, rather than pointing angle for example), determines the cell phone position and/or orientation or motion and commands the fax to print if such is signaled by the cell phone position orientation or motion chosen. The knowledge in space of the cell phone location and its pointing direction (and motion as pointed out above) provides information as to the fact that the fax was the intended target of the effort. Such data can be taught to the system, after the fact even if the fax or any other item desired to be controlled is added later.

Another version has a camera and requisite computer (and or transmission capability to an external computer) in the handheld device, such as a cell phone or whatever. When pointed at an object, the camera can acquire the image of the object and/or any natural features or special datums on the object which are needed to perform the function desired.

One function is just to acquire an image for transmission via for example the cell phones own connection. This is illustrated in FIG. **8**B, where an image of object **849** acquired by camera **850** of cell phone **851** held by user **852** is transmitted over mobile phone link **853** to a remote location and displayed, for example. While this image can be of the user, or someone or something of interest, for example a house, if a real estate agent is making the call, it is also possible to acquire features of an object and use it to determine something.

For example, one purpose is recognition, for example one can point at the object, and let the computer recognize what it is from its TV image. Or point around in space taking multiple TV frames aiming in different directions, and when computer recognition of a desired object in one of the images takes place, transmit certain data to the object. Or it can be used to acquire and transmit to remote locations, only that data from recognized objects.

Thus the invention can provided on a hand held object for a variety of purposes,

To take images of things;

To determine datums on things; and

To automatically read things.

The combination of any or all of these functions in addition with other object functions such as hand held cell phones, dictation units, telephones, wearable computer devices and the like.

An alternative, shown with phantom lines in FIG. **8**A, to the some aspects of the above described operation of the embodiment is to use a laser pointer **801** in for example a cell phone to designate say the fax machine as shown. Then the TV camera **815** simply detects the presence of the laser pointer projected spot **820** on the fax, and via computer memory it is known that this is a device to be energized or connected in connection with the cell phone.

The camera located in a handheld device can also be used to point at a TV screen, such as that on the dashboard of a car, and to utilize data presented there for some purpose. For example, if pointed at a screen saying email message number 5, the camera of the device can be used to obtain this image, recognize it through known character recognition techniques, and process it for transmission if desired. Or it might just say the message to the user of the phone through the speaker of the cell phone. Such a technique is not required if means exist to directly transmit the incoming information to the cell phone, but this may not be possible.

FIG. **9**

FIG. **9** illustrates pointing at a displayed image of an object represented on a screen using a finger or laser pointer, and then manipulating the represented object or a portion thereof using the invention. For example, consider user **901** pointing a laser pointer **905** at an image generated by computer **910** on display **912**, typically a large screen display (e.g., 5 feet diagonal or more) where control features here disclosed are of most value.

The user with the pointer, can point to an image or portion of the displayed image to be controlled, and then using the action of the pointer move the controlling portion of the image, for example a "virtual" slider control **930** projected on the screen whose lever **935** can be moved from left to right, to allow computer **910** sensing the image (for example by virtue of TV camera **940** looking at the screen as disclosed in copending applications) to make the appropriate change, for example in the heat in a room.

Alternatively one can also point at the object using ones fingers and using other aspects of the invention sense the motions of ones fingers with respect to the virtually displayed images on the screen, such as turning of a knob, moving of a slider, throwing a switch etc.

Such controls are not totally physical, as you don't feel the knob, so to speak. But they are not totally virtual either, as you turn it or other wise actuate the control just as if it was physical. For maximum effect, the computer should update the display as you make the move, so that you at least get visual feedback of the knob turning. You could also get an appropriate sound if desired, for example from speaker **950**, like an increase in pitch of the sound as the knob is "moved" clockwise.

FIG. **10**

The above control aspects can in some forms be used in a car as well even with a small display, or in some cases without the display.

Or it can be a real knob which is sensed, for example by determining position of a target on a steering wheel or the fingers turning it tracked (as disclosed in co-pending application references).

For example, consider car steering wheel rim **1000** in FIG. **10**A. In particular, consider hinged targeted switch, **1010** (likely in a cluster of several switches) on or near the top of the wheel, when the car is pointed straight ahead, and actuated by the thumb of the driver **1011**. A camera **1020** located in the headliner **1025**, and read out by microcomputer **1025** senses representative target **1030** on switch **1010**, when the switch is moved to an up position exposing the target to the camera (or one could cover the target with ones fingers, and when you take a finger off, it is exposed, or conversely one can cover the target to actuate the action).

The camera senses that target **1010** is desired to be signaled and accordingly computer **1025** assures this function, such as turning on the radio. As long as the switch stays in the position, the radio is on. However other forms of control can be used where the switch and target snap back to an original position, and the next actuation, turns the radio off. And too, the time the switch is actuated can indicate a function, such as increasing the volume of the radio until one lets off the switch, and the target is sensed to have swung back to its original position and the increase in volume thus terminated.

In operating the invention in this manner, one can see position, velocity, orientation, excursion, or any other attribute of actuation desired. Because of the very low cost involved in incremental additions of functions, all kinds of things not normally sensed can be economically provided. For example the position of a datum **1040** on manually or alternatively automatically movable plastic air outlet **1041** in the dashboard **1042** can be sensed, indicative of the direction of airflow. The computer **1025** can combine this with other data concerning driver or passenger wishes, other outlets, air temperature and the like, to perfect control of the ambiance of the car interior.

It is also noted that the same TV camera used to sense switch positions, wheel position, duct position, seat position (for example using datum **1045**), head rest position (for example using datum **1046**), and a variety of other aspects of physical positions or motions of both the car controls and the driver or passengers. And it can do this without wires or other complicating devices such as rotary encoders which otherwise add to the service complexity and cost.

When the camera is located as shown, it can also see other things of interest on the dashboard and indeed the human driver himself, for example his head **1048**. This latter aspect has significance in that it can be used to determine numerous aspects such as:

1. The identity of the driver. For example, if a certain band of height isn't reached, such as point P on the drivers head, the ignition can be interlocked. Much simpler than face recog-

nition, but effective if properly interlocked to prevent repeated retries in a short time period.

2. The position of the head of the driver in case of an accident. As detailed in reference **4**, a camera or cameras can be used to determine head location, and indeed location of the upper torso if the field of view is large enough. This information can be used to control airbag deployment, or head rest position prior to or during an accident (noting too that headrest position can also be monitored without adding any hardware). Particularly of interest is that the pixel addressing camera of the invention can have the frequency response to be useful in a crash, sensing the movement of the person (particularly severe if unrestrained) within a millisecond or two, and providing a measure of the position for airbag deployment. Additional cameras may also be used to aid the determination, by providing other views or observing other features, for example.

Using a pixel addressing camera for camera **1020** confers additional advantages. For example consider the image of the car interior produced by the camera lens **1021**, on matrix of pixels **1061**, whose addressing and processing is controlled by computer **1025**. In the first instance one can confine the window of view of a certain group of pixels of the total matrix **1061** to be only in the region of the steering wheel, as in window **1065** shown. This allows much faster readout of the more limited number of pixels, and thus of the steering wheel switches, at the expense of not seeing anywhere else in that particular reading. But this may be desirable in some cases, since it may only be required to scan for heater controls or seat positions, every 10 seconds say, while scanning for other more immediate items a hundred times per second or more. A good example are safety related functions. 5 per second might suffice for seeing where the turn signal or windshield washer control was, as an example. Window 1066 dotted lines is illustrative of a window specialized for head, headrest and seat positions, say.

Scans in certain areas of the image can also depend on information obtained. For example one may initiate a scan of a control position, based on the increasing or decreasing frequency of an event occurrence. For example if the persons head is in a different location for a significant number of scans made at 15 second intervals for example, then in case of a crash, this data could be considered unreliable. Thus the camera window corresponding to pixels in the zone of the head location **1048** could be scanned more frequently henceforward, either until the car stopped, or until such action settled down for example. Such action is often the case of a person listening to rock music, for example.

Similarly, if someone is detected operating the heater controls, a scan of predominately heater function controls and related zones like air outlets can be initiated. Thus while normal polling of heater controls might be every 2 seconds say, once action is detected, polling can increase in the window(s) in question to 40 times per second for example. The detection of action can be made first via the camera, or via input from some other input device such as a convention heater knob and electric circuit operable therewith.

Scans in certain areas of the image can also depend on information obtained in other areas of scan, or be initiated by other control actions or by voice. For example, if hard de-acceleration was detected by an accelerometer, but before a crash occurred, the camera could immediately be commanded to begin scanning as fast as possible in the region of the image occupied by the driver and/or any other humans in its field of view. This would be for the purpose of monitoring movements in a crash, if a crash came, in order to deploy an airbag for example.

One might utilize the invention to actuate a function, based on positions of people or other objects in the vehicle. As one example, suppose the drivers hand is resting on a console mounted gear lever. By scanning the image of this region, one can determine from the image the position of the console shift lever, and use the image thereof to control gear change via computer **1025**. However if the driver rests his hands on the windshield wiper stalk, it could in the same manner, become a column mounted gear lever so to speak. Or just be used for up down gear changes, like a paddle shifter on a racing car. In fact in the latter sense, the camera could be instructed to detect ones finger or hand movement to do this function for example, wherever one desired to rest ones hand (within the camera field of view at least). This function is also useful for physically disabled persons wishing to drive the car. And it can be different for different persons as well, via programming of the control functions associated with any given hand, switch or other position or movement.

FIG. **10**B illustrates alternative types of control mechanisms which can be used with the invention, in this case illustrated on the steering wheel of a car, although as can be appreciated, any suitable function or location may be used or created. And too, combinations of functions can be used. The invention is generic to car steering wheel controls, dishwashers, audio systems in ones home, heating and air conditioning elements and virtually all other forms of human related control functions. The key is that the camera computer combination makes a very inexpensive way to share a wide variety of functions with one or just a few basic systems and over a large population base.

As shown in FIG. **10**B, the steering wheel **1070** has two additional types of controls visible to camera **1020** and able to be sensed and generate the appropriate control function via computer. These are rotating device **1072** built to rotate around the steering wheel rim circular cross section, and expose a continuously variable, or digital or step wise increment component to the camera. For example, three bars are shown, short **1075**, medium **1076**, and long **1077**. The computer senses which of the three is visible by comparing the length to pre-stored values (or taught values, see below), and causes the desired action to occur.

The second control **1080** is a sliding device **1081** which can be slid clockwise, or counterclockwise along a circumferential section of the steering wheel at the top, sides or wherever. As before, Its position is determined by camera **1020** again providing more data than just a switch up or down as shown before.

While illustrated on the steering wheel where it is readily at hand, it can be appreciated that the position of either the slider **1081** or the rotary device **1072**, or other similar devices for the purpose at hand could be elsewhere than the wheel, for example on stalk or on a piece of the dash, or other interior component indeed wherever a camera of the invention can view them without excessive obscuration by persons or things in the car. It need not be on a car either, controls of this type can be in the home or elsewhere. Indeed a viewable control datum can even be on a portable component such as ones key chain, phone, or article of clothing apparel, or whatever. Similarly the camera **1020** can view these items for other purposes as well.

The teach-ability of the invention is achieved by showing the camera the code marker in question (e.g., a short bar located on the wheel), and in the computer recording this data along with what it is supposed to signify as a control function for example, turn rear wiper on to first setting. This added functionality of being easily changed after manufacture is an important advantage in some cases, as for example, today after-market addition of wired in accessories is difficult.

Games Using the Invention

The co-pending referenced applications have described games which can be played with target sensing and touch screen based devices, typically but not necessarily, electro-optically based (e.g., TV camera). The cameras of the invention can be used to, for example: Sense the player or players in the game or portions thereof; sense objects held or manipulated by the players (e.g., a ball, a pistol); sense physical tokens used in the game, such as monopoly game tokens; and sense game accessories such as checkerboards, croquet wickets; compare positions of objects with respect to other objects or players.

In addition, the cameras can be used to take images which can be displayed also a major feature given the ability to create life size displays. And the computer of the invention can be used to control the presentation of background image data from stored images, or even images downloaded from the internet for example.

Some or all of these aspects will now be illustrated in some representative game illustrations (again noting that some more are in the co-pending applications).

FIG. **11** Board Game

Even today, popular board games such as Monopoly and the like are being provided in computer playable form, with the "board" represented on the screen of the computer monitor. The invention here builds on this by providing various added features which allow a physical nature of the game just as the real game, but with new aspects and providing physical game play which can be transmitted over the internet to others. These features also can be turned off or on at as desired.

In one version shown in FIG. **11**A, the player tokens such as **1101** and **1102** are observed by camera of the invention **1110** placed directly overhead of the play board **1115**, which can for example be a traditional monopoly board (chess board, checker board, etc). points on the board such as corners **1130**, **1131**, **1132**, and **1133** can also be observed to establish a reference coordinate system for the computer **1140** to track the moves of the markers, either from their natural features, or from specialized datums thereon (e.g., retro-reflective hat top **1141** on marker **1101**). For example a train shape **1102** of a marker can be called from memory, or taught to the computer by showing it to the camera. Rotation invariant image analysis programs such as the PATMAX program from Cognex company can be used to identify the marker in any normal orientation, together with its location on the board (the board itself can be taught to the computer using the camera, but is preferably called up from memory).

The board position and relative scale in the field of view is determined easily by knowing the spacing of the corner points **1130-1133** and using this to calibrate the camera (to provide extra contrast, the corners can have retro-reflective glass bead edging or beading as shown). For example if the points are spaced 20 inch on corners of the board, and the camera is positioned so that 20 inches occupies 80% of its field of view, then the field of view is 25 inches square (for a square matrix of camera pixels), and each pixel of 1000 pixels square, occupies 0.025 inches in the object field.

The play of both players (and others as desired) can be displayed on the monitor **1150**, along with an image of the board (which also can be called from computer memory). But other displays can be provided as well. For example to lend more realism to the game, the display (and if desired sound from speaker **1155** connected to computer **1140**) can also be programmed to show an image or sound that corresponds to the game. For example, when the camera image has provided information that one player has landed on "Boardwalk" (the most valuable property) a big building could be caused to be shown on the screen, corresponding to it also suitable sounds like wow or something provided).

The camera can be used to see monopoly money (or other game accessories) as well, and to provide input so the computer can count it or do whatever.

A large, wall sized for example, screen can add added realism, by allowing one to actually get the feeling of being inside the property purchased, for example.

One of the exciting aspects of this game is that it can be used to turn an existing board game into something different. For example, in the original monopoly the streets are named after those in Atlantic City. By using the computer, and say a DVD disc such as **1160** stored images of any city desired can be displayed, together with sounds. For example, one could land on the Gritti Palace Hotel in Venice, instead of Boardwalk. As shown in FIG. **11**B, the TV camera senses the image of train marker **1101**, and conveys this information to computer **1140**, which causes the display **1150** and speaker of the invention to display the information desired by the program in use.

Making the game in software in this way, allows one to bring it home to any city desired. This is true of a pure (virtual) computer game as well, where the board only exists on the computer screen.

For added fun, for example in a small town context, local stores and properties could be used, together with local images, local personages appearing on the screen hawking them, and the like. A local bank could be displayed to take your money, (even with sounds of the local banker, or their jingle from the radio) etc. This makes the game much more local and interesting for many people. Given the ease of creating such local imagery and sounds with cameras such as digital camcorder **1151** used as an input of display imagery (e.g., from local celebrity **1158**) to the game program, one can make any monopoly experience more interesting and fun at low cost.

The same holds true with other well known games, such as Clue, where local homes could be the mystery solving location, for example. One can also create games to order, by laying out ones own board. If one of the persons is remote, their move can be displayed on the screen **1150**.

In the above, the display has been treated as sort of backdrop or illustration related. However, one can also create a whole new class of games in which the display and/or computer and the board are intertwined. For example as one takes a trip around the monopoly board, several chance related drawings opportunities occur during play. In this new game, such could be internet addresses one draws, which, via modem **1152**, send the board game computer **1140** to any of a large number of potential internet sites where new experiences await, and are displayed in sight and sound on the display.

It should also be noted that the board can be displayed on the screen as well, or alternatively projected on a wall or table (from overhead). A particularly neat mixture of new and old is shown in FIG. **11**B, where the board is displayed on a screen pointed vertically upward just as it would be on a table, and indeed in this case physically resident on a table **1165**. The board is displayed (from software images or cad models of the board in computer **1166**) on a high resolution table top HDTV LCD screen **1167** with a suitable protective plastic shield (not shown for clarity). Play can proceed just as before using physical tokens such as **1101** and **1102**. In this case the display used to augment the game can actually be shown on the same screen as the board, if desired.

The TV camera **1110** in this context is used to see the tokens and any other objects of the game, the people as desired, and the play, as desired. The camera can be used to see the display screen, but the data concerning the board configuration displayed may be best imputed to the computer program from direct data used to create the display.

A beauty of the invention is that it allows the interaction of both computer generated images and simulations, with the play using normal objects, such as one might be accustomed to for example, or which give a "real" feel, or experience to the game.

FIG. **12** Sports Game

FIG. **12** illustrates a generic physical game of the invention using points such as **1201-1205** on the human (or humans) **1210** sensed by a TV camera such as stereo camera pair **1215** and transmitted to the computer of the invention **1220**. While points can be sensed in 2D, this illustration uses as stereo camera pair located on large screen display **1225** as shown to provide a unitary package built into the screen display (pointed out in other co-pending applications). In this particular instance a 3D display is illustrated, though this isn't necessary to obtain value and a good gaming experience. The human optionally wears red and green filter glasses **1235** such that red images on the screen are transmitted to one eye, green to another, so as to provide a 3D effect. Similarly crossed polarized filter glasses (with appropriate display), and any other sort of stereoscopic, or autosteroscopic method can also be used, but the one illustrated is simple, requires no connecting wires to the human, and can be viewed by multiple uses, say in a gym aerobics room.

The game is generic, in that it totally depends on the program of the computer. For example, it can be an exercise game, in which one walks on a treadmill **1250**, but the image displayed on screen **1225** and sound from speakers **1255** and **1266** carry one through a Bavarian forest or the streets of New York as one walks, for example.

Or it can be a parasail game in which one flies over the water near Wakiki beach, with suitable images and sounds. In any case action determined by sensing position, velocity acceleration, or orientation of points **1201-1206** on the player, **1210** is converted by computer **1220** into commands for the display and sound system. Note in the figure this player is shown viewing the same screen as the treadmill walker. This has been shown for illustration purposes, and it is unlikely the same game could be applied to both, but it is possible.

It is noted that fast sensing, such as provided by the pixel addressing camera method disclosed above is highly desirable to allow realistic responses to be generated. This is especially true where velocities or accelerations need to be calculated from the point position data present in the image (and in comparison to previous images).

For example, consider points **1201** and **1202** on player **1210**. If point **1201** moves to **1201***a*, and **1202** moves to **1202***a* indicative of a quick jerk movement to turn the displayed parasail, this movement could occur in a 0.1 second. But the individual point movements to trace the action would have to be sensed in 0.01 second or quicker for example to even approximately determine the acceleration and thus force exerted on the glider, to cause it to move.

It is important to note that the invention is not only generic in so far as the variety of these games are concerned, but it also achieves the above with virtually no mechanical devices requiring maintenance and creating reliability problems which can eliminate profits from arcade type businesses especially with ever more sophistication required of the games themselves.

FIG. **13** Bar Game

FIG. **13** illustrates a game which is in a class of gesture based games, in which the flirting game of FIG. **15** is also an example. In such games one senses the position, velocity or acceleration of a part of a person, or an object associated with the person. This can also include a sequence of positions, itself constituting the gesture. The detected data is then related to some goal of the contest. Consider FIG. **13**, wherein the object in ones hand is monitored using the invention, and a score or other result is determined based on the position, velocity, orientation or other variable of the object determined. For example, in a bar one can monitor the position, orientation, and rate of change thereof of drinking glasses.

A two person game is illustrated, but any reasonable number can play as long as the targets can all be tracked sufficiently for the game (in one test over 200 targets were acquired, but as can be appreciated this uses most of the field of view of the camera, and thus speed improvements made possible by pixel addressing become more difficult.

As shown, a single camera **1301** observes one or more targets such as **1305** on glass **1310** held by contestant **1315**, and target **1320** on glass **1325** of contestant **1330**. On a signal, each drinks, and a score is calculated by program resident in computer **1350** based on the time taken to raise the glass, and place it back empty on table **1355**. A display of the score, and an image desired, for example of the winner (taken with camera **1301** or another camera), or a funny image called from computer memory, is displayed on monitor display **1370**.

If the glass features are sufficiently distinct for reliable and rapid acquisition and tracking, for example as might be provided by an orange color, or a distinct shape, then specialized target features are not required.

Alternatively the velocity, path of movement of the glass (or other object), acceleration, or any other variable from which target data is sufficient to calculate, can be used to determine a score or other information to be presented or used.

FIG. **14**

The referenced co-pending applications have described a game where by laser pointers can be used to designate images on a TV screen. In this case of FIG. **14**A, the TV camera of the invention such as **1410** is used in a two player game to see laser pointer spots such as **1420** and **1421** projected by players **1430** and **1431** respectively, using laser pointers **1440** and **1441** respectively. When one player's spot hits the other, the event is recorded in memory of computer **1450** for further analysis and display.

In a somewhat different context, a person can use a laser pointer to point at an object to designate it for some purpose, for example for action. For example consider FIG. **14**B, in which housewife **1460** who points with laser pointer **1462** so as to provide a laser spot **1465** on dishwasher **1470**. TV camera of the invention **1475** in corner of the kitchen **1480** picks up all laser spots in an image of the room (made easier to process in terms of signal to background imagery if one locates a laser wavelength band-pass interference filter **1481** in front of the TV camera as shown) and compares via computer **1483**, the location of the spot detected in the image to stored memory locations of objects such as the dishwasher **1470** or fridge **1485** in the camera field of view, so as to identify the object needing action. In this case too, housewife may signal via a spatially variant laser pointer projection image (see copending referenced applications for further examples in other applications), or a series of spots in time, what action is desired, for example to turn the washer on. In this case the computer **1483** can cause a command to do so to be sent to the washer.

