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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2008 (12.09.2008)

PCT

(10) International Publication Number
WO 2008/109567 A4

(51) International Patent Classification:
G06K 19/00 (2006.01)

(21) International Application Number:
PCT/US2008/055728

(22) International Filing Date: 3 March 2008 (03.03.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/892,738 2 March 2007 (02.03.2007) US

(71) Applicant *(for all designated States except US)*: **ORGANIC MOTION** [—/US]; 336 West 37th Street, Studio 530, New York, NY 10018 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **TSCHESNOK, Andrew** [US/US]; New York, NY (US).

(74) Agents: **DOBROW, James, F.** et al.; Lowenstein Sandler P.C, 65 Livingston Avenue, Roseland, NJ 07068 (US).

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

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

Published:
— *with international search report*
— *with amended claims*

(88) Date of publication of the international search report:
13 November 2008

Date of publication of the amended claims: 19 February 2009

(54) Title: SYSTEM AND METHOD FOR TRACKING THREE DIMENSIONAL OBJECTS

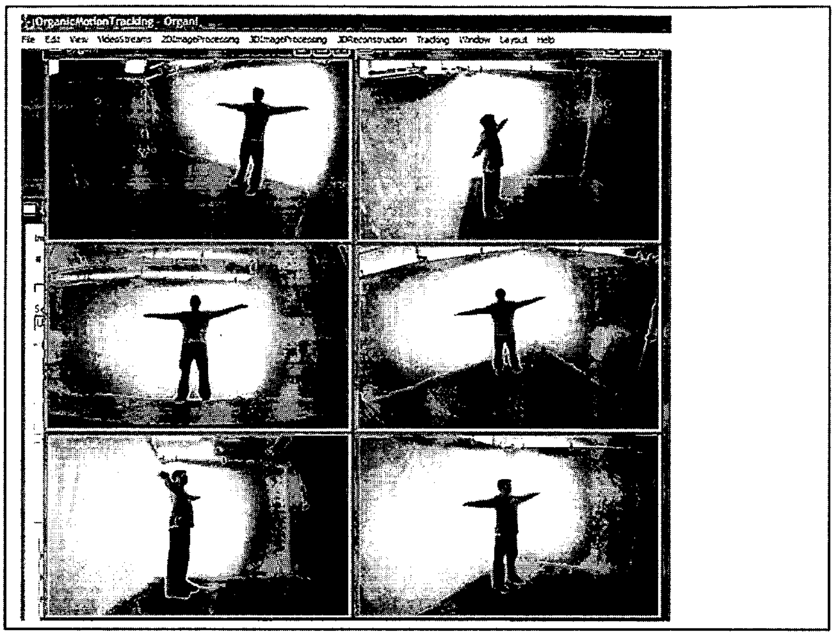

Figure 1

(57) Abstract: Embodiments of the invention are directed to improved systems and methods for three dimensional (3D) image reconstruction. The systems and methods are directed to extracting, digitally reconstructing and tracking 3D objects from multiple two dimensional (2D) video camera sources. The systems and methods are directed to reconstructing a 3D scene via 2D cameras and then re-projecting this data back onto 2D surfaces. This systems and methods can greatly simplify the image processing required to analyze the 3D model by moving the analysis techniques back into the 2D domain.

WO 2008/109567 A4

AMENDED CLAIMS
received by the International Bureau on 30 September 2008 (30.09.2008)

1. A method for creating a 3D representation of a target object in a 3D digital domain, the method comprising:
   (i) placing a target object within a background space;
   (ii) directing at least two video cameras at said target object;
   (iii) acquiring a total image from each of said video cameras directed at the target object;
   (iv) subtracting said background space from each of said total images to produce an target image of the target object;
   (v) performing internal calibration of each of said video cameras;
   (vi) performing relative external calibration of two or more of said video cameras; and
   (vii) reconstructing a 3D image of said target object via projections of one or more of said target images into said 3D digital domain.

2. The method of claim 1, wherein the background subtraction, internal camera calibration and image reconstruction steps are sequentially performed.

3. The method of claim 1, wherein the background subtraction, internal camera calibration and image reconstruction steps are performed substantially simultaneously.

4. The method of claim 1, wherein the relative external camera calibration step is performed prior to the image capture step.

5. The method of claim 1, wherein the internal camera calibration comprises analyzing the distortion of at least one lens on each of said video cameras and compensating for any image shift introduced by the optical properties of said lens.

6. The method of claim 5, wherein said internal camera calibration further comprises calculating properties of said lens, said properties selected from the group consisting of: the focal point of said lens, the center point of said lens as it is attached to said video camera, and the extent to which said lens may be skewed as it is attached to said video camera.

7. The method of claim 6, wherein said step of calculating properties of said lens comprises:
   (i) placing a checkerboard pattern in the background space such that said video camera acquires one or more images of said checkerboard pattern from various angles;
   (ii) ascertaining the intersection points of each square on each of said one or more images of said checkerboard pattern; and
   (iii) using said intersection points to calculate the lens parameters.

