



(86) **Date de dépôt PCT/PCT Filing Date:** 2014/05/26
(87) **Date publication PCT/PCT Publication Date:** 2014/12/04
(85) **Entrée phase nationale/National Entry:** 2016/11/14
(86) **N° demande PCT/PCT Application No.:** IL 2014/050472
(87) **N° publication PCT/PCT Publication No.:** 2014/191990
(30) **Priorité/Priority:** 2013/05/26 (US61/827,602)

(51) **Cl.Int./Int.Cl.** *H04N 5/222* (2006.01),
H04N 5/232 (2006.01), *H04N 5/262* (2006.01),
H04N 7/24 (2011.01)
(71) **Demandeur/Applicant:**
PIXELLOT LTD., IL
(72) **Inventeurs/Inventors:**
TAMIR, MIKY, IL;
OZ, GAL, IL
(74) **Agent:** INTEGRAL IP

(54) **Titre : PROCÉDE ET SYSTÈME DESTINÉS À UNE PRODUCTION TÉLÉVISÉE À FAIBLE COUT**
(54) **Title: METHOD AND SYSTEM FOR LOW COST TELEVISION PRODUCTION**

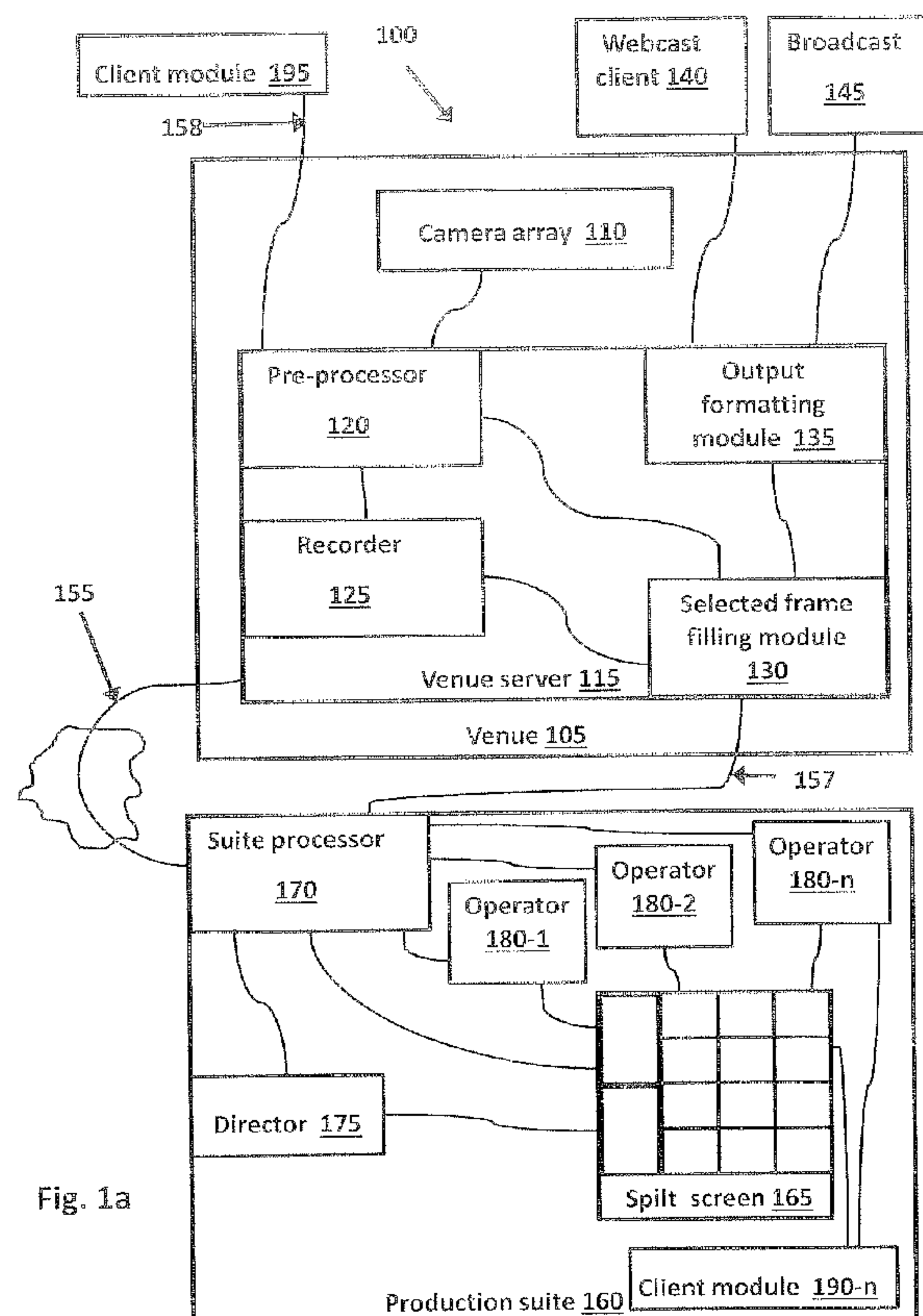


Fig. 1a

(57) **Abrégé/Abstract:**

It is provided a method for producing a stream of desired image frames derived from a panoramic capturing of a scene by an array of video cameras. The production is conducted by a coordinated operation at a venue proximate to the scene and at a production