Any one with a simple laser pointer can make these commands effective. No learning is needed just point at the item desired, with the TV camera and computer of the invention acquiring the data and interpreting it. This is much simpler than remote controls of today, and a major advantage for those who have difficulty or inclination to learn complex electronic devices and procedures. It should be noted that these pointing procedures can easily be combined with voice recognition to further define the desired control activity for example inputting the housewife's voice in this example by virtue of microphone **1476**.

The stored locations can be taught. For example in a setup mode, one can point a laser pointer at the dishwasher, and indicate to the computer that that spot is the dishwasher. The indication can be provided by keyboard, voice recognition or any other means that is satisfactory.

Clearly other items can be monitored or controlled in this manner. The camera can also detect optical indications provided by other means, for example lights in the appliance itself. And one can detect whether light have been left on at night (or not left on) and cause them to be turned off or on as desired.

Such a camera if it is responsive to normal illumination as well as that of the laser wavelength, can also be used to see movements and locations of people. For example, it can look at the top of the stove, and assure that no movement is near the stove **1486**, or objects on it if programmed to do so, thus sounding an alarm if an infant should get near the stove, for example.

The housewife in the kitchen can also point at a board on which preprogrammed actions are represented. For example consider board **1490**, shown in greater detail in FIG. **14**C, in which 3 squares **1491**-**1493** are to represent different functions. Thus if **1491** is programmed (via keyboard, voice or whatever) to represent turning on the clothes dryer in the laundry, when the TV camera sees, and via the computer, identifies spot **1496** projected by the user on square **1491**, it causes the dryer to turn on. Operated in this manner, the board **1490**, in combination with a TV camera of the invention (such as **1475** or a more dedicated one for the board alone) and computer such as **1483** can be considered a form of touch screen, where the user, in this case in the kitchen can point at a portion of the board with a finger, or a laser pointer, and register a choice, much like touching an icon on a conventional computer touch screen.

Similarly, squares or other zones representing choices or the like can be on the item itself. For example, a stove can have four areas on its front, which can be pointed at individually for control purposes, what ever they are (e.g., representing heat settings, burner locations or the like). For security, it could be that only a coded sequence of laser pulses would be seen, or as pointed out in co-pending reference Ser. No. 60/133,673, a spatial code, for example representing the user such as an initial could be projected, and sensed on the object by the TV camera.

The laser pointer can be held in the hand of the user, or, like **1497** attached for example to a finger, such as forefinger **1498**. Or it can be on or in another object, desirably one which is often hand held in the normal course of work, such as a TV remote control, a large spoon, or the like. Or using other aspects of the invention, the finger of the user can be observed to point directly, and the object being pointed at determined. For example if finger **1498** is moved 4 times, it could indicate to the TV camera and thence computer that channel four was desired on a TV display not shown.

If a special pointer is used, it can be any workable optical device, not necessarily a laser. The camera and computer of the invention can also be used to observe the user pointing directly, and compute the pointing vector, as has been described in my co-pending applications.

FIG. **15** A "Flirting" Game

Another game type is where the camera looks at the human, and the humans expressions are used in the game. In this case it is facial expressions, hand or body gestures that are the thing most used.

For example, one idea is to have a scene in a restaurant displayed on a display screen **1500**, preferably a large HDTV screen or wall projection to be as lifelike as possible, and preferably life size as well which lends extra realism to some games, such as this one due to the human element involved.

Let us consider that seated at the table in the restaurant displayed on the screen is a handsome man **1501** whose picture (likely a 3D rendered animation, or alternatively photo-imagery called from memory), and the goal for the girl **1510** playing the game is to flirt with this man until he gets up and comes over to say hello, ask her out or what ever (what he does, could be a function of the score obtained, even!).

Player **1510** seated at table **1511** (for authenticity, for example) is observed by TV camera **1515** (or stereo pair as desired, depending whether 3D information is thought required) and computer of the invention **1520**, which through software determines the position of eyebrows, lips, hands, fingers and any other features needed for the game. If necessary, specialized targets can be used as disclosed herein and elsewhere to augment this discrimination, for example such as optically contrasting nail polish, lipstick, eyeliner or other. Contrast can be in a color sense, or in a reflectivity sense such as even retro-reflective materials such as Scotchlite 7615 by 3M company. Even special targets can be used to enhance expressions if desired.

This can be a fun type game, as the response of the displayed person can be all kinds of things even contrary to the actual gestures if desired. Sounds, such as from speaker **1530** can also be added. And voice recognition of players words sensed by microphone **1550** can also be used, if verbal as well as expressive flirting is used.

While the game here has been illustrated in a popular flirting context, it is more generally described as a gesture based game. It can also be done with another contestant acting as the other player. And For example, the contestants can be spaced by the communication medium of the internet. The displayed characters on the screen (of the other player) can be real, or representations whose expressions and movements change due to sensed data from the player, transmitted in vector or other form to minimize communication bandwidth if desired.

Other games of interest might be:

"Down on the Farm" in which a farmer with live animals is displayed on a life size screen, and the children playing the game are to help the farmer by calling the animals to come over to them. This would use recognition of voice and gesture to make the animal images move and make sounds.

A player can find someone in a display and point at him, like the "Whereas Waldo" puzzle game. Then the subject moves, child runs to peek at him, and to find him, say running down a street whose image is displayed on the screen.

One can also use the camera of the invention to monitor the progress made by a child building blocks, and show an Video displayed image of a real skyscraper progressing as he builds his little version. Note the benefit of group activity like a board game and children's play with each other.

FIG. **16**

FIG. **16** illustrates a version of the pixel addressing camera technique wherein two lines on either side of a 1000 element square array are designated as perimeter fence lines to initiate tracking or other action.

Some "pixel addressing" cameras such as the IVP MAPP 2500 512×512 element camera, are smart, that is can process on the same chip. However, in some cases the control of such a camera may not allow one to actually read just one pixel, say, but rather one must read the whole line on which the pixel rests. Now some processing can be in parallel such that no speed is lost, at least in many instances.

If however, one does have to read a whole line serially into a computer portion, then to fully see a 10×10 pixel round target say, one would have to read at least 10 lines.

If two targets both were located on the same lines, the time involved to read would be the same.

In the same vein, if lines of data must be scanned, then the approach of **2***b* wherein every 20^{th} pixel say is interrogated can be specialized to having such pixels fall on scan lines wherever possible. And where one is restricted to reading all pixels on a scan line and where a target entry zone is anticipated, one can have a scan line oriented to be crossed by such entry. For example in FIG. **16**, the two lines **1601** (line of pixels **3**) and **1602** (line of pixels **997**) of a 1000×1000 element pixel array **1610** are designated as perimeter fence lines, to trigger a target tracking or other function on the entry of a target image on to the array, such as **1615** from either the right or left side in the drawing. This is often the case where entry from top or bottom is precluded by constraints of the application, such as a table top at the bottom, or the height of a person at the top. Or in a stereo example such as FIG. **6**, the baseline defines the direction of excursion of a target as z is varied again calling for crossing of scan lines out of the plane of the paper at some point.

The invention herein has provided an exciting method by which common board games can become more fun. The invention provides a link with that past, as well as all of the benefits of the video and computer revolution, also via the internet.

It is envisioned that the same approach may be applied to many card games as well. It is also thought that the invention will find use in creating ones own games, or in downloading from the internet others creations. For example, common everyday objects can become the tokens of the games, and taught to the game computer by presenting them to the video camera. Similarly, the people playing the game can be taught, including their names and interests.

FIG. **17**

FIG. **17** illustrates a 3D acoustic imaging embodiment of the invention which at low cost may generate accurate 3D images of the insides of objects, when used in conjunction with ultrasonic transducers and particularly a matrix array of ultrasonic transducers.

As shown in FIG. **17**A, the position in xyz of the ultrasonic imaging head **1700** on wand **1701** held in a users hand **1702** is monitored electro-optically as taught in FIG. **1**, using a single camera **1710** and a simple four dot target set **1715** on the head **1700** at the end of the transducer wand **1701** in contact with the object to be examined **1720**. Alternatively, as also taught in FIG. **1**, a stereo pair for example providing higher resolution in angle can be employed.

Computer **1725** combines ultrasonic ranging data from the ultrasound transducer head **1700** and from the sensor of transducer location (in this case performed optically by camera **1710** using the optically visible targets on the transducer head) in order to create a range image of the internal body of the object **1720** which is thus referenced accurately in space to the external coordinate system in the is case represented by the camera co-ordinates xy in the plane of the TV camera scan, and z in the optical axis of the camera.

In many cases it is also desirable to know the pointing angles of the transducer. One instance is where it is not possible to see the transducer itself due to obscuration, in which case the target may alternately be located at the end **1704** of the wand for example. Here the position and orientation of the wand is determined from the target data, and the known length of the wand to the tip is used, with the determined pointing angle in pitch and yaw (obtained from the foreshortening of the target spacings in the camera image field) to calculate the tip position in space.

This pitch and yaw determination also has another use however, and that is to determine any adjustments that need to be made in the ultrasonic transduction parameters or to the data obtained, realizing that the direction of ultrasound propagation from the transducer is also in the pointing direction. And that the variation in ultrasound response may be very dependent on the relation of this direction **1730** with respect to the normal **1735** of the surface **1736** of the object (the normal vector is shown for clarity pointing inward to the object).

The difference in direction can be calculated by using the TV camera (which could be a stereo pair for greater angular resolution) as well to determine the surface normal direction. This can, for example, be done by placing a target set such as **1740** on the surface in the field of the camera as shown. This can be dynamically or statically accomplished using the photogrammetric method described in the Pinkney references.

Differences in direction between the surface normal and the transducer pointing direction are then utilized by software in the computer **1725** of the invention in analysis of the ultrasound signals detected. The pointing angle and the position of the transducer on the surface of the object are used by the computer in predicting the location of various returns from internal points within the object, using a suitable coordinate transformation to relate them to the external coordinate reference of the TV camera.

All data, including transducer signals and wand location is fed to computer **1725** which then allows the 3D image of the inside of the body to be determined as the wand is moved around, by a human, or by a robot. This is really neat as all the images sequentially obtained in this manner can be combined in the computer to give an accurate 3D picture **1745** displayed on monitor **1750**.

In one preferred embodiment as shown in FIG. **17**C, the transducer head **1700** is comprised of a matrix **1755** of 72 individual transducer elements which send and receive ultrasound data at for example, 5 MHZ. This allows an expanded scan capability, since the sensor can be held steady at each discrete location xyz on the object surface, and a 3D image obtained with out movement of the transducer head, by analyzing the outputs of each of the transducers. Some earlier examples are described in articles such as: Richard E. Davidsen, 1996 IEEE Ultrasonics Symposium, A Multiplexed Two-Dimensional Array For Real Time Volumetric and B-Mode; Stephen W. Smith, 1995 IEEE Ultrasonics Symposium, Update On 2-D Array Transducers For Medical Ultrasound, 1995.

If the wand is now moved in space, fine scan resolution is obtained, due to the operation of the individual elements so positioned with out the need to move the wand in a fine pitch manner to all points needed for spatial resolution of this order. This eases the operators task, if manually performed, and makes robotization of such examination much easier from a control point of view.

Consider FIG. **17**B which illustrates a transducer as just described, also with automatic compensation at each point for pointing angle, robotically positioned by robot, **1785** with respect to object **1764**. In this case a projection technique such as described in U.S. Pat. No. 5,854,491 is used to optically determine the attitude of the object surface, and the surface normal direction **1760** from the position of target set **1765** projected on the surface by diode laser set **1770**, and observed by TV Camera **1775** located typically near the working end of the robot. Differences between the normal direction and the transducer propagation direction (typically parallel to the housing of the transducer) is then used by computer **1777** to correct the data of the ultrasonic sensor **1780** whose pointing direction in space is known through the joint angle encoders and associated control system **1782** of robot **1785** holding the sensor. Alternatively the pointing direction of this sensor can be monitored by an external camera such as **1710** of FIG. **17**A.

It should be noted that the data obtained by TV camera **1775** concerning the normal to the surface and the surface range from the robot/ultrasonic sensor, can be used advantageously by the control system **1782** to position the robot and sensor with respect to the surface, in order to provide a fully automatic inspection of object **1764**. Indeed the camera sensor operating in triangulation can be used to establish the coordinates of the exterior surface of object **1764** as taught for example in U.S. Pat. No. 5,854,491, while at the same time, the acoustic sensor can determine the range to interior points which can be differentiated by their return signal time or other means. In this manner, a complete 3D map of the total object, interior and exterior, can be obtained relative to the coordinate system of the Robot, which can then be transformed to any coordinate system desired.

The invention has a myriad of applications beyond those specifically described herein. The games possible with the invention in particular are limited only by the imagination.

What is claimed is:

1. A method for controlling a handheld computing device comprising the steps of:

holding said device in one hand;

moving at least one finger in space in order to signal a command to said device;

electro-optically sensing light reflected from said at least one finger using a sensing means associated with said device;

determining from said sensed light the movement of said finger, and

using said sensed finger movement information, controlling said device in accordance with said command.

2. A method according to claim **1**, wherein at least one camera is utilized to effect said electro-optical sensing.

3. A method according to claim **1**, including the further step of acquiring an image of at least a portion of the user of the device.

4. A method according to claim **1**, wherein said movement is sensed in 3 dimensions.

5. A method according to claim **1**, wherein movement of one finger relative to another finger is sensed.

6. A method according to claim **1**, wherein movement of two fingers is sensed.

7. Handheld computer apparatus comprising:

a housing;

a camera means associated with said housing for obtaining an image using reflected light of at least one object positioned by a user operating said object;

computer means within said housing for analyzing said image to determine information concerning a position or movement of said object; and

means for controlling a function of said apparatus using said information.

8. Apparatus according to claim **7**, wherein said object is a finger.

9. Apparatus according to claim **7**, further including a display function which is controlled.

10. Apparatus according to claim **9**, wherein said display is 3D display.

11. Apparatus according to claim **7**, further including means for transmitting information.

12. Apparatus according to claim **7**, further including a light source for illuminating said object.

13. Apparatus according to claim **7**, wherein said apparatus is a cellular phone.

14. A method for controlling a handheld computing device comprising the steps of:

providing a computer within said device;

associating a camera with said device, said camera viewing at least a portion of the body of a user operating said device or an object held by said user, in order provide image data concerning said portion or object;

using said computer, analyzing said image data to determine information concerning a user input command; and

from said determined information, controlling a function of said device.

15. A method according to claim **14**, wherein reflected light from said body portion or object is imaged by said camera.

16. A method according to claim **14**, wherein said information includes the position of the portion or object.

17. A method according to claim **14**, wherein said information includes the change in position of the portion or object.

18. A method according to claim **14**, wherein said information includes the velocity or path of the portion or object.

19. A method according to claim **14**, wherein said information is obtained in 3 dimensions.

20. A method according to claim **14**, wherein said information includes the pointing direction of the portion or object.

21. A method according to claim **14**, wherein a display is controlled.

22. A method according to claim **21**, wherein a virtual image on said display is moved or changed.

23. A method according to claim **21**, wherein said display is a 3D display.

24. A method according to claim **23**, wherein said display is a stereoscopic display.

25. A method according to claim **14**, including the further step of transmitting data to a further device.

26. A method according to claim **14**, wherein said camera operates at 30 frames per second or greater.

27. A method according to claim **14**, wherein said controlled function relates to a game.

28. A method according to claim **14**, including the further step of acquiring a picture of the user of the handheld device.

29. A method according to claim **14**, wherein two fingers of the user are sensed and a pinching action determined.

30. A method according to claim **14**, wherein said body portion indicates an expression of said user.

31. A method according to claim **14**, wherein said information provides an aid to speech recognition.

* * * * *

EXHIBIT C

US008878949B2

(12) **United States Patent**
Pryor

(10) **Patent No.: US 8,878,949 B2**
(45) **Date of Patent: *Nov. 4, 2014**

(54) **CAMERA BASED INTERACTION AND INSTRUCTION**

(71) Applicant: **Gesture Technology Partners, LLC**, Sylvania, OH (US)

(72) Inventor: **Timothy R. Pryor**, Sylvania, OH (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **13/961,452**

(22) Filed: **Aug. 7, 2013**

(65) **Prior Publication Data**

US 2014/0028855 A1 Jan. 30, 2014

Related U.S. Application Data

(63) Continuation of application No. 13/459,670, filed on Apr. 30, 2012, now Pat. No. 8,654,198, which is a continuation of application No. 12/891,480, filed on Sep. 27, 2010, now Pat. No. 8,189,053, which is a continuation of application No. 11/376,158, filed on Mar. 16, 2006, now Pat. No. 7,804,530, which is a continuation of application No. 09/568,552, filed on May 11, 2000, now Pat. No. 7,015,950.

(60) Provisional application No. 60/133,671, filed on May 11, 1999.

(51) **Int. Cl.**

H04N 5/232	(2006.01)
G06F 3/01	(2006.01)
G06F 3/038	(2013.01)
H04N 5/222	(2006.01)

(52) **U.S. Cl.**
CPC ***H04N 5/23296*** (2013.01); ***G06F 3/017*** (2013.01); ***G06F 3/0386*** (2013.01); ***H04N 5/222*** (2013.01); ***H04N 5/232*** (2013.01); ***H04N 5/23219*** (2013.01)
USPC **348/211.99**; 348/211.4

(58) **Field of Classification Search**
CPC .. H04N 5/23238; H04N 5/247; H04N 5/3415
USPC 348/211.4, 211.5, 211.8, 211.9, 222.1, 348/239
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,909,002	A	9/1975	Levy
4,219,847	A	8/1980	Pinkney et al.
4,339,798	A	7/1982	Hedges et al.
4,631,676	A	12/1986	Pugh
4,791,589	A	12/1988	Blazo et al.
4,843,568	A	6/1989	Krueger et al.

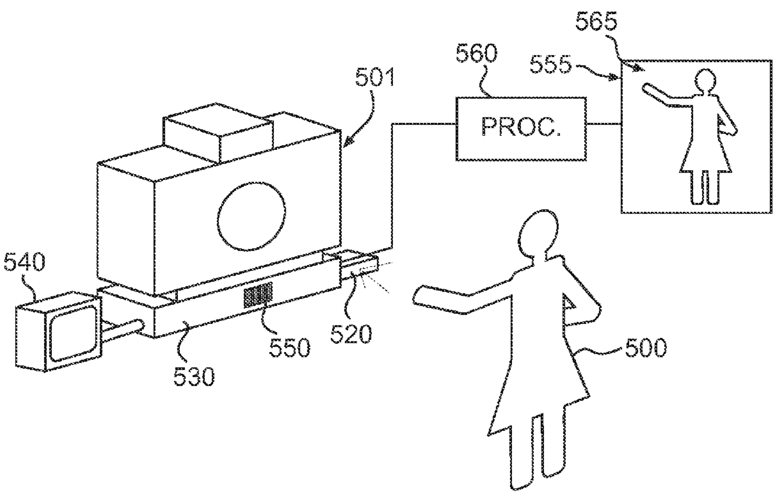
(Continued)

Primary Examiner — Tuan Ho
(74) *Attorney, Agent, or Firm* — Warner Norcross & Judd LLP

(57) **ABSTRACT**

Disclosed are methods and apparatus for instructing persons using computer based programs and/or remote instructors. One or more video cameras obtain images of the student or other participant. In addition images are analyzed by a computer to determine the locations or motions of one or more points on the student. This location data is fed to computer program which compares the motions to known desired movements, or alternatively provides such movement data to an instructor, typically located remotely, who can aid in analyzing student performance. The invention preferably is used with a substantially life-size display, such as a projection display can provide, in order to make the information displayed a realistic partner or instructor for the student. In addition, other applications are disclosed to sports training, dance, and remote dating.