8. The method of claim 7, wherein at least thirty or more images of said checkerboard pattern are acquired.

9. The method of claim 1, wherein at least five or more video cameras are directed at said target object.

10. The method of claim 1, wherein at least ten or more video cameras are directed at said target object.

11. The method of claim 1, wherein at least one of said video cameras is selected from the group of: a digital video camera, an analog video camera, a black-and-white video camera, and a color video camera.

12. The method of claim 1, wherein said background subtraction comprises:
   (a) placing an active background in said background space behind said target object;
   (b) causing said active background to glow;
   (c) determining whether each pixel of said image is lighter than a predetermined threshold;
   (d) characterizing each pixel which is determined to be brighter than said predetermined threshold as a background pixel, and characterizing each pixel which is determined to be less bright than said predetermined threshold as a foreground pixel; and
   (e) retaining each foreground pixel as said image of the target object.

13. The method of claim 1, wherein said background subtraction comprises:

(a) capturing a background image of said background space prior to placing said target object within said background space;

(b) securing a foreground image of said target after placing said target object within said background space;

(c) performing a pixel-per-pixel subtraction of the background image from the foreground image;

(d) characterizing all pixels remaining in the foreground image as a foreground pixel; and

(e) retaining each foreground pixel as said image of the target object.

14. The method of claim 1, wherein said background subtraction comprises:

(a) capturing a background image of said background space prior to placing said target object within said background space;

(b) securing a foreground image of said target after placing said target object within said background space;

(c) converting each of said background image and said foreground image to a hue-luminosity-saturation color-space;

(d) characterizing a pixel in the color-space according to one or more rules selected from the group consisting of:

(i) characterizing a pixel of said foreground image as a foreground pixel if said pixel in said foreground image is lighter in saturation than a corresponding pixel in said background image;

(ii) characterizing a pixel of said foreground image as a background pixel if said pixel in said foreground image is darker in saturation than a corresponding pixel in said background image, the hue value of said pixel in said foreground image is within a predetermined range of said corresponding pixel in said background image, and the difference in saturation and luminance between said pixel in said foreground image and said corresponding pixel in said background image is not outside a predetermined range;

(iii) characterizing a pixel of said foreground image as a foreground pixel if said pixel in said foreground image if the hue value of said pixel in said foreground image is outside a predetermined range of said corresponding pixel in said background image; and

(iv) characterizing a pixel as the foreground if the foreground and background differ in saturation beyond a predefined difference, clipping the difference and setting the background the foreground to set the pixel type; and

(e) retaining each foreground pixel as said image of the target object.

15. The method of claim 1, wherein said background subtraction comprises using region-based techniques which analyze changes in gradients in said background space versus a foreground space.

16. The method of claim 1, wherein the relative external camera calibration comprises:

(a) placing a checkerboard pattern in the background space such that each of said video cameras acquires one or more images of said checkerboard pattern from various angles;

(b) capturing one or more images of said checkerboard pattern on each of said video cameras;

(c) ascertaining the intersection points of each square on each of said one or more images of said checkerboard pattern; and

(d) using said intersection points to calculate the position of each of said video cameras.

17. The method of claim 16, further comprising building a map of the positional relationships between each of said video cameras based on each of said video camera's relative position to said checkerboard pattern.

18. The method of claim 17, wherein said map building of positional relationships comprises:

(a) determining which of said video cameras is in a best location to capture said images of said checkerboard; and

(b) orienting all others of said video cameras to said video camera determined to be in a best location to capture said images of said checkerboard.

19. The method of claim 18, further comprising calculating the standard deviation and removing outliers from the calculation.

20. The method of claim 1, wherein said image reconstruction via projections of one or more of said acquired images into said 3D digital domain comprises projecting said image acquired from each of said video cameras into said 3D digital domain in the direction, and from the position, of said video camera, and with the lens properties of said video camera.

21. The method of claim 20, wherein said image reconstruction is accomplished using hardware that supports shader language.

22. The method of claim 20, wherein said image reconstruction further comprises:
(a) conceptualizing one or more virtual projectors in said 3D digital domain in a configuration analogous to the configuration of said one or more video cameras;
(b) inserting one or more virtual 2D slices into said 3D digital domain;
(c) comparing each of said virtual 2D slices to each of said projected images;
(d) incrementing a counter each time one of said projected images intersects one of said virtual 2D slices;
(e) comparing said counter to the number of said one or more video cameras; and
(f) deciding whether said target object is considered to exist at the location of said virtual 2D slice in said 3D digital domain.

23. The method of claim 22, wherein said image reconstruction further comprises dynamically configuring the characteristics of each of said virtual 2D slices to best fit the characteristics of said target object.

24. The method of claim 22, wherein each of said virtual 2D slices are inserted into said 3D digital domain such that each of said virtual 2D slices is parallel to every other of said virtual 2D slices.