(57) Abrégé(suite)/Abstract(continued):

suite located remotely from the venue. The method includes a venue server recording a stream of raw video images captured simultaneously by the video- cameras, and converting the stream of raw video images at a first data rate to a stream of panoramic views at a second data rate. The second data rate is less than one third of the first data rate. The method further includes transmitting the stream of panoramic views to a remote production suite, and there selecting a stream of desired frames including respective regions of interest. The method further includes transmitting data on the stream of desired frames to the venue server, and the venue server producing out of the recorded stream of raw video images, a stream of high resolution frames. The high resolution frames are characterized by non perceptible boundaries between neighboring images derived from respective neighboring raw video images.

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

(19) World Intellectual Property
Organization
International Bureau(43) International Publication Date
4 December 2014 (04.12.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/191990 A1

(51) International Patent Classification:

H04N 5/222 (2006.01) H04N 7/24 (2011.01)
H04N 5/262 (2006.01) H04N 5/232 (2006.01)

(21) International Application Number:

PCT/IL2014/050472

(22) International Filing Date:

26 May 2014 (26.05.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/827,602 26 May 2013 (26.05.2013) US

(71) Applicant: **PIXELLOT LTD.** [IL/IL]; 6 Hanechoshet St., 69710 Tel-Aviv (IL).(72) Inventors: **TAMIR, Miky**; 9 Mautitius Vitale St., 6965100 Tel-Aviv (IL). **OZ, Gal**; 7 Emek Hachula, 44622 Kfar Saba (IL).(74) Agent: **FRIEDMAN, Mark**; Dr. Mark Friedman Ltd., Moshe Aviv Tower, 54th Floor, 7 Jabotinsky St., 5252007 Ramat Gan (IL).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, 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, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: METHOD AND SYSTEM FOR LOW COST TELEVISION PRODUCTION