18 Claims, 7 Drawing Sheets



(56) **References Cited**

U.S. PATENT DOCUMENTS

4,908,704	A	3/1990	Fujioka et al.
4,988,981	A	1/1991	Zimmerman et al.
5,008,946	A	4/1991	Ando
5,088,928	A	2/1992	Chan
5,227,986	A	7/1993	Yokota et al.
5,249,053	A	9/1993	Jain
5,297,061	A	3/1994	Dementhon et al.
5,365,597	A	11/1994	Holeva
5,376,796	A	12/1994	Chan et al.
5,388,059	A	2/1995	DeMenthon
5,454,043	A	9/1995	Freeman
5,491,507	A	2/1996	Umezawa et al.
5,534,921	A	7/1996	Sawanobori
5,572,251	A	11/1996	Ogawa
5,581,276	A	12/1996	Cipolla et al.
5,594,469	A	1/1997	Freeman et al.
5,616,078	A	4/1997	Oh
5,624,117	A	4/1997	Ohkubo et al.
5,781,647	A	7/1998	Fishbine et al.
5,781,650	A	7/1998	Lobo et al.
5,828,770	A	10/1998	Leis et al.
5,845,006	A	12/1998	Sumi et al.
5,853,327	A	12/1998	Gilboa
5,878,174	A	3/1999	Stewart et al.
5,904,484	A	5/1999	Burns
5,926,168	A	7/1999	Fan
5,940,126	A	8/1999	Kimura
5,982,352	A	11/1999	Pryor
5,999,840	A	12/1999	Grimson et al.
6,052,132	A	4/2000	Christian et al.
6,098,458	A	8/2000	French et al.
6,108,033	A	8/2000	Ito et al.
6,148,100	A	11/2000	Anderson et al.
6,160,899	A	12/2000	Lee et al.
6,204,852	B1	3/2001	Kumar et al.
6,252,598	B1	6/2001	Segen
6,342,917	B1	1/2002	Amenta
6,346,929	B1	2/2002	Fukushima et al.
6,359,647	B1	3/2002	Sengupta et al.
6,363,160	B1	3/2002	Bradski et al.
6,373,472	B1	4/2002	Palalau et al.
6,442,465	B2	8/2002	Breed et al.
6,508,709	B1	1/2003	Karmarkar
6,529,617	B1	3/2003	Prokoski
6,597,817	B1	7/2003	Silverbrook
6,663,491	B2	12/2003	Watabe et al.
6,750,848	B1	6/2004	Pryor
6,775,361	B1	8/2004	Arai et al.
6,788,336	B1	9/2004	Silverbrook
6,911,972	B2	6/2005	Brinjes
7,489,863	B2	2/2009	Lee
7,564,476	B1 *	7/2009	Coughlan et al. 348/14.08

* cited by examiner

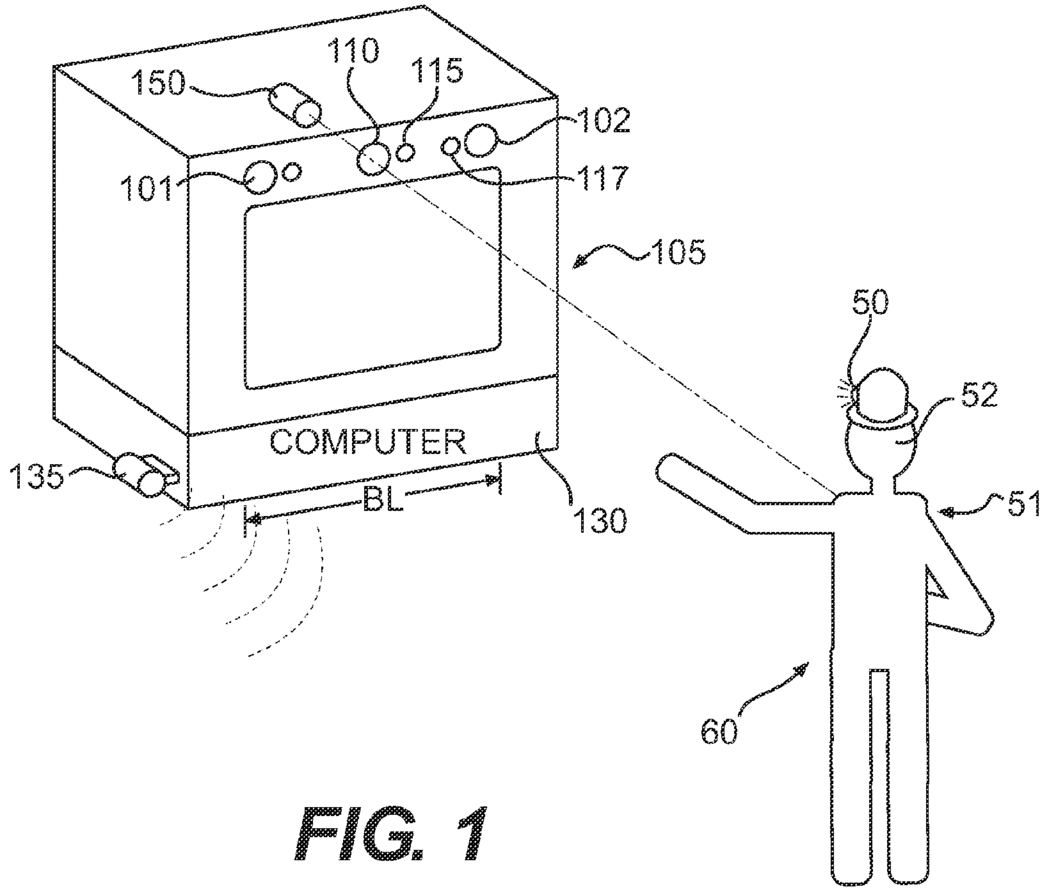


FIG. 1

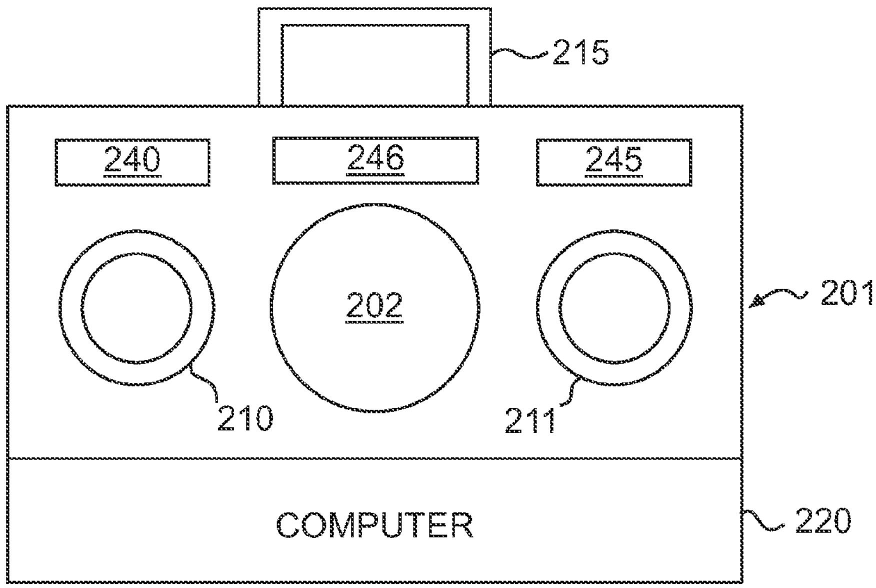


FIG. 2A

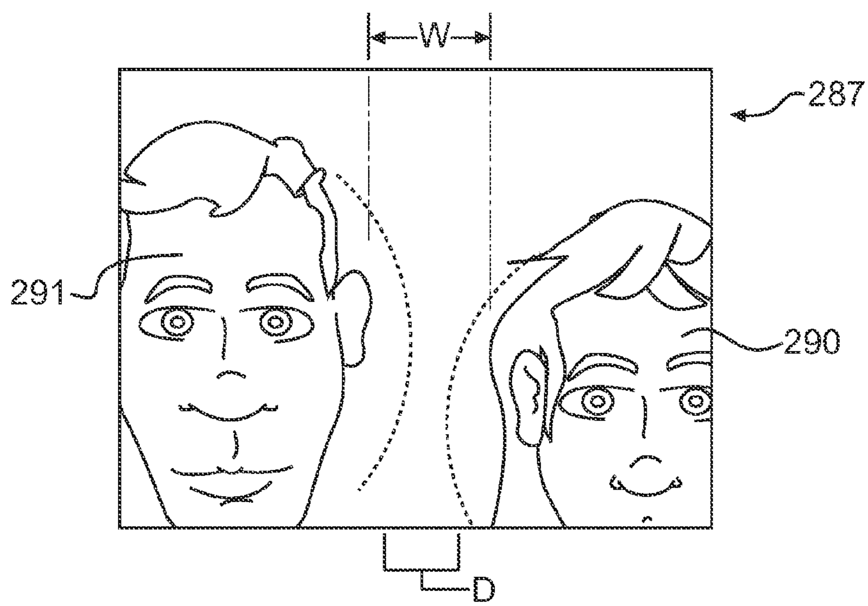


FIG. 2D

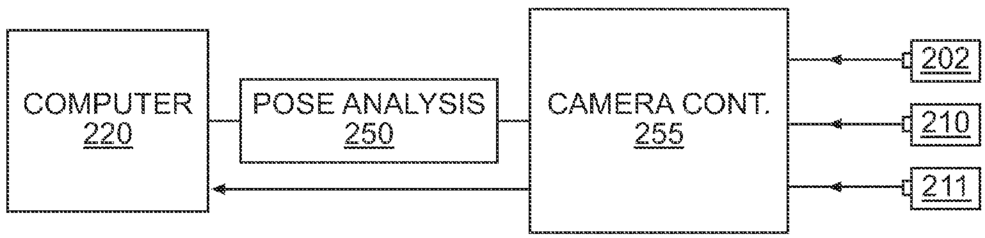

COMPUTER
220
POSE ANALYSIS
250
CAMERA CONT.
255
202
210
211

FIG. 2B

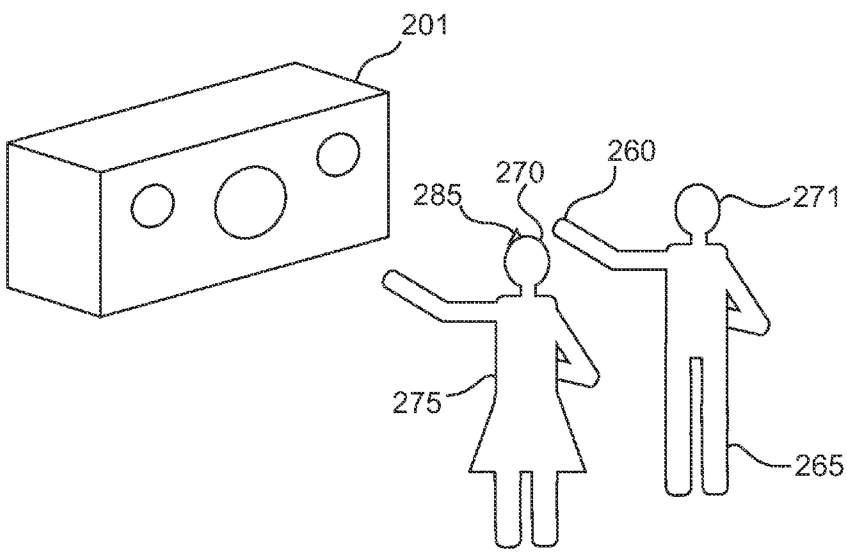

201
285
270
260
271
275
265

FIG. 2C

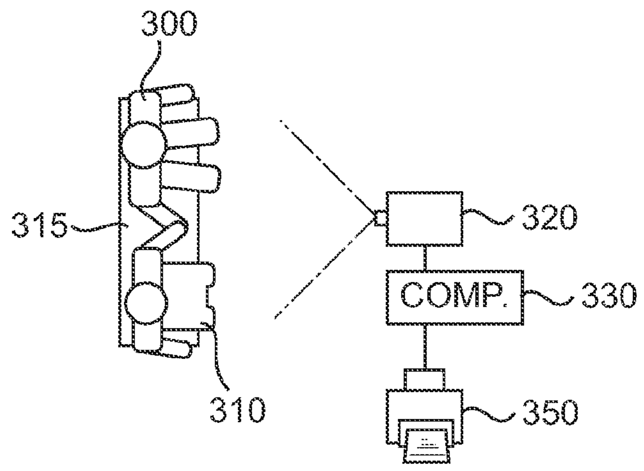

300
315
320
COMP.
330
310
350

FIG. 3

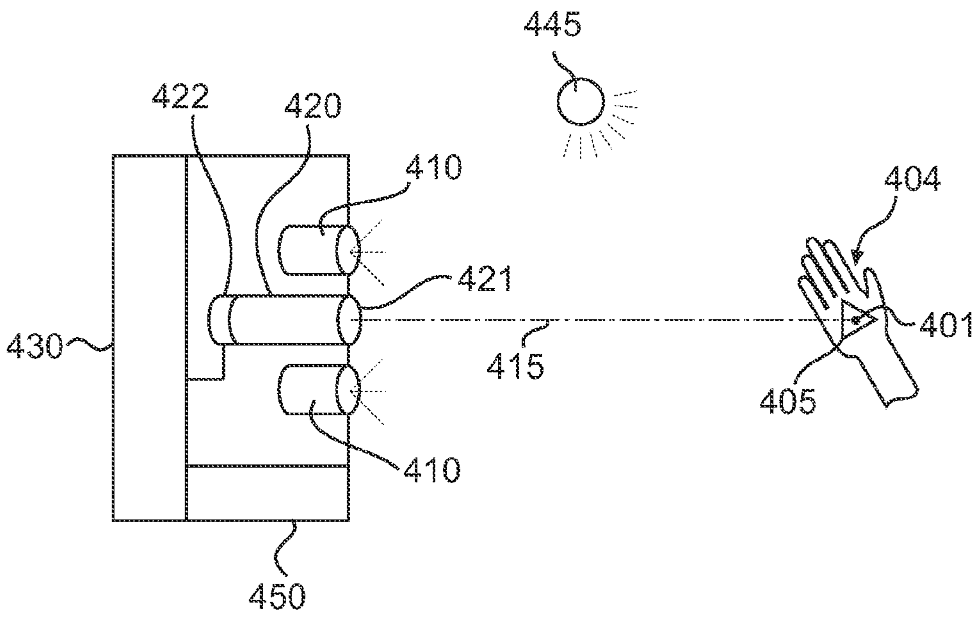

445
422
420
410
404
421
430
401
415
405
410
450

FIG. 4

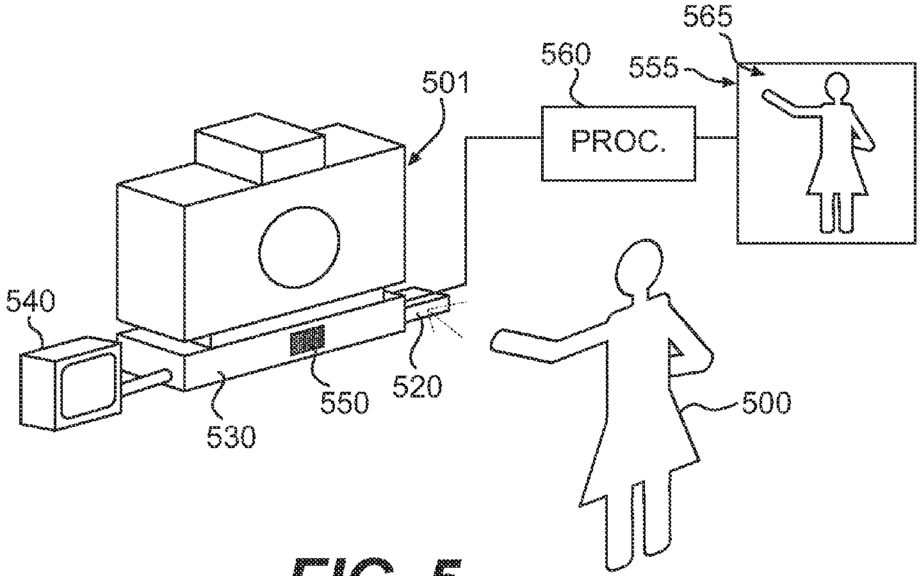

565
560
555
501
PROC.
540
520
550
500
530

FIG. 5

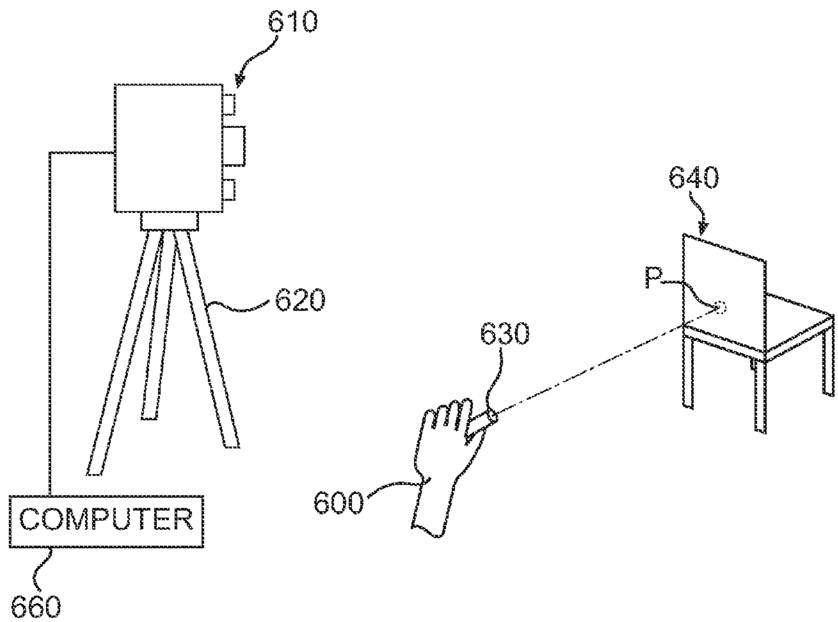

610
640
620
P
630
600
COMPUTER
660

FIG. 6

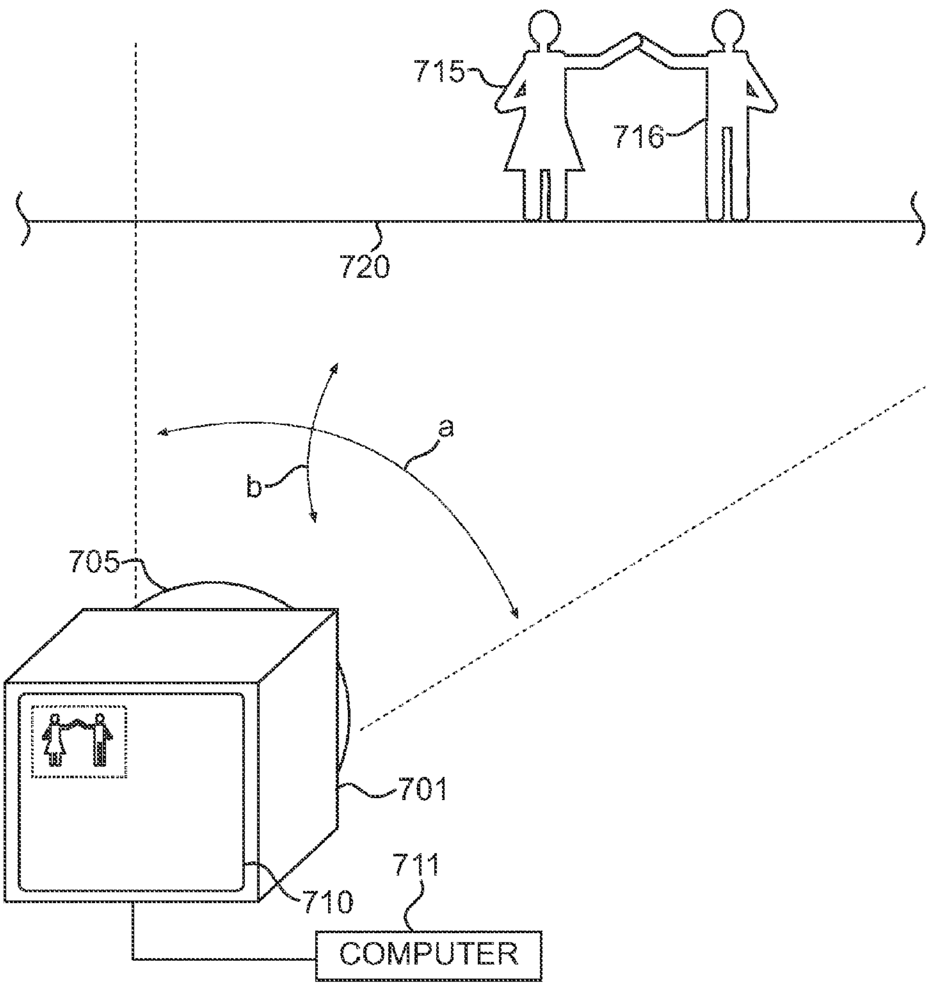


FIG. 7

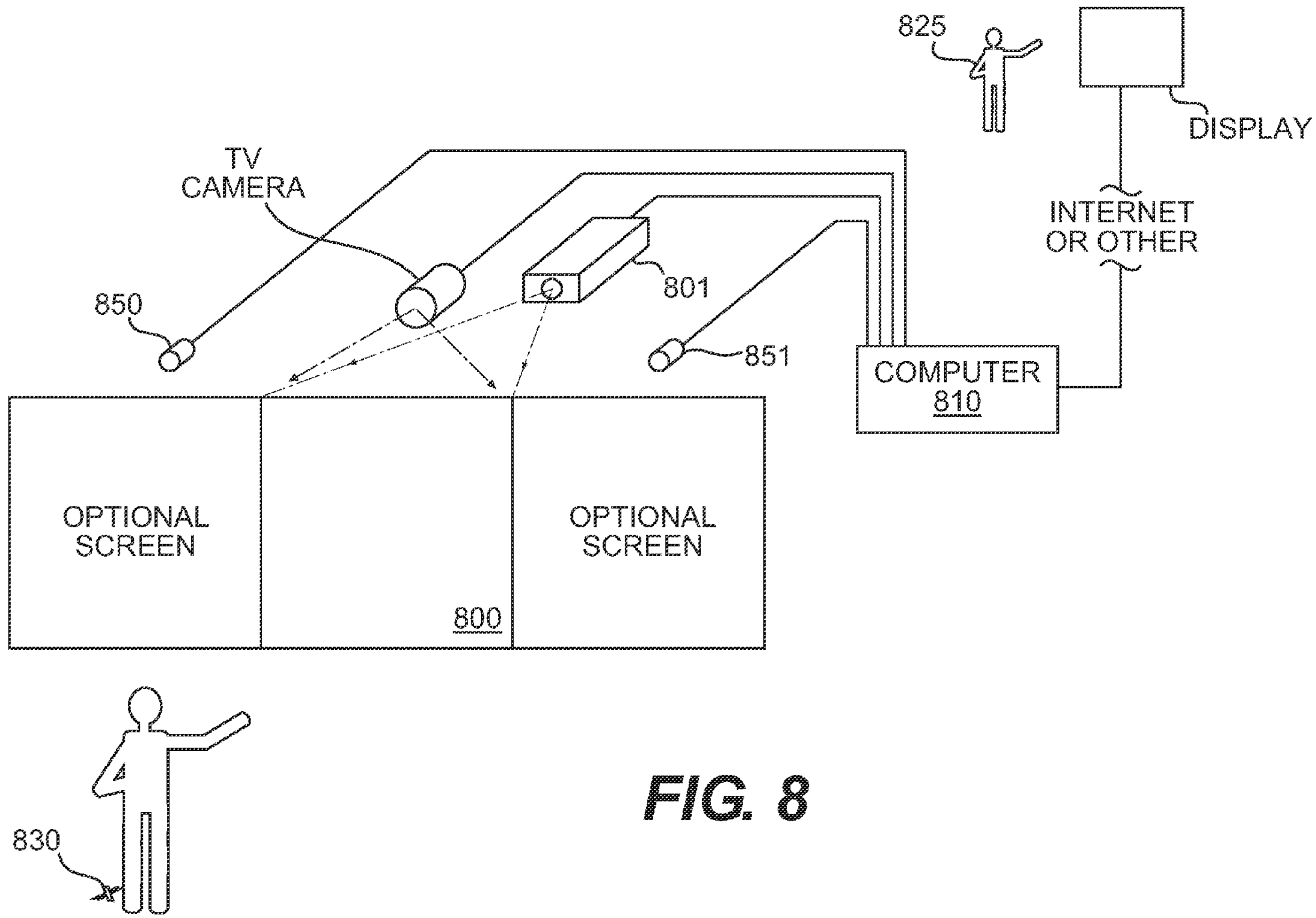

825
DISPLAY
TV
CAMERA
INTERNET
OR OTHER
801
850
851
COMPUTER
810
OPTIONAL
SCREEN
800
OPTIONAL
SCREEN
830

FIG. 8

CAMERA BASED INTERACTION AND INSTRUCTION

Method and apparatus are disclosed to enhance the quality and usefulness of picture taking for pleasure, commercial, or other business purposes. In a preferred embodiment, stereo photogrammetry is combined with digital image acquisition to acquire or store scenes and poses of interest, and/or to interact with the subject in order to provide data to or from a computer. Other preferred embodiments illustrate applications to control of display systems.