25. The method of claim 1, wherein the method further comprises tracking said target object in said 3D digital domain.

26. The method of claim 25, wherein the method further comprises detecting a surface texture of said target object; and tracking said surface texture of said target object.

27. The method of claim 25, wherein said tracking of said target object comprises:
   (a) performing 2D image analysis on each of said acquired images projected into said 3D digital domain;
   (b) detecting the presence of said target object based on the known characteristics of said target object; and
   (c) creating a rag doll virtual representation of said target object in said 3D digital domain.

28. The method of claim 26, wherein the 2D image analysis comprises:
   (a) assessing each virtual 2D slice for one or more image representation constructs;
   (b) calculating the outline of each of said image representation constructs;
   (c) fitting an ellipse to each of said image representation constructs;
   (d) calculating the centerpoint, x- and y-radii, and an angle of rotation for each of said image representation constructs; and
   (e) applying the known properties of each virtual 2D slice to each of said image representation constructs located within said virtual 2D slice to generate real-world coordinates of each of said image representation constructs.

29. The method of claim 26, wherein detection of the target object comprises:
   (a) conceptualizing one or more virtual projectors in said 3D digital domain in a configuration analogous to the configuration of said one or more video cameras;
   (b) inserting one or more virtual 2D slices into said 3D digital domain at one or more locations predetermined based on the known characteristics of said target object;
   (c) comparing each of said virtual 2D slices to each of said projected images;
   (d) incrementing a counter each time one of said projected images intersects one of said virtual 2D slices;

(e) comparing said counter to the number of said one or more video cameras;

(g) deciding whether said target object is considered to exist at the location of said virtual 2D slice in said 3D digital domain;

(g) assessing each virtual 2D slice for one or more image representation constructs;

(h) calculating the outline of each of said image representation constructs;

(i) fitting an ellipse to each of said image representation constructs;

(j) calculating the centerpoint, x- and y-radii, and an angle of rotation for each of said image representation constructs; and

(k) applying the known properties of each virtual 2D slice to each of said image representation constructs located within said virtual 2D slice to generate real-world coordinates of each of said image representation constructs.

30. The method of claim 29, wherein said target object is a human, and the method further comprises:

(a) inserting a virtual 2D slice parallel to the floor and at floor height;

(b) inserting a virtual 2D slice parallel to the floor and at ankle height;

(c) inserting a virtual 2D slice parallel to the floor and at knee height

(d) inserting a virtual 2D slice parallel to the floor and at hip height;

(e) inserting a virtual 2D slice parallel to the floor and at chest height;

(f) inserting one or more virtual 2D slices parallel to the floor and at head height.

31. The method of claim 29, further comprising determining the orientation of said human by analyzing the relationship between one or more image representation constructs located within said virtual 2D slice parallel to the floor and at floor height and one or more image representation constructs located within said 2D slice parallel to the floor and at ankle height.

32. The method of claim 29, wherein the determination of the orientation of said human further comprises analyzing the relationship between one or more image representation constructs located within said virtual 2D slice parallel to the floor and at hip height and

one or more image representation constructs located within said 2D slice parallel to the floor and at chest height.

33. The method of claim 32, further comprising determining the height of said human.

34. The method of claim 33, further comprising inserting one or more virtual 2D slices as end caps, such that the edge of said target object can be determined.

35. A system for creating and optionally tracking a 3D representation of a target object, said system comprising:
   (i) at least two video cameras directed at said target object;
   (ii) one or more video camera image acquisition modules, coupled to one or more of said video camera; and
   (iii) at least one processor-based device, comprising a direct memory access controller and a graphics processing unit, wherein said processor-based device is coupled to said one or more video camera acquisition modules, and wherein said processor-based device is configured to at least: acquire images from the image acquisition modules, perform internal camera calibration functions, perform background subtraction processing, and perform image reconstruction processing.

36. The system of claim 1, wherein at least five or more video cameras are directed at said target object.

37. The system of claim 36, wherein at least ten or more video cameras are directed at said target object.

38. The system of claim 1, wherein at least one of said video cameras is selected from the group of: a digital video camera, an analog video camera, a black-and-white video camera, and a color video camera.

39. The system of claim 1, wherein each of said at least two video cameras are oriented around said target object so as to provide the maximum number of unique perspectives available with such a number of said video cameras.

40. The system of claim 1, wherein at least one of said one or more video camera image acquisition modules is coupled to one of said video cameras via a connection selected from the group of: Firewire, Ethernet, wireless, and analog.

41. The system of claim 1, wherein said graphics processing unit of said processor-based device is a DirectX-capable, consumer-grade 3D graphics card capable of running vertex and pixel shaders.

42. The system of claim 41, wherein said graphics processing unit is a Redeon 9700 Graphics Processor.

43. The system of claim 1, wherein said processor-based device is coupled to said one or more video camera acquisition modules via direct memory access transfer.