BACKGROUND

Representative of USA patents on Digital cameras are U.S. Pat. Nos. 5,534,921, 5,249,053 and many others which describe use of matrix array (CCD or otherwise) based cameras to take pictures of humans or other objects. The images taken are generally comprised of 400,000 or more pixels which are often compressed to smaller record sizes for data storage, for later retrieval and display. Video cameras or Camcorders are also increasingly able to take still photographs as well, and record or transmit them to computers.

Aside from exposure control (to keep the light reaching the detector array within the dynamic range of same), and range finding (to effect the best lens focus given the object distance in question) there are few cases known to the inventor where the camera taking the picture actually determines some variable in the picture and uses it for the process of obtaining the picture.

One such example that does not take a picture of humans but rather of data, is exemplified by U.S. Pat. No. 4,791,589, where a certain wave form signature on an oscilloscope is searched for by processing the digital camera image, and when it is seen, the image stored.

More apropos the function of "Picture Taking" as the general public knows it and of interest as the primary focus of the instant invention, is U.S. Pat. No. 5,781,650 by Lobo, et al which describes analysis after the fact of recorded images to determine facial content and thus the age of the subject. This disclosure also alludes to a potential point and shoot capability also based on the age classification of the individuals whose picture is desired.

There is no known picture taking reference based on object position and orientation with respect to the camera, or other objects that I am aware of.

SUMMARY OF THE INVENTION

High Resolution Digital still cameras employing matrix photodetector array chips to scan the image produced by the camera lens are now commonplace, and will be even more so in a few years as chips and memories become very inexpensive, and pixel density approaches 2000×2000 pixels, rivaling photographic film. Even today Camcorders having 700×500 pixel image chips are common for video based data and stills.

This invention is aimed at improvements in utilization of these cameras and others which make use of a computer based camera's ability to analyze, in real time if desired, the images obtained. Indeed a picture taking system may be composed of a combination of cameras, some used for purposes other than the recording of the picture proper.

It is a goal of the invention to provide a method for taking pictures when certain poses of objects, sequences of poses, motions of objects, or any other states or relationships of objects are represented. It is also a goal to allow this to be done in a self timer like mode, when desired scene situations or

specific dates or other circumstances exist. In some cases, information as to what is desired may be entered remotely, even over the internet, or radio telephone.

It is also a goal of the invention to provide a method for selecting from a digital or other picture memory, pictures obtained when certain pre programmed poses of objects, sequences of poses, or relationships of objects are represented.

It is a further goal of the invention to provide means by which users engaged in digital camera based activities, or other activities, using a computer can have their pictures taken.

It is a still further goal to provide all such functions in a 2D or 3D context, and using simple equipment capable of widespread use.

It is another goal of the invention to feed back data to a subject or subjects having his or her, or their picture taken, in order that they assume another pose or engage in another activity, or juxtaposition of subject positions.

While this invention is primarily aimed at the general picture taking public at large, it is realized that commercial photographers and cine-photographers, for example in the coming trend to digital "Hollywood" movie making, may benefit greatly from the invention herein, as it potentially allows more cost effective film production by giving the director the ability to expose the camera to the presence of masses of data, but only saving or taking that data which is useful, and if desired, to signal the creation of further data based on data obtained. All this with little or no human intervention as desired, thus saving on the cost of direction, film crews, and other labor or venue related costs.

DRAWINGS DEPICTING PREFERRED EMBODIMENTS OF THE INVENTION

FIG. **1** illustrates means by which users engaged in digital camera based activities, or other activities, using a computer can have their pictures taken.

FIGS. **2**A-**2**D illustrate a method for taking pictures when certain pre programmed poses of objects, sequences of poses, or relationships of objects are represented.

FIG. **3** illustrates a self timer like mode, or when specific dates or other circumstances exist, including a system embodiment for taking pictures in shopping malls or other locales and providing instant print or other hardcopy capability (e.g. on a tee shirt).

FIG. **4** illustrates means to provide all such functions in a 2D or 3D context, using simple equipment capable of widespread use. Various retroreflective artificialtarget configurations are also disclosed.

FIG. **5** illustrates a method to feed back data to a subject having his or her picture taken, in order that the subject assumes another pose or engage in another activity.

FIG. **6** illustrates a commercial version of the invention useful for police departments and real estate agents, among others.

FIG. **7** illustrates an embodiment of the invention used for photography of stage performances.

FIG. **8** illustrates an embodiment of the invention used for ballet instruction and other teaching and interaction activities also with remotely located instructors or players.

EMBODIMENTS OF THE INVENTION

FIG. 1

Illustrated in FIG. **1** of the invention is means by which users engaged in digital camera based activities, or other activities, using a computer can have their pictures taken, and in this context, FIG. **1** resembles that of co-pending referenced application 9 above. A single camera, or a set, such as a stereo pair are employed to see portions of an object, such as a person, a part of a person such as a hand, leg, foot, fingers, or head, and/or to view datums on an object, portion of an object, or an object held by the person or with which the person interacts. In addition, multiple persons and objects can be seen.

Where a single camera is employed, 2D measurements of object location relative to the camera (x and y perpendicular to the camera axis) are all that is possible, unless datums of known shape or spacing are used on the object viewed. Where a stereo pair or more of cameras are employed, 3D (xyz) data of a single point can be provided, for example retro-reflector **50** on the head **52** of person **51**. In both cases where 3 or more datums are used on an object, 6 Degree of freedom data can be obtained, allowing object orientation in 3 angular axes as well as range in 3 axes to be obtained. With two or more cameras, such 3D data may also be obtained using other features of objects such as edges of arms and the likely using known photogrammetric techniques.

The cameras used may also be used to take pictures of an object, or another specialized camera used for that purpose in conjunction with those used to determine the location of object features. Both examples are illustrated in this application.

As shown in this figure, two cameras **101** and **102** are used as a stereo pair, with each camera located at opposite sides of a TV monitor **105**, used for either computer or Television display or both. This is a desirable configuration commercially and discussed the co-pending application references above. In this particular case, an additional camera **110** is shown in the middle of the other two, said added camera used for picture taking, internet telephony and/or other purposes. An optional auxiliary LED light source **115** (or **116** or **117**) for illuminating a user **60** or other object is also shown.

All three cameras are connected to the computer **130** by means of a USB (Universal Serial Bus) daisy chain, or IEEE 1394 firewire connections (faster). Each is accessed, as needed for position and orientation determination, or picture taking.

Even using a single camera in two dimensions (as is normal today), some position and orientation data or sequences of same can be achieved using modern image processing techniques. (See for example the invention disclosed in U.S. Pat. No. 4,843,568 of Myron Krueger). However, accurate sensing and control of systems, such as cameras herein is difficult today with processors cost effective enough to be used by the public at large, and artificial target augmentation of image points is often desirable.

It is thus possible using the invention to be taking pictures of users of interactive computer systems for whatever purpose. This allows one to automatically capture images of children at play, for example with a computer system such as a computer game. It also enables many other functions which are described below. And it can be used in the field, where the computer, stereo position sensing and picture taking camera, may be co-located together in the same housing.

It is noted that where retro-reflectors are used, (as opposed to choosing for example less contrasting datums, for example natural object features such as edges of fingers, or clothing features, or targets such as colored dots) then each of the two cameras for stereo location determination needs lights to illuminate retro-reflectors substantially co-located with the camera axes. These lights can alternatively provide general lighting for any other camera or cameras to use in taking photographs or other purposes.

It is noted that cameras **101** and **102** need not have the image of the retro-reflector or other discernable target be in precise focus, indeed it is often helpful to have a some blur due to defocusing so as to aid sub pixel position solution of datum location. If the LEDs or other light sources are in the near infrared, and the camera lenses are focused in the visible, this occurs naturally, unless the lens is also near infrared chromatic corrected.

An optional laser pointer (or other suitable illumination source), comprised of diode laser and collimating optics **150** is also usable with the invention to illuminate object portions from which 3D data is desired (such as the neck region of person **51** as shown), or in the simpler case to designate which areas of a picture are to be focused, or zoomed in on or transmitted or recorded—with or without consideration of 3-D position data of the object. This can be fixed as shown, or optionally hand held by the user, for example in left hand (dotted lines) and used by him or her to designate the point to be measured in 3D location. (see also references above). In addition a person taking pictures, such as a photography can without looking through the viewfinder of the camera, point to appoint on the subject, which is then dealt with by camera typically by focusing the lens system such that the point is in the desired state of focus (usually but not necessarily when the laser spot on the subject appears smallest in diameter and/or of highest contrast). Such as system is particularly useful for cameras with wide fields of view, or those mounted on pan tilt mechanisms, where the mechanism can also be activated to position the camera axis to take the picture with the laser spot for example centered in the camera field.

In the laser designated case, it is generally the laser spot or other indication on the surface that is imaged, (although one can also instruct, for example using voice recognition software in computer **130** inputted via voice activated microphone **135**, the camera processor to obtain and store if desired the image of the area around the spot projected onto the object as well or alternatively), and if the spot is desired, it is often useful that cameras **101** and **102** have band-pass filters which pass the laser wavelength, and any led illumination wavelengths used for retro-reflector illumination for example, but block other wavelengths to the extent possible at low cost. It is noted that the discrimination in an image can also be made on color grounds—i.e. with red diode lasers and red LEDs, the system can analyze the image areas containing reds in the image, for example—with the knowledge that the answer can't lie at any shorter wavelengths (e.g. green, yellow, blue).

By using two cameras **101** and **102**, a superior ranging system for the laser spot location on the subject results, since the baseline distance "BL" separating the cameras for triangulation based ranging purposes can be sufficient to provide accurate measurement of distance to the object.

FIGS. **2**A-**2**D

As we begin to consider the apparatus of FIG. **1**, it is clear one could do much more to enhance picture taking ability than hereto fore described and contained in the prior art. And it can be done with apparatus capable of field use.

FIGS. **2**A-**2**D for example, illustrates a method for taking pictures when certain pre programmed or otherwise desired poses of objects, sequences of poses, or relationships of objects are represented. No such ability is available to photographers today.

Consider still camera system **201**, patterned after that of FIG. **1** and comprising 3 cameras and associated image scanning chips. The central camera, **202**, is for picture taking and has high resolution and color accuracy. The two cameras on either side, **210** and **211**, may be lower resolution (allowing lower cost, and higher frame rate, as they have less pixels to scan in a given frame time), with little or no accurate color capability, as they are used to simply see object positions or special datum positions on objects (which may be distinguished however by taught colors for example as taught in some of my co-pending inventions).

Cost wise the distinction between cameras is important. Today low cost CMOS chips and lenses capable of the providing stereo measurements as described above are $15 or less. High quality CCD color detector arrays and lenses for high quality photo images are over $100, and in many cases $1000 or more.

An optical viewfinder **215** is one of many ways to indicate to the user what scene information is being gathered by the camera system. The user can in this invention specify with a viewfinder based readout, the area of the field that is desired. Use of the viewfinder in this manner, whether looked through or displayed on a screen, is for example an alternative to designating an area on the actual object using a laser pointer for the purpose.

The camera system **201** further contains a computer **220** which processes the data from cameras **210** and **211** to get various position and/or orientation data concerning a person (or other object, or persons plural, etc). Integral light sources as described in FIG. **1** above may also be provided such as LED arrays **240** and **245** and xenon flash **246**.

In general, one can use the system to automatically "shoot" pictures for example, when any or all of the following occur, as determined by the position and orientation determining system of the camera of the invention:

1. Subject in a certain pose.
2. Subject in a sequence of poses.
3. Portion of Subject in a sequence of poses (e.g. gestures).
4. Subject or portion(s) in a specific location or orientation.
5. Subject in position relative to another object or person. For example, this could be bride and groom kissing in a wedding, boy with respect to cake on birthday, and sports events sequences of every description (where the camera can even track the object datums in the field and if desired adjust shutter speed based on relative velocity of camera to subject).
6. Ditto all of above with respect to both persons in certain poses or gesture situations.
7. When a subject undertakes a particular signal comprising a position or gesture—i.e. a silent command to take the picture (this could be programmed, for example, to correspond to raising one's right hand).

In addition it is noted that the invention acts as a rangefinder, finding range to the subject, and even to other subjects around the subject, or to all parts of interest on an extensive subject. This allows a desired lens focus to be set based on any or all of this data, as desired. It also allows a sequence of pictures to be taken of different objects or object portions, at different focal depths, or focus positions. The same holds true for exposure of these locations as well.

It is also possible to use the above criteria for other purposes, such as determining what to record (beyond the recording that is implicit in taking pictures), or in determining what to transmit. The latter is important vis a vis internet activity, where available internet communication bandwidth limits what can be transmitted (at least today). In this case video telephony with the invention comprehends obtaining only those images you really care about in real time. So instead of transmitting low resolution image data at 20 frames a second, you can transmit say 5 (albeit asynchronously gathered) frames of high resolution preferred data. (This doesn't solve flicker problems, but it does mean that poor quality or extraneous material isn't sent!). Criteria such as degree of image motion blur or image focus can also be used in making transmission decisions.

FIG. **2**B illustrates a block diagram showing a pose analysis software or hardware module **250** analyzing processed image data (for example utilizing camera image data processed by visionbloks software from Integral Vision Corp.) from the computer **220** (which may be the same physical microprocessor, such as a Intel Pentium 2 in a Dell inspiron 3500 laptop computer, or different) and determining from same when a certain pose for example has been seen. When this occurs, a signal is sent to the camera control module **255** to hold the last frame taken by camera **202**, and to display it to the photographer, digitally store it, or transmit it—to someone else, or another data store or display. Such transmission can be by data link, internet, cell phone, or any other suitable means.

Another criteria could be that two or more preselected poses were seen one after the other, with a time delay between them, also pre-selected if desired.

FIG. **2**C illustrates a specific case whereby a point on one person, say hand **260** of man **265** having head **271**, is determined, and a picture is taken by camera system **201** of the invention when this point comes within a distance of approximately 6 inches (or any other desired amount including contact—i.e. zero distance) from another person or object, say the head **270** of woman **275**. To obtain the data, one can look for hand or head indications in the image using known machine vision techniques, and/or in a more simple case put a target marker such as colored triangle **285** or other type on the hand or head or both and look for it.

The use of the natural features of the subjects heads, which are distinguishable by shape and size in a known field containing two persons, is now illustrated. For example, image morphology or template matching in the image field of the solid state TV camera **202** can be used to distinguish the head shapes from background data and data concerning the rest of the features such as hands, etc. of subjects **265** and **275** (or conversely hand shapes if desired can be found and heads excluded, or the hand of the right person, versus the head of the left, and so forth).

As shown in FIG. **2**D, when the image field **287** of camera **202** after processing contains the two head images, **290** and **291**, spaced a distance "W". When W is not within a tolerance D, the picture is not taken; whereas if the heads are close enough, within D as illustrated in dotted lines, the picture is taken.

Criteria as mentioned can include proximity of other parts of the body, or objects associated with the subjects (which themselves can be objects). In addition, the motion or relative motion of objects can be the criteria. For example, one could take program the device to take the picture when on two successive frames the condition shown in FIG. **2**D exists where the heads are apart in frame **1**, but closer in frame **2** (probably corresponding to a movement say of the boy to kiss the girl). Clearly other sequences are possible as well, such as movement taking place in several frames followed by a sequence of frames in which no movement occurs. Other means to determine motion in front of the camera can also be used in this context, such as ultrasonic sensors.

It is also noted that the actual position or movement desired can be "Taught" to the computer **220** of the picture taking system. For example, a boy and girl in a wedding could approach each other and kiss beforehand. The sequence of frames of this activity (a "gesture" of sorts by both parties) is recorded, and the speed of approach, the head positions and any other pertinent data determined. When the photographer thinks the picture is right, the computer of the camera system is instructed to take the picture—for example it could be at the instant when after a suitable approach, two head images become joined into one—easily recognizable with machine vision processing software under uniform background conditions. Then in the future, when such a condition is reached in the camera field of view, pictures are taken and stored, or transmitted. This allows a camera to free run whose image field for example takes in the head table at a wedding party, taking only the shots thought to be of most interest. Numerous conditions might be programmed in, or taught in—another at the same party, would be anyone at the head table proposing a toast to the bride and groom, with arm and glass raised. If video is taken, it might be taken from the point at which the arm rises, until after it comes down. Or with suitable voice recognition, when certain toast type words are heard, for example.

Application to "3-D" Pictures

Where it is desired to take "3-D" pictures, it can be appreciated that each camera, **210** and **211** can take images of the scene in place of camera **202**, and that both cameras **210** and **211** outputs can be stored for later presentation in a 3D viewing context, using known display techniques with appropriate polarized glasses or switchable LCD goggles for example. In this case the camera outputs can serve double duty if desired, each both recording picture data, as well as determining position of one or more points on the object or objects desired.

In addition, or alternatively, one can use in this 3D picture case, the camera **202** (or even a stereo camera pair in place of **202**) as a means for determining position and orientation independently from the stereo picture taking cameras.

If not used for immediate position information, camera **202** does not have to be digital and could employ film or other media to record information.

FIG. 3

In a manner resembling that of FIGS. **2**A-**2**D above, the invention can also serve to aid a person to take his or her own picture—a modern "Self timer" if you will. For example any or all of the criteria such as the items 1-7 above, can be used as criteria for the picture to be taken of oneself. This is in addition to other more normal things like taking pictures after a certain time, or on a certain date or time interval, etc. This has particular appeal for taking pictures of one's self, or in any other situation where the photographer is not present (e.g. unattended recording of animals, children, etc.). Similarly, a hand signal or other signal to the camera can be used to trigger the picture to be taken, using the computer camera combination to determine the hand position or movement. This can also be done by voice using microphone input and suitable voice recognition software in the computer.

Today, in a conventional context, one can as a photographer, choose to shoot a fashion model or other subject, and when you see a pose you like record the picture. But as one's own photographer, this is much more difficult, unless you stream in video and search through the poses after the fact. But even then, you don't know that the poses were what was desired, as no feedback exists during the shoot.

With the invention, you may program the system to take only those poses which you think you want to get. And it can instruct the subject, when a picture is taken (and the lack thereof indicating to do something different to obtain the desired effect resulting in a picture). The effect desired can be changed in midstream to adjust for changing wants as well, by changing the program of the computer (which could be done using hardware switches, inserting a disc, or otherwise entered as a command). In addition, as mentioned above, the gesture or pose desired, can be taught to the system, by first photographing a variety of acceptable positions or sequences, and putting bounds on how close to these will be accepted for photographing.

A specialized case is shown in FIG. **3**, for self taking instant picture or printout device for use in a shopping mall Kiosk or other venue. In this case two sweethearts **300** and **310** are on a bench **315** in front of the digital or other camera **320**. When the computer **330** detects from processing the image (or images) of the invention that their faces are in close proximity (for example using the centroid of mass of their head as the position indicator, or even facial features such as described in the Lobo et al patent reference), the computer then instructs the camera to record the picture. A push button or other selector on the device allows the subjects to select what criteria they want—for example when their heads are together for 5 seconds or more, or not together, or hands held, or whatever. Or when their faces are within a certain distance criteria, such as one inch.

Alternatively, camera **320** may be a video camera and recorder which streams in hundreds or even thousands of frames of image data, and the selection of a group is made automatically by the invention in rapid fashion afterwards, with the subjects selecting their prints from the pre-selected (or taught as above) images as desired. Or the machine itself can make the final selection from the group, sort of as a random slot machine for pictures so to speak, and print the picture using inkjet printer **350** for example. Such a situation could be provided at less cost for example, with an incentive to add in your own criteria for an extra cost, and get pictures to choose from more along the lines desired. Note that in addition to, or instead of prints, they could have magnetic or other machine readable media to take home too.

FIG. 4

FIG. **4** illustrates means to provide all such functions in a 2D or 3D context, using simple equipment capable of widespread use.

For example, the simplest case is to use the same single camera such as **110**, to both take the picture, and to determine location, according to the invention, of one or more points on the object or objects for purposes of controlling the picture taking, recording, or transmission process in some way.

As has been disclosed in the aforementioned referenced co-pending applications, one can view using the single camera, one or more such points in two dimensions, or in three dimensions under certain conditions when spaced points on the object have known spacing between them on the surface of the object.

Identifying points from raw images is processing intensive, as is determination movement gestures of such images, such as an image of an arm or hand in a varying clothing and background situations. But determining the location or movement of one or more artificial targets such as a colored retro-reflector is easy, accurate and fast, based on brightness (under substantially coaxial illumination) and color—and possibly shape as well if the target is of some distinguishable shape.

For example, consider retro-reflector (e.g. glass bead Scotchlight 7615 tape by 3M company) **401**, on the hand of a subject **404**, the retro-reflector having a red reflection filter **405** matched to the wavelength of the LEDs **410** used with (and angularly positioned on or near the axis **415** of) camera **420** comprising lens **421** and detector array **422** used to take the picture of the object desired. When it is desired to determine the position of the hand **404**, the red LED's are turned on by camera controller **430**, and a bright reflection is seen in the image at the point in question due to the retro-reflection effect.

Where stereo pairs of cameras are used, as in FIG. **1** or **2**A, two reflections are seen whose disparity in location from one camera to the other gives the z distance (range direction) from the camera. In this case light sources are located with each camera of the stereo pair in order that for each camera, the retro-reflectors are properly illuminated with light emanating from point or points angularly near the camera in question.

The LEDs can be illuminated on alternate camera frames, or at any other time when "picture" type image data is not desired. In this case the camera does not under room lights **445** say, normally see the retro-reflection signal, which is desirable as the bright spot of **401** from the image of the human desired. Processor **450** processing the data, can even be used to subtract out from the recorded image, the shape of the retro-reflector, which might be a noticeably different shape than found in practice (e.g. a triangle). The image can be filled in where the subtraction occurred with color, brightness, contrast and texture or other characteristics of the surroundings. This is particularly easy if the target (retro-reflector or otherwise) is placed on the human or object in a region of small variation in characteristics needed to be filled in, e.g. the back of one's hand, say. The key is that after processing, the image look like it did without addition of the artificial target.

If the LEDs are turned on by the camera controller during picture taking, color processing can be used to remove from the stored image of the scene, any indications of bright zones at the LED wavelength used, filling in with color of the surrounding area as desired.

Clearly both processing techniques just described or others can be used. And the methods work well with stereo pairs of cameras too.

Retro-reflective or other distinguishable artificial targets can be provided in different decorative designs for wrist, back of hand, rings, forehead, hats, etc. For example, 3 targets in a heart or triangle shape, a square box of 4 targets, or a box or pyramid with line targets on its edges, and so forth.

Colored targets can be made of cloth, plastic, or the like, including Colored plaids, polka dots, etc. Or coatings or Filters or evaporated on filters may be placed in front of a target such as a plastic retroreflector in order to render it of a given color (if it wasn't made of colored material in the first place).

Decorative line outlines (also possible in retroreflective bead material) can also be used as target datums, for example down the seam of glove fingers, or shoes, or belts, dress beading, etc.

FIG. **5**

FIG. **5** illustrates further one of many methods by which the invention may be used to feed back data to a subject (or subjects) having his or her picture taken, in order that the subject assume another pose or engage in another activity.

For example consider FIG. **5**. A girl **500** is having her picture taken by the camera of the invention **501** (in this case a single digital camera version such as illustrated in FIG. **4**), and her positions, orientations or sequences of same, including motions between points are analyzed as described above, in this case by computer **530**. The computer has been programmed to look for funny movements and positions, defined here as when the arms are in unusual positions (clearly a subjective issue, programmed as to tolerances, or taught to the system by the person in control of the situation).

The girl then poses for the camera. When the camera of the invention takes the picture according to its preprogrammed criteria (in this case, for example, defined as when her arms are over her head, and after a significant movement has occurred), it lets her know by lighting light **520** connected by wires not shown to computer **530**. During the photo shoot, then she begins to learn what it is looking for (if she hasn't been already told) and does more of the same. If desired, and optional video display **540** or voice out put speaker **550**, both connected to computer **530**, indicate to her what is desired. This could also be a particular type of pose, e.g. "Cheesecake" based on historic classical poses learned from photo art (note that she can also make comments for recording too, with optional microphone input not shown. As pointed out above, voice recognition software, such as IBM Via Voice" can be used to recognize commands from the subject or photographer, and cause other results).

It can be more sophisticated yet. For example, if the computer **530** and any associated software as needed may be used to analyze the model's lips and her smile. In this manner, the invention can be used to photograph all "smiling" poses for example. Or poses where the smile is within certain boundaries of lip curvature even. Similarly, the camera or cameras of the invention can be used, with suitable image analysis software to determine when the subject's eyes are open a certain amount, or facing the camera for example.

FIG. **3** above has alluded to possible use of the invention data processing to determine position and/or orientation data from recorded picture frames, after the picture is taken. A method for selecting from memory pictures obtained when certain pre programmed poses of objects sequences of poses, or relationships of objects are represented.

Selection can be according to criteria for example 1-7 above, but there are some differences. First if the data is taken normally from a single camera such as that of **202** above, 3D information is not available. This being the case, conventional 2D machine vision type image processing (e.g. "Vision Bloks" software from Integral Vision Corp.) can be used to extract object features and their locations in the images retained.

A second version alternatively could employ a single picture taking camera, but by employing 3 dot or other suitable targets on the photographed object in the camera field, could calculate 3D data related to the object (position and orientation in up to 6 axes can be so calculated by the computer of the invention using target location data in the camera image field) .

A third version, records data from the camera, or in the case of the FIG. **2**A device, all three cameras—all recorded for example on digital media such that the processing can be done after the fact, just as it would have been live.

Another application can be to monitor the relative change in successive pictures as seen by one or more relatively low resolution cameras and when such change is minimal, cue the high resolution camera requiring a longer exposure to become enabled. In this manner blur of the high resolution camera image is avoided. This is useful in taking pictures of children, for example. This comparison of images can be made without actually measuring distances, but rather by looking for images which are not different within an acceptance band, one to another, thus indicating the motion is largely stopped. This can be determined by subtracting one image from the other and determining the amount of pixels above a threshold. The more, the less the images are alike. Other techniques can be used as well, such as correlation techniques.

In some instances it is desirable to have, in taking pictures, a display such as **555**, preferably (but not necessarily) life size. This display can be not only used to display the image **565** of the person whose picture is being taken, but as well can display still (or video) images called up from computer memory or other media storage such as DVD discs, and the like. One use of the displayed images is to indicate to the subject a desired pose for example. This can be done by itself, or interactively using the invention. A computer generated and rendered 3D image can also be created using suitable 3D solid modeling software (such as CAD KEY) to show an approximate pose to the model.

For example the invention disclosed above, allows one to automatically observe the expressions, gestures and continence of a person, by determining the shape of their smile, the direction of eye gaze, and the positions or motion of parts of the body such as the head, arms, hands, etc. Analysis using pre programmed algorithms or taught sequences can then lead to a determination as to what information to display on display **555** controlled in image content by display processor **560**.

As one instance, suppose computer image analysis of data from camera **501** of the invention has determined that the person **500** is not smiling enough, and is in too stationary a pose. A signal from computer **510** is provided to display processor **560** so as to display on display **555** an image of someone (perhaps the same subject at an earlier time, or a computer generated likeness of a subject) having the characteristics desired. The person looks at this display, and sees someone smiling more for example, and in one scenario, tries to mimic the smile. And so forth. Alternatively, voice generation software, such as included in IBM VIAVOICE can be used to computer generate a voice command, "Smile More" for example, rather than show a visual illustration of the effect desired.

FIG. **6**

Let us now discuss some other applications of picture taking enabled by the invention. One embodiment can be used to determine location of items in a scene, for example furniture in a house, for which homicide studies or insurance fraud could be an issue (see also FIG. **1** above, as well as referenced co-pending applications).

For example, a detective (whose arm **600** is shown) arrives at a murder scene in a room, and he sets the stereo camera **610** of the invention disclosed in FIG. **2**C on a tripod **620** (or other suitable location) and systematically designates, using laser pointer **630**, any object desired, such as chair **640** impacted by the laser beam at point P. The camera/computer system of the invention locates the designated point takes a picture of the room, or a portion thereof, including the zone of the designated point P which stands out in the picture due to the laser spot brightness. Optionally, the stereo pair of cameras of the invention can digitize rapidly the xyz coordinates of point p, which can be superposed if desired on the image of the scene including point p itself and its immediate surroundings. This data can be processed by computer **660** as desired and either recorded or transmitted to a remote location along with the images as desired using known communication means. This work can be done outdoors, as well as inside. Numerous points to be digitized can be sensed and/or indicated, as desired.

The same digitization procedure can be used to digitize a room for a real estate person for example, to develop a data base on a house for sale. And many other such applications exist.

Finally it should be noted that the invention solves many famous problems of picture taking, for example of children. The digital camera images of the invention can be processed for example using appropriate software such as Vision Bloks to determine if the child's eyes are open (determined for example by recognizing the eye iris in the face area), and if so to take the picture, or after the fact, to select the picture from a group. Or a signal can be given by the system to the child to "open your eyes" so to speak. To determine if the eye is open, the image can be processed for example to look for the white of the eye, or to look for red reflections from the eye. This can even be done with deep red, or near IR light sources like LEDs which do not bother the child.

Similarly, if the child (or other subject) is in motion, when you want him still, the picture can be analyzed until he is still, and then the picture taken or selected. This can be determined from comparison of successive frames, from motion blur or other characteristics of motion in the image. Or a signal as above can be given to the child to "sit still" (a famous command in picture taking annals).

FIG. **7**

The invention can also be used for commercial photography and for producing motion pictures. One advantage is that very high resolution images at suitable exposure levels of critical scenes can be taken, but not too many which would overload the memory capacity of a camera system. A means to enhance this is now described.

It is noted that a camera having an ability to read individual pixels as desired, or at least to choose the lines of pixels to be read, can achieve high rates of scan if one knows apriori where to look apriori for data. Or if one say scans every 20th pixel in either direction xy of the camera, to determine where frame to frame changes are occurring (due to change in pixel brightness or color). Once change is determined one can often isolate those areas to the ones of interest. For example, even in a "Still" picture, the head often moves (similar to the lovers on the bench in the shopping mall mentioned above). Every 20th pixel, cuts the number of pixels by 400 times, and raises a normal 30 hz scan rate to over 1000 scans per second—more than needed in many cases.

When the area of interest is found, the pixels in that area are all scanned for example.

Such pixel addressing cameras can also be used for determining the position and change in position of features used to determine, and track, pose and other variables, as has also been discussed in co-pending applications, particularly Camera Based Man-Machine Interfaces U.S. Ser. No. 60/142,777, incorporated herein by reference. Of special interest is that same high resolution camera can be used to take the picture desired, while at the same time be used to find or track the object at high speed.

Such high speed tracking can be interspersed with the taking of pictures. For example if in photographing a ballet, it may be desired only to take pictures of the prima ballerina, who typically is the one, with any male dancer, that is moving the most. By determining the zone to be measured, one can sense quickly what zone should looked at, and high resolution photographs obtained from that zone. This allows one to use a very large format camera in a fixed location (e.g. 5000×5000 pixels) to cover the image of the whole stage via suitable optics, but to only take and store the pixels in a 1000×700 zone of interest movement, or positional or gesture interest for example, providing a 35 times increase in the frame rate needed today with such large pixel cameras. This allows their practical use, without resort to human cameramen, or pan/tilt mechanisms.

Similar logic holds for quarterbacks in a football game, who often run faster than any defense men around them and can be differentiated accordingly (along with any other issues such as uniform color, design or the like). If possible, it is desirable to have a clearly defined target, such as a retroreflective or bright colored target on one's helmet for example. Indeed helmet color can be chosen accordingly.

This is illustrated in FIG. **7** wherein camera **701** composed of lens **705** and an addressable version of a Kodak MegaPixel detector array **710** having 4000×4000 elements and under the control of computer **711** is used to scan the image of a pair of dancers **715** and **716** on stage **720**. The field of view of the camera equal to area ab covers the whole stage. But the area scanned out from array **710** is confined to the region in which the dancers were last seen, which is defined as a zone a'b' equal to in this case 500×500 pixels. This still allows DVD type resolutions to be achieved, without pan or tilt of the camera. Similarly such techniques can be used for video conferencing, sports, and other activities as well.

It should be noted that in the above embodiments the words picture and photograph are interchangeable, as are photographing or photography and picture-taking. The camera used for same is preferably but not necessarily a solid state TV camera whose pixels are scanned serially or randomly under program command.

FIG. **8**

The invention can also be used to sense positions of people for instructional purposes. Data as to a dancer's movements for example can be obtained, and appropriate images, or data or both transmitted without excessive bandwidth requirements to a remote location for comment or interaction by a trained professional. Combined with life-size screen displays this allows a life like training experience to be gained at low cost, since one professional can watch 10 students in different locations say, each trying her movements alone in the intervening moments. In addition such training can occur in the home, as if one had a private tutor or coach.

For example consider FIG. **8**. A class of ballet students is practicing near a "mirror" which in this case is comprised life size digital display screen **800** illuminated from the rear by a Sharp brand projector **801** driven by computer **810**. By sliding a real mirror in an out the mirror can be a mirror, or a display. If desired, this display can be extensive, and for example using 3 projectors to cover 3 adjacent screens each 6 feet high×9 feet long for example, such that a total length of a large studio is comprised.

A master instructor **825** (possibly remotely located via the internet or other communication means) can observe the students via TV camera (or cameras). By viewing the students the instructor can make corrections via audio, or by calling up imagery which represents the appropriate moves—for example from a professional doing the same Swan Lake number. In addition, the TV cameras of the invention can monitor the actual location and movements of the student, or students, and their relationship to each other, and if desired to various markers such as **830** on the floor of the studio, placed there to assist in choreographing the piece.

In addition, if the various gesture and position monitoring aspects of the invention are utilized as described above and in co-pending applications it is possible to have the instructions computer generated using dancers movements as input to a computer analysis program. This is particularly useful if dance routines which are classical in nature, are being attempted, which have known best forms which can be computer modeled.

In another version, an assistant can be on the scene say working with ten students in a local studio, while the master is remote.

It is also possible with the invention to provide input image data to projector computer **810**, even from remote internet located sources, which represents other people dancing for example. These can be images of the master, or others in the class—even if all in different locations. OR the images can be those of others who have performed a particular routine in the past, for example Dance of the Sugar plum fairy in the Nutcracker. This imagery could be from the Bolshoi ballet performance of the same dance, displayed in small town ballet studio or home—to illustrate the moves required. The use of life size projection not only gives a feel to this imagery, but further allows, I have discovered, a unique experience for the performer. Namely that the person can perform "with" the troupe displayed. In some cases, in ballet for example, this sometimes can be more useful than watching one's self in the mirror (typical in ballet studios).

By using the cameras of the invention, such as stereo pair **850** and **851** to determine student positions, it is also possible to control the display in many ways. For example as the student got closer to the display, the persons in the display could appear to come closer to the student. Conversely, it might be desirable to have them move away from the student to keep a constant apparent distance between them for example. And if the student is twirling left, the figures in the ballet depicted on the screen can be caused to turn right (as they are "in the mirror" so to speak) to match the movement of the student in approximate form at least.

In addition it is often desirable for learning purposes to Control speed of music and video display to match sensed movements of pupil, or from remote master person. Use display techniques which can produce variable motion display, such as variable speed DVD disc or read data in to ram. In addition it is desirable that overlaid could be masters voice.

The invention can be advantageously used in many performing arts, not just ballet. For example, live theatre, where actors from Hamlet performances of the past can interact with those practicing. Or where instructors of Skating or Gymnastics, other activities can also interact.

Sports as well is amenable to the technique, but the size of the "studio" or gym becomes an issue. Basketball for example fits the space aspect of the projection screens and the fields of view of the invention cameras as here described.

Ability of masters remotely located, and use of copyrighted performance material of famous performers and troupes allows one to franchise the studio concept of the invention. For example each town could have a Bolshoi studio franchise of this type.

It is noted that this same arrangement can serve other purposes beyond instruction. One is the possibility of remote dating, in which sensed movement of one partner is communicated, along with voice and visual expression to the other. In addition, is possible, as disclosed in co-pending applications, to build the displays described above in the form of a touch screen in which contact of one partner with the display of the other remotely transmitted from afar can occur.

If one uses large scale touch screens with optional added sensor inputs. As would be the ballet studio example of FIG. **8** if equipped with touch screen capability, then one can provide a mechanism for marketing of people relative (i.e. life size) objects such as automobiles in facilities such as Auto showrooms. Thus a ballet studio for example, can be used for other purposes, not just instructional, but for selling cars for example, where the display screen is displaying new models (including ones that are figments of design imagination, and where customer input is desired as in a focus group) and where customer inputs voice and action can be detected if desired by the invention. Or in reverse, an underused car showroom can be converted—on demand—into a site which can be used for, among other things, instructional purposes in performing arts, sports and the like. This gives a reason for being to the show room that transcends selling cars, and helps attract people to the facility. If a car was displayed, on a touch screen, one could walk up to the full size display of the car, and touch the door handle, which would cause the touch screen to sense that same had occurred, and indicate to the computer to cause the display to display the door opening to expose the interior.

The invention claimed is:

1. A portable device comprising:

a device housing including a forward facing portion, the forward facing portion of the device housing encompassing an electro-optical sensor having a field of view and including a digital camera separate from the electro-optical sensor; and

a processing unit within the device housing and operatively coupled to an output of the electro-optical sensor, wherein the processing unit is adapted to:

determine a gesture has been performed in the electro-optical sensor field of view based on the electro-optical sensor output, and

control the digital camera in response to the gesture performed in the electro-optical sensor field of view, wherein the gesture corresponds to an image capture command, and wherein the image capture command causes the digital camera to store an image to memory.

2. The portable device of claim **1** wherein the determined gesture includes a hand motion.

3. The portable device of claim **1** wherein the determined gesture includes a pose.

4. The portable device of claim **1** wherein the electro-optical sensor is fixed in relation to the digital camera.

5. The portable device of claim **1** further including a forward facing light source.

6. The portable device of claim **1** wherein the electro-optical sensor defines a resolution less than a resolution defined by the digital camera.

7. The portable device of claim **1** wherein the electro-optical sensor includes at least one of a CCD detector and a CMOS detector.

8. A computer implemented method comprising:

providing a portable device including a forward facing portion encompassing a digital camera and an electro-optical sensor, the electro-optical sensor having an output and defining a field of view;

determining, using a processing unit, a gesture has been performed in the electro-optical sensor field of view based on the electro-optical sensor output, wherein the determined gesture corresponds to an image capture command; and

capturing an image to the digital camera in response to the determined gesture corresponding to the image capture command.

9. The method according to claim **8** wherein the determined gesture includes a hand motion.

10. The method according to claim **8** wherein the determined gesture includes a pose.

11. The method according to claim **8** wherein the electro-optical sensor includes first and second sensors in fixed relation relative to the digital camera.

12. The method according to claim **8** wherein the electro-optical sensor defines a resolution less than a resolution defined by the digital camera.

13. An image capture device comprising:

a device housing including a forward facing portion, the forwarding facing portion encompassing a digital camera adapted to capture an image and having a field of view and encompassing a sensor adapted to detect a gesture in the digital camera field of view; and

a processing unit operatively coupled to the sensor and to the digital camera, wherein the processing unit is adapted to:

detect a gesture has been performed in the electro-optical sensor field of view based on an output of the electro-optical sensor, and

correlate the gesture detected by the sensor with an image capture function and subsequently capture an image using the digital camera, wherein the detected gesture is identified by the processing unit apart from a plurality of gestures.

14. The image capture device of claim **13** wherein the detected gesture includes a hand motion.

15. The image capture device of claim **13** wherein the detected gesture includes a pose.

16. The image capture device of claim **13** further including a forward facing light source.

17. The image capture device of claim **13** wherein the sensor defines a resolution less than a resolution defined by the digital camera.

18. The image capture device of claim **13** wherein the sensor is fixed in relation to the digital camera.

* * * * *

EXHIBIT D

US008553079B2

(12) **United States Patent**
Pryor

(10) **Patent No.: US 8,553,079 B2**
(45) **Date of Patent: Oct. 8, 2013**

(54) **MORE USEFUL MAN MACHINE INTERFACES AND APPLICATIONS**

(71) Applicant: **Timothy R. Pryor**, Sylvania, OH (US)

(72) Inventor: **Timothy R. Pryor**, Sylvania, OH (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **13/714,748**

(22) Filed: **Dec. 14, 2012**

(65) **Prior Publication Data**

US 2013/0169535 A1 Jul. 4, 2013

Related U.S. Application Data

(63) Continuation of application No. 12/700,055, filed on Feb. 4, 2010, which is a continuation of application No. 10/866,191, filed on Jun. 14, 2004, now abandoned, which is a continuation of application No. 09/433,297, filed on Nov. 3, 1999, now Pat. No. 6,750,848.

(60) Provisional application No. 60/107,652, filed on Nov. 9, 1998.

(51) **Int. Cl.**
H04N 9/47 (2006.01)
H04N 7/18 (2006.01)

(52) **U.S. Cl.**
USPC **348/77**; 348/155

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,909,002 A	9/1975	Levy	
4,219,847 A	8/1980	Pinkney et al.	
4,339,798 A	7/1982	Hedges et al.	
4,631,676 A	12/1986	Pugh	
4,791,589 A	12/1988	Blazo et al.	
4,843,568 A	6/1989	Krueger et al.	
4,908,704 A	3/1990	Fujioka et al.	
4,988,981 A	1/1991	Zimmerman et al.	
5,008,946 A	4/1991	Ando	
5,088,928 A	2/1992	Chan	
5,227,986 A	7/1993	Yokota et al.	
5,249,053 A	9/1993	Jain	
5,297,061 A	3/1994	Dementhon et al.	
5,365,597 A	11/1994	Holeva	
5,376,796 A	12/1994	Chan et al.	
5,388,059 A	2/1995	DeMenthon	
5,454,043 A	9/1995	Freeman	
5,459,793 A *	10/1995	Naoi et al.	382/165
5,491,507 A	2/1996	Umezawa et al.	
5,528,263 A *	6/1996	Platzker et al.	345/156
5,534,921 A	7/1996	Sawanobori	
5,572,251 A *	11/1996	Ogawa	348/207.99
5,581,276 A	12/1996	Cipolla et al.	
5,594,469 A	1/1997	Freeman et al.	
5,616,078 A *	4/1997	Oh	463/8
5,624,117 A	4/1997	Ohkubo et al.	
5,781,647 A	7/1998	Fishbine et al.	
5,781,650 A	7/1998	Lobo et al.	
5,808,672 A *	9/1998	Wakabayashi et al.	348/220.1
5,828,770 A	10/1998	Leis et al.	
5,845,006 A	12/1998	Sumi et al.	
5,853,327 A	12/1998	Gilboa	
5,864,334 A *	1/1999	Sellers	345/168

(Continued)

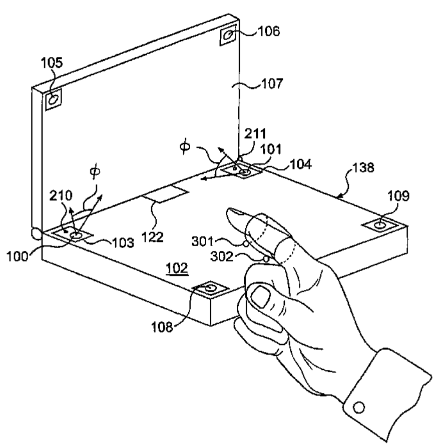
Primary Examiner — Peling Shaw

(74) *Attorney, Agent, or Firm* — Warner Norcross & Judd LLP

(57) **ABSTRACT**

A method for determining a gesture illuminated by a light source utilizes the light source to provide illumination through a work volume above the light source. A camera is positioned to observe and determine the gesture performed in the work volume.

30 Claims, 7 Drawing Sheets



(56) **References Cited**

U.S. PATENT DOCUMENTS

5,878,174 A 3/1999 Stewart et al.
5,904,484 A 5/1999 Burns
5,926,168 A 7/1999 Fan
5,936,610 A * 8/1999 Endo 345/157
5,940,126 A 8/1999 Kimura
5,982,352 A 11/1999 Pryor
5,999,840 A 12/1999 Grimson et al.
6,052,132 A 4/2000 Christian et al.
6,098,458 A 8/2000 French et al.
6,108,033 A 8/2000 Ito et al.
6,148,100 A 11/2000 Anderson et al.
6,160,899 A 12/2000 Lee et al.
6,204,852 B1 3/2001 Kumar et al.
6,252,598 B1 * 6/2001 Segen 715/863
6,342,917 B1 1/2002 Amenta
6,346,929 B1 * 2/2002 Fukushima et al. 345/8
6,359,647 B1 3/2002 Sengupta et al.
6,363,160 B1 3/2002 Bradski et al.
6,373,472 B1 4/2002 Palalau et al.
6,442,465 B2 8/2002 Breed et al.
6,508,709 B1 1/2003 Karmarkar
6,529,617 B1 3/2003 Prokoski
6,597,817 B1 7/2003 Silverbrook
6,663,491 B2 12/2003 Watabe et al.
6,750,848 B1 6/2004 Pryor
6,775,361 B1 8/2004 Arai et al.
6,788,336 B1 9/2004 Silverbrook
6,911,972 B2 6/2005 Brinjes
7,489,863 B2 2/2009 Lee

* cited by examiner

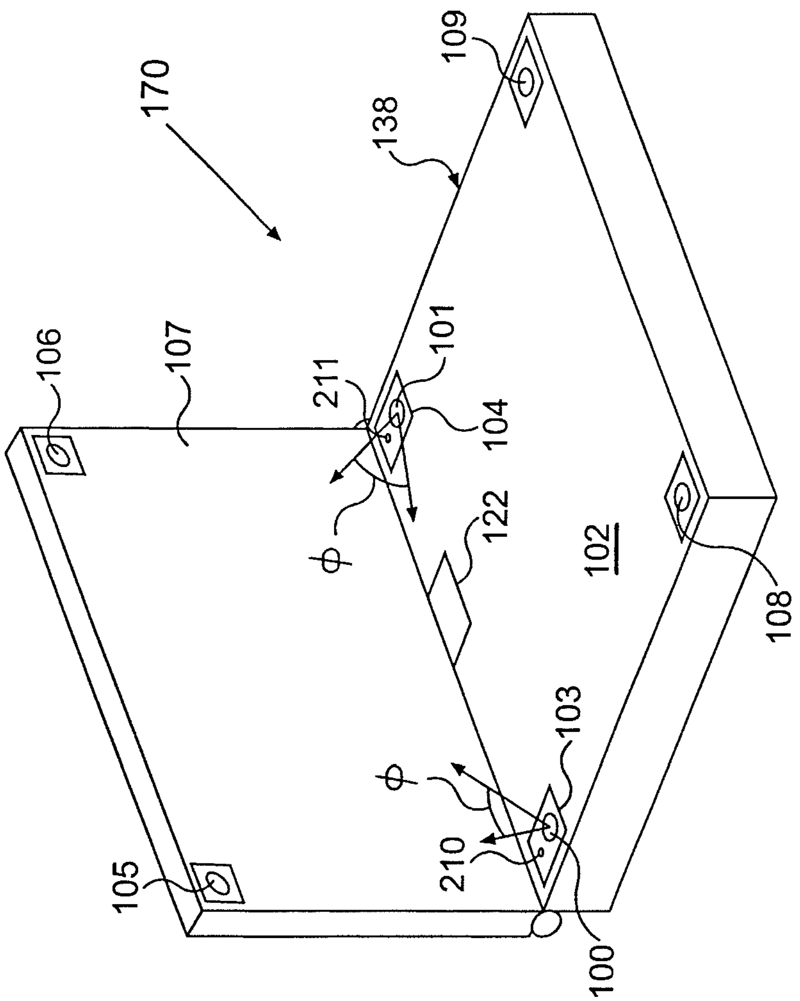
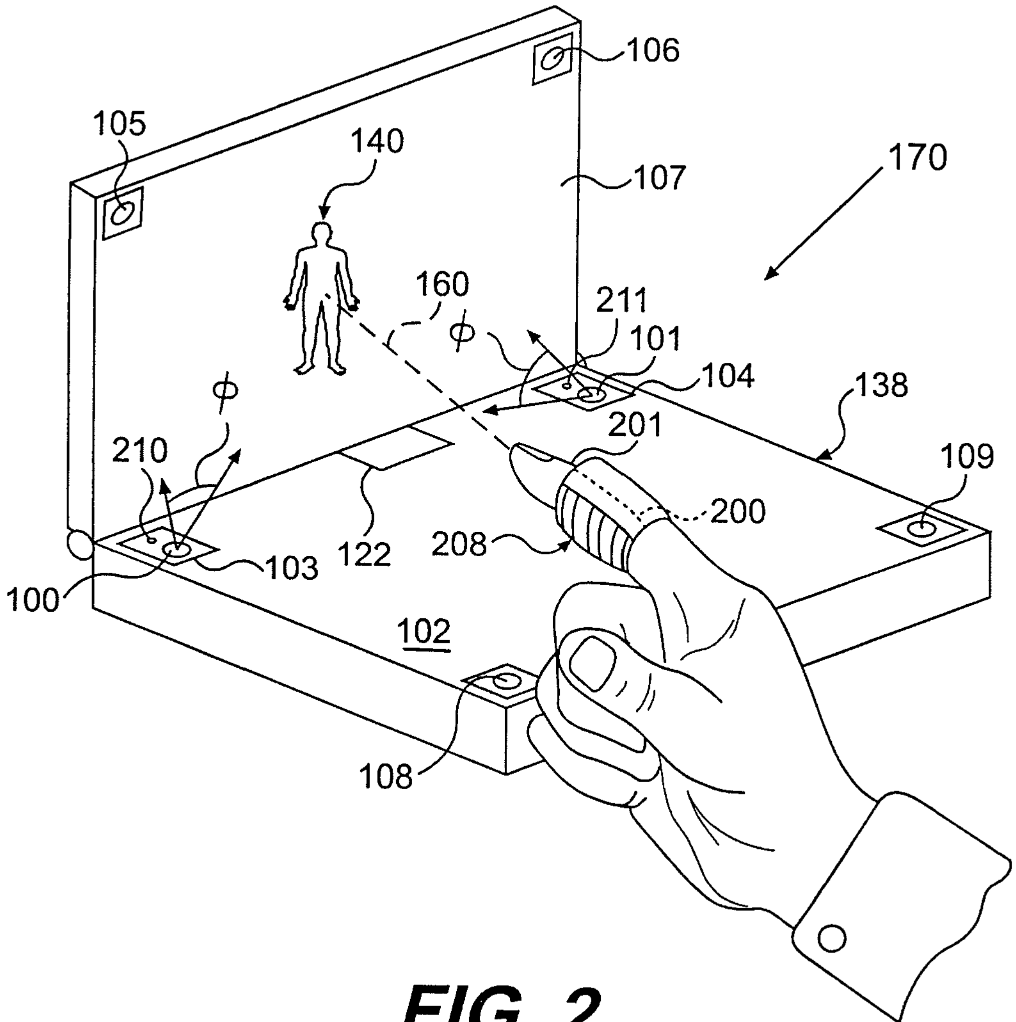
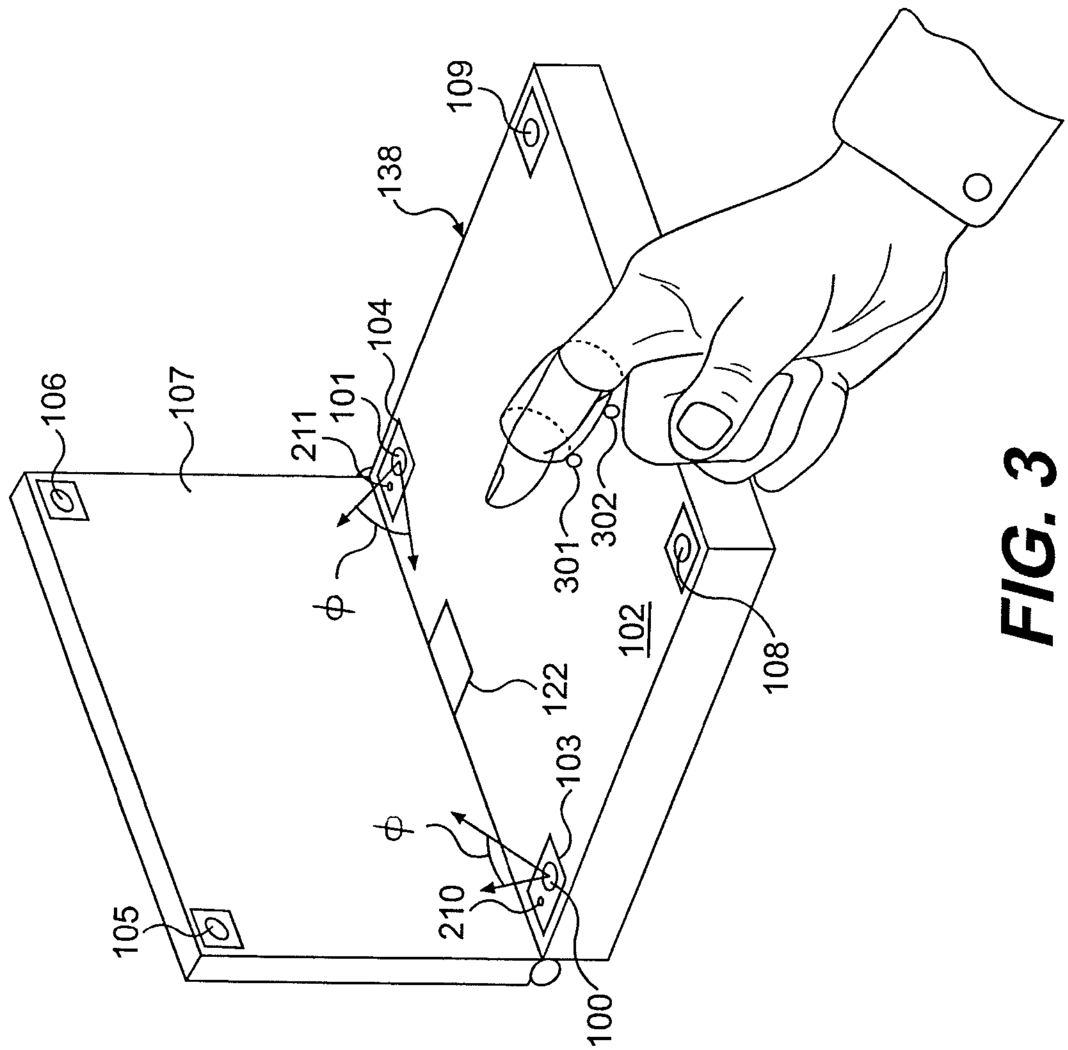


FIG. 1



106
105
140
107
170
160
ϕ
211
101
104
138
ϕ
210
201
109
200
208
122
103
100
102
108

FIG. 2



100
101
102
103
104
105
106
107
108
109
122
138
210
211
301
302
ϕ
ϕ

FIG. 3

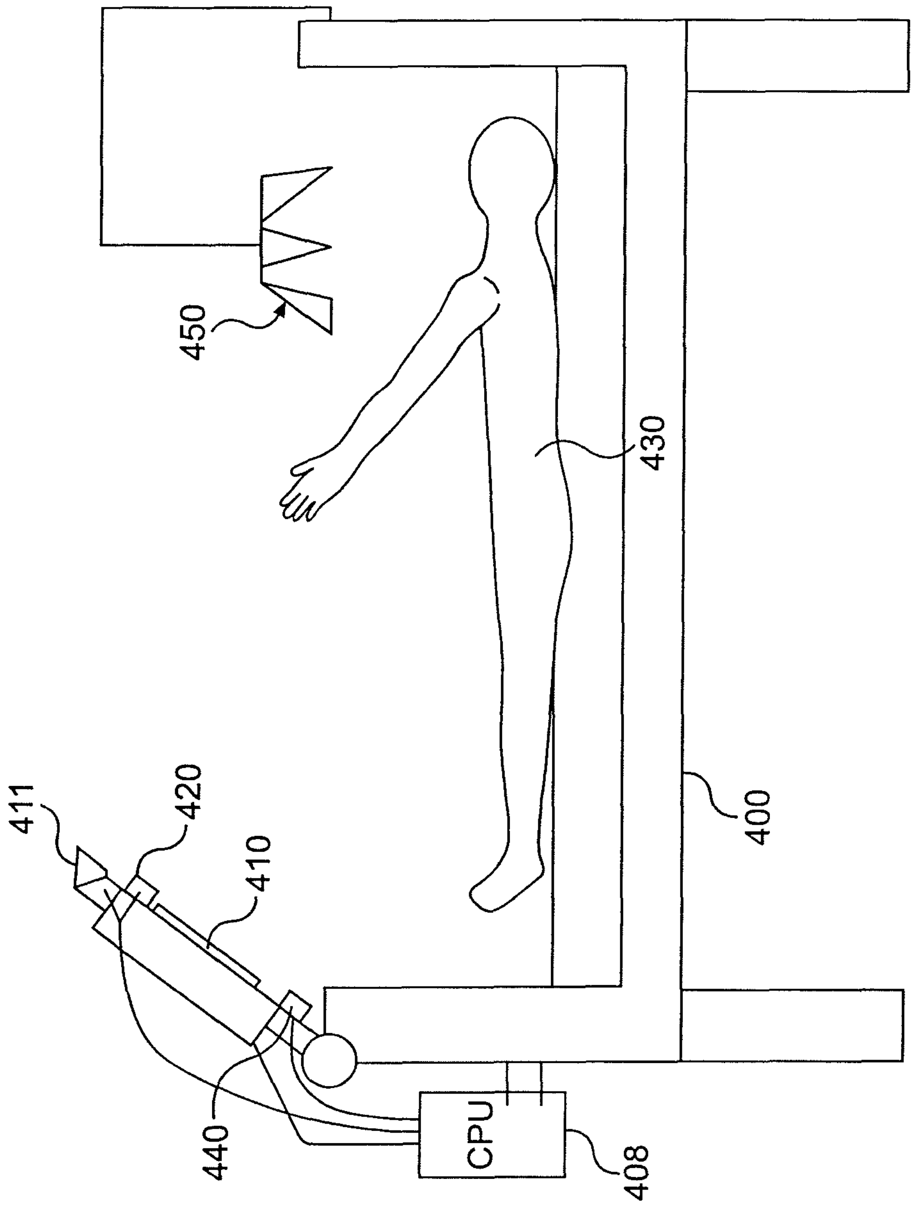


FIG. 4

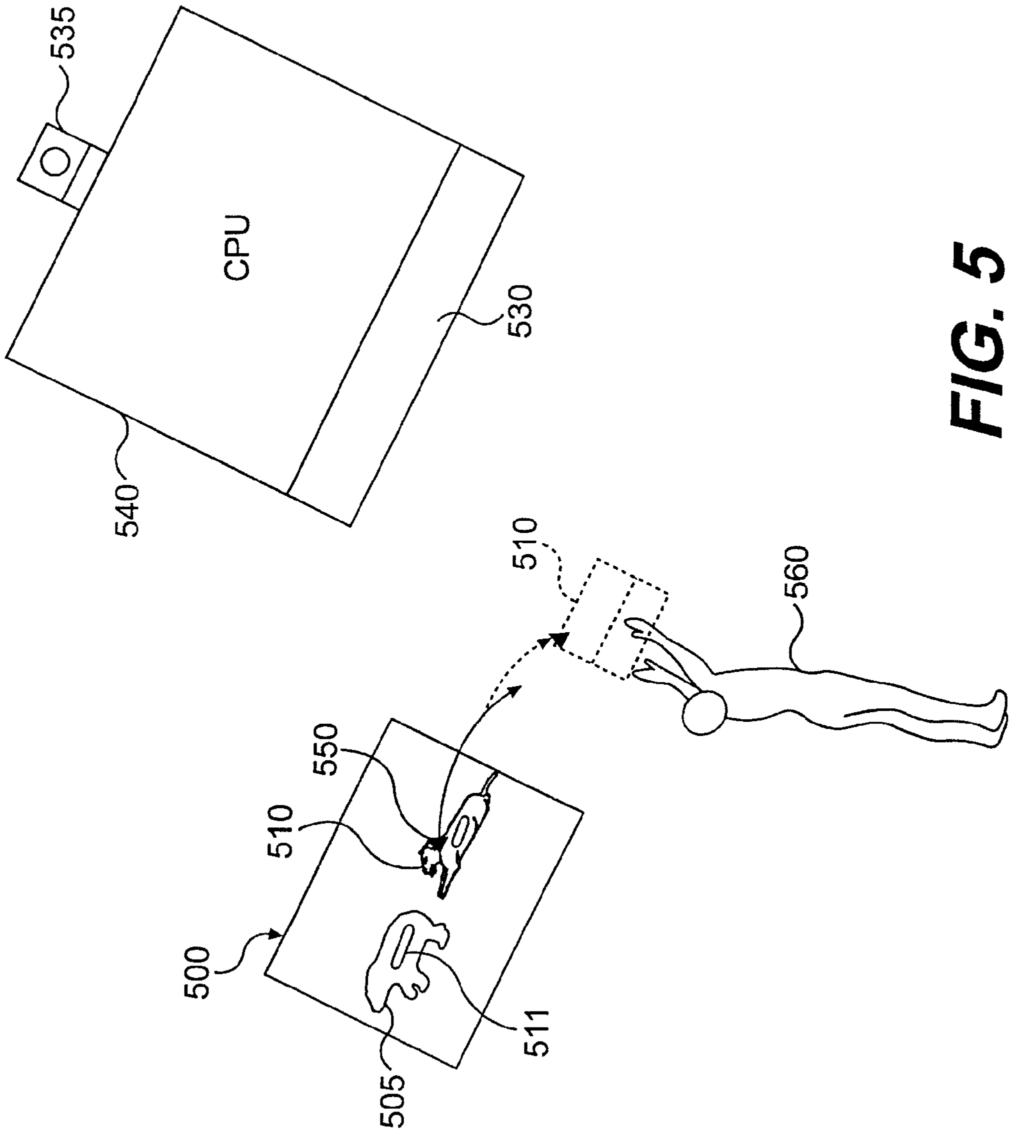
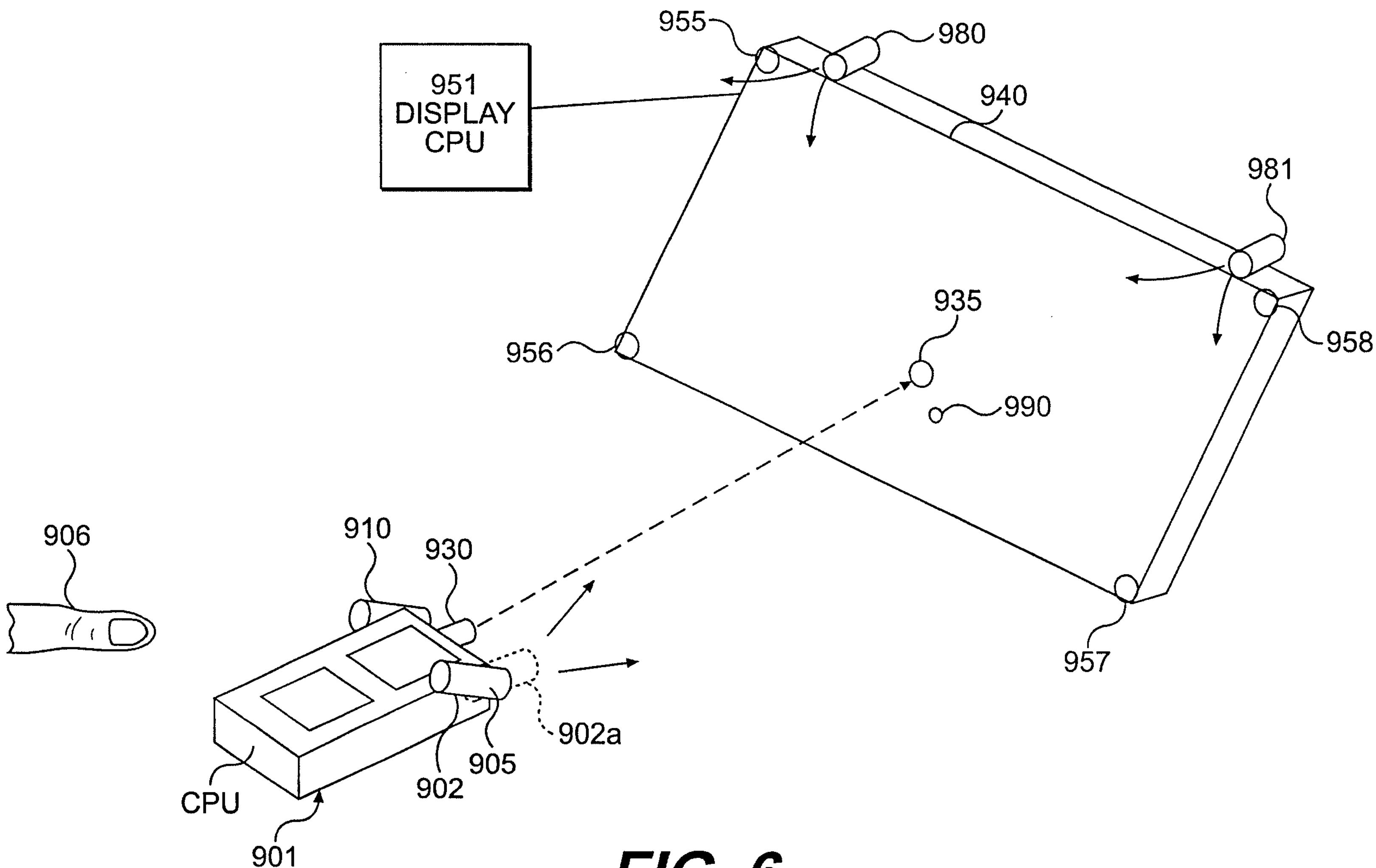


FIG. 5



951
DISPLAY
CPU
955
980
940
981
935
956
958
990
906
910
930
957
902a
905
902
CPU
901

FIG. 6

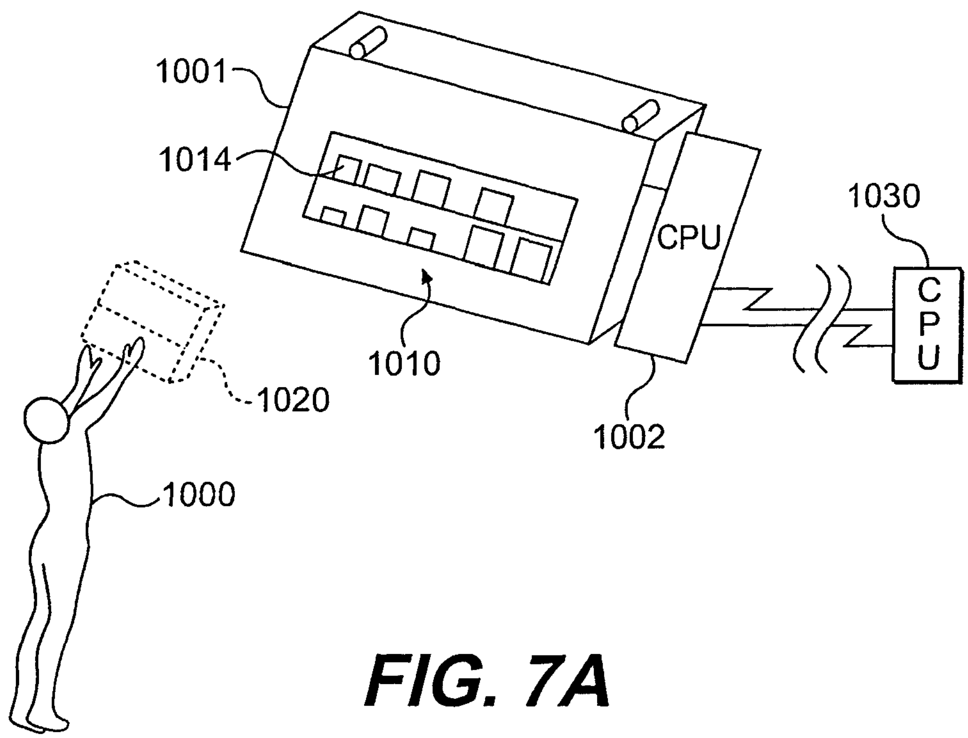

1001
1014
CPU
1030
CPU
1010
1020
1002
1000

FIG. 7A

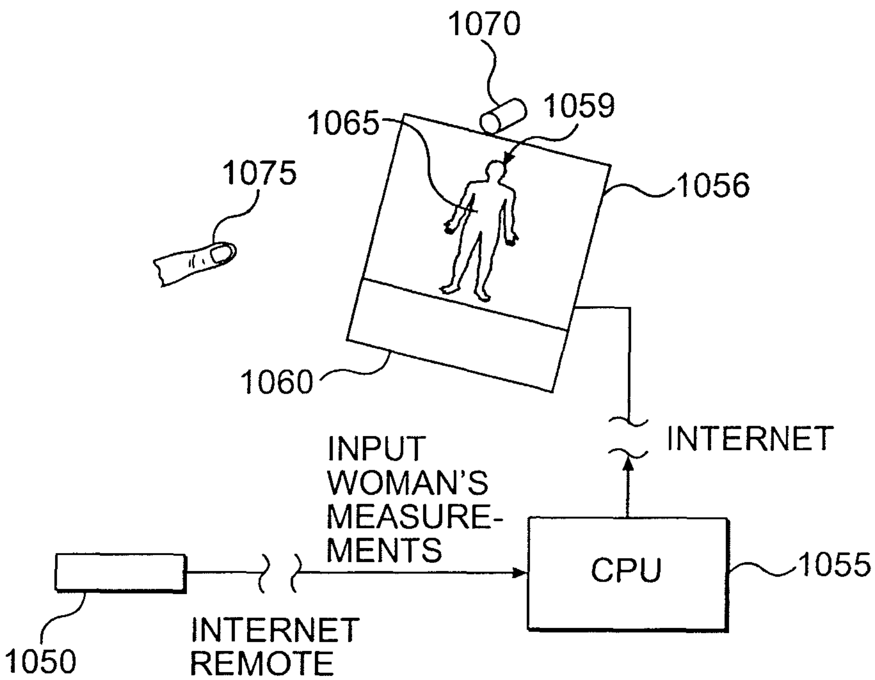

1070
1059
1065
1075
1056
1060
INPUT
WOMAN'S
MEASURE-
MENTS
INTERNET
CPU
1055
1050
INTERNET
REMOTE

FIG. 7B

MORE USEFUL MAN MACHINE INTERFACES AND APPLICATIONS

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. patent application Ser. No. 12/700,055, filed Feb. 4, 2010, which is a continuation of U.S. patent application Ser. No. 10/866,191, filed Jun. 14, 2004, which is a continuation of U.S. patent application Ser. No. 09/433,297, filed Nov. 3, 1999 (now U.S. Pat. No. 6,750,848), which claims benefit of U.S. Provisional Application No. 60/107,652, filed Nov. 9, 1998. These applications are hereby incorporated by reference.

REFERENCES TO RELATED APPLICATIONS BY THE INVENTORS

U.S. patent application Ser. No. 09/138,339, filed Aug. 21, 1998.

U.S. Provisional Application No. 60/056,639, filed Aug. 22, 1997.

U.S. Provisional Application No. 60/059,561, filed Sep. 19, 1998.

Man Machine Interfaces: Ser. No. 08/290,516, filed Aug. 15, 1994, and now U.S. Pat. No. 6,008,800.

Touch TV and Other Man Machine Interfaces: Ser. No. 08/496,908, filed Jun. 29, 1995, and now U.S. Pat. No. 5,982,352.

Systems for Occupant Position Sensing: Ser. No. 08/968,114, filed Nov. 12, 1997, now abandoned, which claims benefit of Ser. No. 60/031,256, filed Nov. 12, 1996.

Target holes and corners: U.S. Ser. No. 08/203,603, filed Feb. 28, 1994, and Ser. No. 08/468,358 filed Jun. 6, 1995, now U.S. Pat. No. 5,956,417 and U.S. Pat. No. 6,044,183.

Vision Target Based Assembly: U.S. Ser. No. 08/469,429, filed Jun. 6, 1995, now abandoned; Ser. No. 08/469,907, filed Jun. 6, 1995, now U.S. Pat. No. 6,301,763; Ser. No. 08/470,325, filed Jun. 6, 1995, now abandoned; and Ser. No. 08/466,294, filed Jun. 6, 1995, now abandoned.

Picture Taking Method and Apparatus: Provisional Application No. 60/133,671, filed May 11, 1998.

Methods and Apparatus for Man Machine Interfaces and Related Activity: Provisional Application No. 60/133,673 filed May 11, 1998.

Camera Based Man-Machine Interfaces: Provisional Patent Application No. 60/142,777, filed Jul. 8, 1999.

The copies of the disclosure of the above referenced applications are incorporated herein by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates to simple input devices for computers, particularly, but not necessarily, intended for use with 3-D graphically intensive activities, and operating by optically sensing object or human positions and/or orientations. The invention in many preferred embodiments, uses real time stereo photogrammetry using single or multiple TV cameras whose output is analyzed and used as input to a personal computer, typically to gather data concerning the 3D location of parts of, or objects held by, a person or persons.

This continuation application seeks to provide further detail on useful embodiments for computing. One embodiment is a keyboard for a laptop computer (or stand alone keyboard for any computer) that incorporates digital TV cameras to look at points on, typically, the hand or the finger, or objects held in the hand of the user, which are used to input data to the computer. It may also or alternatively, look at the head of the user as well.

Both hands or multiple fingers of each hand, or an object in one hand and fingers of the other can be simultaneously observed, as can alternate arrangements as desired.

2. Description of Related Art

My referenced co-pending applications incorporated herein by reference discuss many prior art references in various pertinent fields, which form a background for this invention.

BRIEF DESCRIPTION OF FIGURES

FIG. **1** illustrates a laptop or other computer keyboard with cameras according to the invention located on the keyboard surface to observe objects such as fingers and hands overhead of the keyboard.

FIG. **2** illustrates another keyboard embodiment using special datums or light sources such as LEDs.

FIG. **3** illustrates a further finger detection system for laptop or other computer input.

FIG. **4** illustrates learning, amusement, monitoring, and diagnostic methods and devices for the crib, playpen and the like.

FIG. **5** illustrates a puzzle toy for young children having cut out wood characters according to the invention.

FIG. **6** illustrates an improved handheld computer embodiment of the invention, in which the camera or cameras may be used to look at objects, screens and the like as well as look at the user along the lines of FIG. **1**.

FIGS. **7**A-B illustrate new methods for internet commerce and other activities involving remote operation with 3D virtual objects display.

DESCRIPTION OF THE INVENTION

FIG. **1**

A laptop (or other) computer keyboard based embodiment is shown in FIG. **1**. In this case, a stereo pair of cameras **100** and **101** located on each side of the keyboard are used, desirably having cover windows **103** and **104** mounted flush with the keyboard surface **102**. The cameras are preferably pointed obliquely inward at angles Φ toward the center of the desired work volume **170** above the keyboard. In the case of cameras mounted at the rear of the keyboard (toward the display screen), these cameras are also inclined to point toward the user at an angle as well.

Alternate camera locations may be used such as the positions of cameras **105** and **106**, on upper corners of screen housing **107** looking down at the top of the fingers (or hands, or objects in hand or in front of the cameras), or of cameras **108** and **109** shown.

One of the referenced embodiments of the invention is to determine the pointing direction vector **160** of the user's finger (for example pointing at an object displayed on screen **107**), or the position and orientation of an object held by the user. Alternatively, finger position data can be used to determine gestures such as pinch or grip, and other examples of relative juxtaposition of objects with respect to each other, as has been described in co-pending referenced applications. Positioning of an object or portions (such as hands or fingers of a doll) is also of use, though more for use with larger keyboards and displays.

In one embodiment, shown in FIG. **2**, cameras such as **100**/**101** are used to simply look at the tip of a finger **201** (or thumb) of the user, or an object such as a ring **208** on the finger. Light from below, such as provided by single central light **122** can be used to illuminate the finger that typically looks bright under such illumination.

It is also noted that the illumination is directed or concentrated in an area where the finger is typically located such as in work volume **170**. If the light is of sufficient spectral content, the natural flesh tone of the finger can be observed—and recognized by use of the color TV cameras **100**/**101**.

As is typically the case, the region of the overlapping cameras viewing area is relatively isolated to the overlapping volumetric zone of their fields **170** shown due to focal lengths of their lenses and the angulation of the camera axes with respect to each other. This restricted overlap zone helps mitigate against unwanted matches in the two images due to information generated outside the zone of overlap. Thus there are no significant image matches found of other objects in the room, since the only flesh-toned object in the zone is typically the finger or fingers of the user. Or alternatively, for example, the user's hand or hands. Similarly objects or targets thereon can be distinguished by special colors or shapes.

If desired, or required, motion of the fingers can be also used to further distinguish their presence vis-a-vis any static background. If for example, by subtraction of successive camera frames, the image of a particular object is determined to have moved it is determined that this is likely the object of potential interest which can be further analyzed directly to determine if is the object of interest.

In case of obscuration of the fingers or objects in the hand, cameras in additional locations such as those mentioned above, can be used to solve for position if the view of one or more cameras is obscured.

The use of cameras mounted on both the screen and the keyboard allows one to deal with obscurations that may occur and certain objects may or may not be advantageously delineated in one view or the other.

In addition, it may be in many cases desirable to have a datum on the top of the finger as opposed to the bottom because on the bottom, it can get in the way of certain activities. In this case the sensors are required on the screen looking downward or in some other location such as off the computer entirely and located overhead has been noted in previous application.

To determine finger location, a front end processor like that described in the target holes and corners co-pending application reference incorporated U.S. Ser. Nos. 08/203,603 and 08/468,358 can be used to also allow the finger shape as well as color to be detected.

Finger gestures comprising a sequence of finger movements can also be detected by analyzing sequential image sets such as the motion of the finger, or one finger with respect to another such as in pinching something can be determined. Cameras **100** and **101** have been shown at the rear of the keyboard near the screen or at the front. They may mount in the middle of the keyboard or any other advantageous location.

The cameras can also see one's fingers directly, to allow typing as now, but without the physical keys. One can type in space above the plane of the keyboard (or in this case plane of the cameras). This is useful for those applications where the keyboard of conventional style is too big (e.g., the hand held computer of FIG. **6**).

FIG. **2**

It is also desirable for fast reliable operation to use retro-reflective materials and other materials to augment the contrast of objects used in the application. For example, a line target such as **200** can be worn on a finger **201**, and advantageously can be located if desired between two joints of the finger as shown. This allows the tip of the finger to be used to type on the keyboard without feeling unusual—the case perhaps with target material on tip of the finger.

The line image detected by the camera can be provided also by a cylinder such as retroreflective cylinder **208** worn on the finger **201** which effectively becomes a line image in the field of view of each camera (assuming each camera is equipped with a sufficiently coaxial light source, typically one or more LEDs such as **210** and **211**), can be used to solve easily using the line image pairs with the stereo cameras for the pointing direction of the finger that is often a desired result. The line, in the stereo pair of images provides the pointing direction of the finger, for example pointing at an object displayed on the screen **140** of the laptop computer **138**.

FIG. **3**

It is also possible to have light sources on the finger that can be utilized such as the 2 LED light sources shown in FIG. **3**. This can be used with either TV camera type sensors or with PSD type analog image position sensors as disclosed in references incorporated.

In particular the ring mounted LED light sources **301** and **302** can be modulated at different frequencies that can be individually discerned by sensors imaging the sources on to a respective PSD detector. Alternatively, the sources can simply be turned on and off at different times such that the position of each point can be independently found allowing the pointing direction to be calculated from the LED point data gathered by the stereo pair of PSD based sensors.

The "natural interface keyboard" here described can have cameras or other sensors located at the rear looking obliquely outward toward the front as well as inward so as to have their working volume overlap in the middle of the keyboard such as the nearly full volume over the keyboard area is accommodated.

Clearly larger keyboards can have a larger working volume than one might have on a laptop. The pair of sensors used can be augmented with other sensors mounted on the screen housing. It is noted that the linked dimension afforded for calibration between the sensors located on the screen and those on the keyboard is provided by the laptop unitary construction.

One can use angle sensing means such as a rotary encoder for the laptop screen tilt. Alternatively, cameras located on the screen can be used to image reference points on the keyboard as reference points to achieve this. This allows the calibration of the sensors mounted fixedly with respect to the screen with respect to the sensors and keyboard space below. It also allows one to use stereo pairs of sensors that are not in the horizontal direction (such as **101**/**102**) but could for example be a camera sensor such as **100** on the keyboard coupled with one on the screen, such as **106**.

Knowing the pointing angles of the two cameras with respect to one another allows one to solve for the 3D location of objects from the matching of the object image positions in the respective camera fields.

As noted previously, it is also of interest to locate a line or cylinder type target on the finger between the first and second joints. This allows one to use the fingertip for the keyboard activity but by raising the finger up, it can be used as a line target capable of solving for the pointed direction for example.

Alternatively one can use two point targets on the finger such as either retroreflective datums, colored datums such as rings or LED light sources that can also be used with PSD detectors which has also been noted in FIG. **2**.

When using the cameras located for the purpose of stereo determination of the position of the fingers from their flesh tone images it is useful to follow the preprocessing capable of processing data obtained from the cameras in order to look for the finger. This can be done on both color basis and on the basis of shape as well as motion.

In this invention, I have shown the use of not only cameras located on a screen looking downward or outward from the screen, but also cameras that can be used instead of or in combination with those on the screen placed essentially on the member on which the keyboard is incorporated. This allows essentially the keyboard to mounted cameras which are preferably mounted flush with the keyboard surface to be unobtrusive, and yet visually be able to see the users fingers, hands or objects held by the user and in some cases, the face of the user.

This arrangement is also useful for 3D displays, for example where special synchronized glasses (e.g., the "Crystal Eyes" brand often used with Silicon Graphics work stations) are used to alternatively present right and left images to each eye. In this case the object may appear to be actually in the workspace **170** above the keyboard, and it may be manipulated by virtually grasping (pushing, pulling, etc.) it, as has been described in co-pending applications.

FIG. **4**: Baby Learning and Monitoring System

A baby's reaction to the mother (or father) and the mother's analysis of the baby's reaction is very important. There are many gestures of babies apparently indicated in child psychology as being quite indicative of various needs, wants, or feelings and emotions, etc. These gestures are typically made with the baby's hands.

Today this is done and learned entirely by the mother being with the baby. However with an Electro-optical sensor based computer system, such as that described in co-pending applications located proximate to or even in the crib (for example), one can have the child's reactions recorded, not just in the sense of a video tape which would be too long and involved for most to use, but also in terms of the actual motions which could be computer recorded and analyzed also with the help of the mother as to what the baby's responses were. And such motions, combined with other audio and visual data can be very important to the baby's health, safety, and learning.

Consider for example crib **400** with computer **408** having LCD monitor **410** and speaker **411** and camera system (single or stereo) **420** as shown, able to amuse or inform baby **430**, while at the same time recording (both visually, aurally, and in movement detected position data concerning parts of his body or objects such as rattles in his hand) his responses for any or all of the purposes of diagnosis of his state of being, remote transmission of his state, cues to various programs or images to display to him or broadcast to others, or the like.

For one example, baby's motions could be used to signal a response from the TV either in the absence of the mother or with the mother watching on a remote channel. This can even be over the Internet if the mother is at work.

For example, a comforting message could come up on the TV from the mother that could be prerecorded (or alternatively could actually be live with TV cameras in the mother's or father's workplace for example on a computer used by the parent) to tell the baby something reassuring or comfort the baby or whatever. Indeed the parent can be monitored using the invention and indicate something back or even control a teleoperater robotic device to give a small child something to eat or drink for example. The same applies to a disabled person.

If the father or mother came up on the screen, the baby could wave at it, move its head or "talk" to it but the hand gestures may be the most important.

If the mother knows what the baby is after, she can talk to baby or say something, or show something that the baby recognizes such as a doll. After a while, looking at this live one can then move to talking to the baby from some prerecorded data.

What other things might we suppose? The baby for example knows to puts its hand on the mother's cheek to cause the mother to turn to it. The baby also learns some other reflexes when it is very young that it forgets when it gets older. Many of these reflexes are hand movements, and are important in communicating with the remote TV based mother representation, whether real via telepresense or from CD Rom or DVD disk (or other media, including information transmitted to the computer from afar) and for the learning of the baby's actions.

Certainly just from the making the baby feel good point-of-view, it would seem like certain motherly (or fatherly, etc.) responses to certain baby actions in the form of words and images would be useful. This stops short of physical holding of the baby which is often needed, but could act as a stop gap to allow the parents to get another hour's sleep for example.

As far as the baby touching things, I've discussed in other applications methods for realistic touch combined with images. This leads to a new form of touching crib mobiles that could contain video imaged and or be imaged themselves—plus if desired—touched in ways that would be far beyond any response that you could get from a normal mobile.

For example, let us say there is a targeted (or otherwise TV observable) mobile **450** in the crib above the baby. Baby reaches up and touches a piece of the mobile which is sensed by the TV camera system (either from the baby's hand position, the mobile movement, or both, and a certain sound is called up by the computer, a musical note for example. Another piece of the mobile and another musical note. The mobile becomes a musical instrument for the baby that could play either notes or chords or complete passages, or any other desired programmed function.

The baby can also signal things. The baby can signal using agitated movements would often mean that it's unhappy. This could be interpreted using learned movement signatures and artificial intelligence as needed by the computer to call for mother even if the baby wasn't crying. If the baby cries, that can be picked up by microphone **440**, recognized using a voice recognition system along the lines of that used in IBM Via Voice commercial product for example. And even the degree of crying can be analyzed to determine appropriate action.

The computer could also be used to transmit information of this sort via the internet email to the mother who could even be at work. And until help arrives in the form of mother intervention or whatever, the computer could access a program that could display on a screen for the baby things that the baby likes and could try to soothe the baby through either images of familiar things, music or whatever. This could be useful at night when parents need sleep, and anything that would make the baby feel more comfortable would help the parents.

It could also be used to allow the baby to input to the device. For example, if the baby was hungry, a picture of the bottle could be brought up on the screen. The baby then could yell for the bottle. Or if the baby needed his diaper changed, perhaps something reminiscent of that. If the baby reacts to such suggestions of his problem, this gives a lot more intelligence as to why he is crying and while mothers can generally tell right away, not everyone else can. In other words, this is pretty neat for babysitters and other members of the household so they can act more intelligently on the signals the baby is providing.

Besides in the crib, the system as described can be used in conjunction with a playpen, hi-chair or other place of baby activity.

As the child gets older, the invention can further be used also with more advanced activity with toys, and to take data from toy positions as well. For example, blocks, dolls, little cars, and moving toys even such as trikes, scooters, drivable toy cars and bikes with training wheels.

The following figure illustrates the ability of the invention to learn, and thus to assist in the creation of toys and other things.

FIG. **5**: Learning Puzzle Roy

Disclosed in FIG. **5** is a puzzle toy **500** where woodcut animals such as bear **505** and lion **510** are pulled out with handle such as **511**. The child can show the animal to the camera and a computer **530** with TV camera (or cameras) **535** can recognize the shape as the animal, and provide a suitable image and sounds on screen **540**.

Alternatively, and more simply, a target, or targets on the back of the animal can be used such as triangle **550** on the back of lion **511**. In either case the camera can solve for the 3D, and even 5 or 6D position and orientation of the animal object, and cause it to move accordingly on the screen as the child maneuvers it. The child can hold two animals, one in each hand and they can each be detected, even with a single camera, and be programmed in software to interact as the child wishes (or as he learns the program).

This is clearly for very young children of two or three years of age. The toys have to be large so they can't be swallowed.

With the invention in this manner, one can make a toy of virtually anything, for example a block. Just hold this block up, teach the computer/camera system the object and play using any program you might want to represent it and its actions. To make this block known to the system, the shape of the block, the color of the block or some code on the block can be determined. Any of those items could tell the camera which block it was, and most could give position and orientation if known.

At that point, an image is called up from the computer representing that particular animal or whatever else the block is supposed to represent. Of course this can be changed in the computer to be a variety of things if this is something that is acceptable to the child. It could certainly be changed in size such as a small lion could grow into a large lion. The child could probably absorb that more than a lion changing into a giraffe for example since the block wouldn't correspond to that. The child can program or teach the system any of his blocks to be the animal he wants and that might be fun.

For example, he or the child's parent could program a square to be a giraffe where as a triangle would be a lion. Maybe this could be an interesting way to get the child to learn his geometric shapes!

Now the basic block held up in front of the camera system could be looked at just for what it is. As the child may move the thing toward or away from the camera system, one may get a rough sense of depth from the change in shape of the object. However this is not so easy as the object changes in shape due to any sort of rotations.

Particularly interesting then is to also sense the rotations if the object so that the animal can actually move realistically in 3 Dimensions on the screen. And perhaps having the de-tuning of the shape of the movement so that the child's relatively jerky movements would not appear jerky on the screen or would not look so accentuated. Conversely of course, you can go the other way and accentuate the motions.

This can, for example, be done with a line target around the edge of the object is often useful for providing position or orientation information to the TV camera based analysis software, and in making the object easier to see in reflective illumination.

Aid to Speech Recognition

The previous co-pending application entitled "Useful man machine interfaces and applications" referenced above, discussed the use of persons movements or positions to aid in recognizing the voice spoken by the person.

In one instance, this can be achieved by simply using ones hand to indicate to the camera system of the computer that the voice recognition should start (or stop, or any other function, such as a paragraph or sentence end, etc.).

Another example is to use the camera system of the invention to determine the location of the persons head (or other part), from which one can instruct a computer to preferentially evaluate the sound field in phase and amplitude of two or more spaced microphones to listen from that location—thus aiding the pickup of speech—which often times is not able to be heard well enough for computer based automatic speech recognition to occur.

Digital Interactive TV

As you watch TV, data can be taken from the camera system of the invention and transmitted back to the source of programming. This could include voting on a given proposition by raising your hand for example, with your hand indication transmitted. Or you could hold up 3 fingers, and the count of fingers transmitted. Or in a more extreme case, your position, or the position of an object or portion thereof could be transmitted—for example you could buy a coded object—whose code would be transmitted to indicate that you personally (having been pre-registered) had transmitted a certain packet of data.

If the programming source can transmit individually to you (not possible today, but forecast for the future), then much more is possible. The actual image and voice can respond using the invention to positions and orientations of persons or objects in the room—just as in the case of prerecorded data—or one to one internet connections. This allows group activity as well.

In the extreme case, full video is transmitted in both directions and total interaction of users and programming sources and each other becomes possible.

An interim possibility using the invention is to have a program broadcast to many, which shifts to prerecorded DVD disc or the like driving a local image, say when your hand input causes a signal to be activated.

Handwriting Authentication

A referenced co-pending application illustrated the use of the invention to track the position of a pencil in three dimensional space such that the point at which the user intends the writing point to be at, can be identified and therefore used to input information, such as the intended script.

As herein disclosed, this part of the invention can also be used for the purpose of determining whether or not a given person's handwriting or signature is correct.

For example, consider authentication of an Internet commercial transaction. In this case, the user simply writes his name or address and the invention is used to look at the movements of his writing instrument and determine from that whether or not the signature is authentic. (A movement of one or more of his body parts might also or alternatively be employed). For example a series of frames of datum location on his pen can be taken, to determine one or more positions on it as a function of time, even to include calculating of its pointing direction, from a determined knowledge in three axes of two points along the line of the pen axis. In this case a particular pointing vector sequence "signature" would be learned for this person, and compared to later signatures.

What is anticipated here is that in order to add what you might call the confirming degree of authenticity to the signature, it may not be necessary to track the signature completely. Rather one might only determine that certain aspects of the movement of the pencil are the authentic ones. One could have people write using any kind of movement, not just their signature having their name. The fact is that people are mostly used to writing their name and it would be assumed that that would be it. However, it could well be that the computer asks the user to write something else that they would then write and that particular thing would be stored in the memory.

Optionally, one's voice could be recognized in conjunction with the motion signature to add further confirmation.

This type of ability for the computer system at the other end of the Internet to query a writer to write a specific thing in a random fashion adds a degree of cryptographic capacity to the invention. In other words, if I can store the movements in my hand to write different things, then clearly this has some value.

The important thing though is that some sort of representation of the movements of the pencil or other instrument can be detected using the invention and transmitted.

FIG. **6**: Hand Held Computer

FIG. **6** illustrates an improved handheld computer embodiment of the invention. For example, FIG. **8** of the provisional application referenced above entitled "camera based man machine interfaces and applications" illustrates a basic hand held device and which is a phone, or a computer or a combination thereof, or alternatively to being hand held, can be a wearable computer for example on one's wrist.

In this embodiment, we further disclose the use of this device as a computer, with a major improvement being the incorporation of a camera of the device optionally in a position to look at the user, or an object held by the user—along the lines of FIG. **1** of the instant disclosure for example.

Consider hand held computer **901** of FIG. **6**, incorporating a camera **902** which can optionally be rotated about axis **905** so as to look at the user or a portion thereof such as finger **906**, or at objects at which it is pointed. Optionally, and often desirably, a stereo pair of cameras to further include camera **910** can also be used. It too may rotate, as desired. Alternatively fixed cameras can be used as in FIG. **1**, and FIG. **8** of the referenced co-pending application, when physical rotation is not desired, for ruggedness, ease of use, or other reasons (noting that fixed cameras have fixed fields of view, which limit versatility in some cases).

When aimed at the user, as shown, it can be used, for example, to view and obtain images of:

Ones self-facial expression etc., also for image reasons, id etc., combined effect.

Ones fingers (any or all), one finger to other and the like. This in turn allows conversing with the computer in a form of sign language which can replace the keyboard of a conventional computer.

One or more pbjects in one's hand. Includes a pencil or pen, and thus can be used rather than having a special touch screen and pencil if the pencil itself is tracked as disclosed in the above figure. It also allows small children to use the device, and those who cannot hold an ordinary stylus.

One's Gestures.

The camera **902** (and **910** if used, and if desired), can also be optionally rotated and used to viewpoints in space ahead of the device, as shown in dotted lines **902***a*. In this position for example it can be used for the purposes described in the previous application. It can also be used to observe or point at (using optional laser pointer **930**) points such as **935** on a wall, or a mounted LCD or projection display such as **940** on a wall or elsewhere such as on the back of an airline seat.

With this feature of the invention, there is no requirement to carry a computer display with you as with a infrared connection (not shown) such as known in the art one can also transmit all normal control information to the display control computer **951**. As displays become ubiquitous, this makes increasing sense—otherwise the displays get bigger the computers smaller trend doesn't make sense if they need to be dragged around together. As one walks into a room, one uses the display or displays in that room (which might themselves be interconnected).

The camera unit **902** can sense the location of the display in space relative to the handheld computer, using for example the four points **955**-**958** on the corners of the display as references. This allows the handheld device to become an accurate pointer for objects displayed on the screen, including control icons. And it allows the objects on the screen to be sensed directly by the camera—if one does not have the capability to spatially synchronize and coordinate the display driver with the handheld computer.

The camera can also be used to see gestures of others, as well as the user, and to acquire raw video images of objects in its field.

A reverse situation also exists where the cameras can be on the wall mounted display, such as cameras **980** and **981** can be used to look at the handheld computer module **901** and determine its position and orientation relative to the display.

Note that a camera such as **902**, looking at you the user, if attached to hand held unit, always has reference frame of that unit. If one works with a screen on a wall, one can aim the handheld unit with camera at it, and determine its reference frame to the handheld unit. Also can have two cameras operating together, one looking at wall thing, other at you (as **902** and **902***a*) in this manner, one can dynamically compare ref frames of the display to the human input means in determining display parameters. This can be done in real time, and if so one can actually wave the handheld unit around while still imputing accurate data to the display using ones fingers, objects or whatever.

Use of a laser pointer such as **930** incorporated into the handheld unit has also been disclosed in the referenced co-pending applications. For example, a camera on the hand held computer unit such as **902** viewing in direction **902***a* would look at laser spot such as **990** (which might or might not have come from the computers own laser pointer **930**) on the wall display say, and recognized by color and size/shape reference to edge of screen, and to projected spots on screen.

FIGS. **7**A-B: Internet and Other Remote Applications

FIG. **7**A illustrates new methods for internet commerce and other activities involving remote operation with 3D virtual objects displayed on a screen. This application also illustrates the ability of the invention to prevent computer vision eye strain.

Let us first consider the operation of the invention over the internet as it exists today in highly bandwidth limited form dependent on ordinary phone lines for the most part. In this case it is highly desirable to transmit just the locations or pointing vectors of portions (typically determined by stereo photo-grammetry of the invention) of human users or objects associated therewith to a remote location, to allow the remote computer **10** to modify the image or sound transmitted back to the user.

Another issue is the internet time delay, which can exist in varying degrees, and is more noticeable, the higher resolution of the imagery transmitted. In this case, a preferred arrangement is to have real time transmission of minimal position and vector data (using no more bandwidth than voice), and to transmit back to the user, quasi stationary images at good resolution. Transmission of low resolution near real time images common in internet telephony today, does not convey the natural feeling desired for many commercial applications to now be discussed. As bandwidth becomes more plentiful these restrictions are eased.

Let us consider the problem posed of getting information from the internet of today. A user **1000** can go to a virtual library displayed on screen **1001** controlled by computer **1002** where one sees a group **1010** of books on stacks. Using the invention as described herein and incorporated referenced applications to determine my hand and finger locations, I the user, can point at a book such as **1014** in a computer sensed manner, or even reach out and "grab" a book, such as **1020** (dotted lines) apparently generated in 3D in front of me.

My pointing, or my reach and grab is in real time, and the vector (such as the pointing direction of ones finger at the book on the screen, or the position and orientation closing vectors of one's forefinger and thumb to grab the 3D image **1020** of the book) indicating the book in question created is transmitted back by internet means to the remote computer **1030** which determines that I have grabbed the book entitled War and Peace from the virtual shelf. A picture of the book coming off the shelf is then generated using fast 3D graphical imagery such as the Merlin VR package available today from Digital Immersion company of Sudbury, Ontario. This picture (and the original picture of the books on the shelves) can be retransmitted over the internet at low resolution (but sufficient speed) to give a feeling of immediacy to the user. Or alternatively, the imagery can be generated locally at higher resolution using the software package resident in the local computer **1002** which receives key commands from the distant computer **1030**.

After the book has been "received" by the user, it then can be opened automatically to the cover page for example under control of the computer, or the users **10** hands can pretend to open it, and the sensed hands instruct the remote (or local, depending on version) computer to do so. A surrogate book such as **1040** can also be used to give the user a tactile feel of a book, even though the real book in questions pages will be viewed on the display screen **1001**. One difference to this could be if the screen **1001** depicting the books were life size, like real stacks. Then one might wish to go over to a surrogate book incorporating a separate display screen—just as one would in a real library, go to a reading table after removing a book from a stack.

Net Grocery stores have already appeared, and similar applications concern picking groceries off of the shelf of a virtual supermarket, and filling ones shopping cart. For that matter, any store where it is desired to show the merchandise in the very manner people are accustomed to seeing it, namely on shelves or racks, generally as one walks down an aisle, or fumbles through a rack of clothes for example. In each case, the invention, which also can optionally use voice input, as if to talk to a clothing sales person, can be used to monitor the person's positions and gestures.

The invention in this mode can also be used to allow one to peruse much larger objects. For example, to buy a car (or walk through a house, say) over the internet, one can lift the hood, look inside, etc., all by using the invention to monitor the 3D position of your head or hands and move the image of the car presented accordingly. If the image is presented substantially life-size, then one can be monitored as one physically walks around the car in one's room say, with the image changing accordingly. In other words just as today.

Note that while the image can be apparently life-size using virtual reality glasses, the natural movements one is accustomed to in buying a car are not present. This invention makes such a natural situation possible (though it can also be used with such glasses as well).

It is noted that the invention also comprehends adding a force based function to a feedback to your hands, such that it feels like you lifted the hood, or grabbed the book, say. For this purpose holding a surrogate object as described in co-pending applications could be useful, in this case providing force feedback to the object.

If one looks at internet commerce today, some big applications have turned out **10** to be clothes and books. Clothes are by far the largest expenditure item, and let's look closer at this.

Consider too a virtual mannequin, which can also have measurements of a remote shopper. For example, consider diagram **78**, where a woman's measurements are inputted by known means such as a keyboard **1050** over the internet to a CAD program in computer **1055**, which creates on display screen **1056** a 3D representation of a mannequin **1059** having the woman's shape in the home computer **1060**. As she selects a dress **1065** to try on, the dress which let's say comes in 10 sizes, 5 to 15, is virtually "tried on" the virtual mannequin and the woman **1070** looks at the screen **1056** and determines the fit of a standard size 12 dress. She can rapidly select larger or smaller sizes and decide which she thinks looks and/or fits better.

Optionally, she can signal to the computer to rotate the image in any direction, and can look at it from different angles up or down as well, simply doing a rotation in the computer. This signaling can be conventional using for example a mouse, or can be using TV based sensing aspects of the invention such as employing camera **1070** also as shown in FIG. **1** for example. In another such case, she can reach out with her finger **1075** for example, and push or pull in a virtual manner the material, using the camera to sense the direction of her finger. Or she can touch herself at the points where the material should be taken up or let out, with the camera system sensing the locations of touch (typically requiring at least a stereo pair of cameras or other electro-optical system capable of determining where her fingertip is in 3D space. Note that a surrogate for the tried on dress in this case, could be the dress she has on, which is touched in the location desired on the displayed dress.

The standard size dress can then be altered and shipped to her, or the requisite modifications can be made in the CAD program, and a special dress cut out and sewed which would fit better.

A person can also use her hands via the TV cameras of the invention to determine hand location relative to the display to take clothes off a virtual manikin which could have a representation of any person real or imaginary. Alternatively she can remotely reach out using the invention to a virtual rack of clothes such as **1090**, take an object off the rack, and put it on the manikin. This is particularly natural in near life-size representation, just like being in a store or other venue. This ability of the invention to bring real life experience to computer shopping and other activity that is a major advantage.

The user can also feel the texture of the cloth if suitable haptic devices are 15 available to the user, which can be activated remotely by the virtual clothing program, or other type of program.

Modifications of the invention herein disclosed will occur to persons skilled in the art, and all such modifications are deemed to be within the scope of the invention as defined by the appended claims.

The invention claimed is:

1. A computer implemented method comprising:

providing a light source adapted to direct illumination through a work volume above the light source;

providing a camera oriented to observe a gesture performed in the work volume, the camera being fixed relative to the light source; and

determining, using the camera, the gesture performed in the work volume and illuminated by the light source.

2. The method according to claim **1** wherein the light source includes a light emitting diode.

3. The method according to claim **1** wherein the light source includes a plurality of light emitting diodes.

4. The method according to claim **1** wherein detecting a gesture includes analyzing sequential images of the camera.

5. The method according to claim **1** wherein the detected gesture includes at least one of a pinch gesture, a pointing gesture, and a grip gesture.

6. The method according to claim **1** further including determining the pointing direction of a finger in the work volume.

7. The method according to claim **1** further including providing a target positioned on a user that is viewable in the work volume.

8. The method according to claim **1** further including determining the three-dimensional position of a point on a user.

9. The method according to claim **1** wherein the camera and the light source are positioned in fixed relation relative to a keypad.

10. The method according to claim **9** the camera, the light source and the keypad form part of a laptop computer.

11. A computer apparatus comprising:

a light source adapted to illuminate a human body part within a work volume generally above the light source;

a camera in fixed relation relative to the light source and oriented to observe a gesture performed by the human body part in the work volume; and

a processor adapted to determine the gesture performed in the work volume and illuminated by the light source based on the camera output.

12. The computer apparatus of claim **11** further including a display and a keyboard, wherein the work volume is above the keyboard and in front of the display.

13. The computer apparatus of claim **12** wherein the display is pivotable relative to the keyboard.

14. The computer apparatus of claim **11** wherein the light source includes a light emitting diode.

15. The computer apparatus of claim **11** wherein the light source includes a plurality of light emitting diodes.

16. The computer apparatus of claim **12** wherein the display includes a three-dimensional display.

17. The computer apparatus of claim **11** further including a target that is viewable by the camera when in the work volume.

18. The computer apparatus of claim **11** wherein the determined gesture includes a pinch gesture.

19. The computer apparatus of claim **11** wherein the determined gesture includes a pointing gesture.

20. The computer apparatus of claim **11** wherein the determined gesture includes a grip gesture.

21. A computer implemented method comprising:

providing a camera oriented to observe a gesture performed in a work volume above the camera;

providing a light source in fixed relation relative to the camera and adapted to direct illumination through the work volume; and

detecting, using the camera, a gesture performed by at least one of a user's fingers and a user's hand in the work volume.

22. The method according to claim **21** wherein the light source includes a light emitting diode.

23. The method according to claim **21** wherein the light source includes a plurality of light emitting diodes.

24. The method according to claim **21** wherein detecting a gesture includes analyzing sequential images of the camera.

25. The method according to claim **21** wherein the detected gesture includes at least one of a pinch gesture, a pointing gesture, and a grip gesture.

26. The method according to claim **21** further including determining the pointing direction of one of the user's fingers using the first and second cameras.

27. The method according to claim **21** further including providing a target positioned on the user that is viewable by the camera.

28. The method according to claim **21** further including determining the three-dimensional position of a point on at least one of the user's hand and the user's fingers.

29. The method according to claim **21** further including providing a three-dimensional display viewable by the user.

30. The method according to claim **21** wherein the camera and the light source are positioned in fixed relation relative to a keypad.

* * * * *