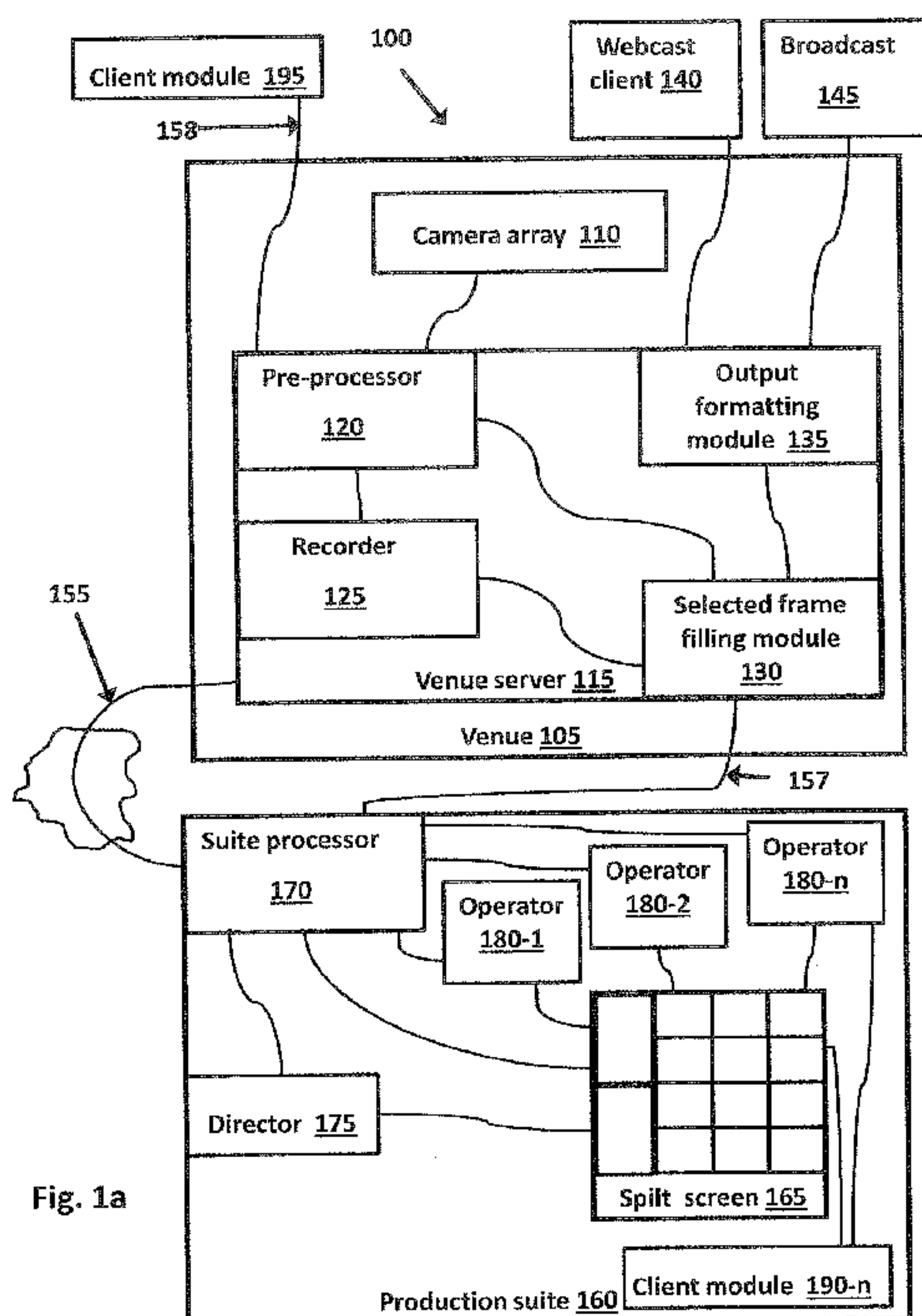


Fig. 1a

(57) Abstract: It is provided a method for producing a stream of desired image frames derived from a panoramic capturing of a scene by an array of video cameras. The production is conducted by a coordinated operation at a venue proximate to the scene and at a production suite located remotely from the venue. The method includes a venue server recording a stream of raw video images captured simultaneously by the video-cameras, and converting the stream of raw video images at a first data rate to a stream of panoramic views at a second data rate. The second data rate is less than one third of the first data rate. The method further includes transmitting the stream of panoramic views to a remote production suite, and there selecting a stream of desired frames including respective regions of interest. The method further includes transmitting data on the stream of desired frames to the venue server, and the venue server producing out of the recorded stream of raw video images, a stream of high resolution frames. The high resolution frames are characterized by non perceptible boundaries between neighboring images derived from respective neighboring raw video images.

METHOD AND SYSTEM FOR LOW COST TELEVISION PRODUCTION

CROSS REFERENCE TO RELATED APPLICATION

- 5 This application claims the benefit of priority of U.S. Provisional Application Serial No. 61/827,602, filed on May 26, 2013, entitled " METHOD AND SYSTEM FOR LOW COST TELEVISION PRODUCTION" by Miky Tamir and Gal Oz, the content of which is relied upon and incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

10 Field of the invention

The invention is in the field of an efficient low cost replacement to outdoors (mainly sports) television productions and more particularly cloud based (remote) real time and offline productions.

Description of related art

- 15 A television coverage of sports scenes is spreading to a variety of sports events, which in the past were not covered. However, the cost of full venue production including a large team and several cameras prohibits high quality coverage. On the other hand, the home users expect to get a TV production which covers the whole scene, focuses on events occurring in various locations during a game, and having replays of
20 dramatic moments. Thus, the objective of the current invention is to suggest a low cost efficient television production, to replace current outdoor productions of sports scenes that are too expensive in equipment and manpower.

BRIEF SUMMARY OF THE INVENTION

- It is provided according to preferred embodiment of the present application, a method
25 for producing a stream of desired image frames derived from a panoramic capturing of a scene by an array of video cameras. The production is conducted by a coordinated operation at a venue proximate to the scene and at a production suite located remotely from the venue. The method includes a venue server recording a

stream of raw video images captured simultaneously by the video cameras, and converting the stream of raw video images at a first data rate to a stream of panoramic views at a second data rate. The second data rate is less than one third of the first data rate. The method further includes transmitting the stream of panoramic
5 views to a remote production suite, and there selecting a stream of desired frames including respective regions of interest. The selected frames are portions of respective panoramic views. The method further includes transmitting data on the stream of desired frames to the venue server, and the venue server producing out of the recorded stream of raw video images, a stream of high resolution frames in
10 accordance with the stream of desired frames. The high resolution frames are characterized by non perceptible boundaries between neighboring images derived from respective neighboring raw video images.

In some embodiments, the method includes determining several preset fixed frames for capturing respective fixed regions of interest.

15 In some embodiments, the method includes selecting a frame from frames like a preset fixed frame determined in advance for capturing a fixed region of interest, a virtual moving camera following an object in the scene, a virtual moving camera performing a pre-defined motion, a frame determined by an algorithm for automatic detection and inclusion of a predefined event, and a frame associated with an event
20 which occurred in a certain location in the scene previously.

In some embodiments, the method includes transmitting to the venue server, for appending to a desired frame, instructions regarding data items like a graphical item, a desired artistic transition (like fade, dissolve, etc) between frames, an audio item and statistical data regarding identified events occurring in the scene.

25 In some embodiments, several arrays of video cameras are capturing the scene from two or more locations in the venue, and respective streams of panoramic views are transmitted to several production suites.

In some embodiments, a moving video camera is used in conjunction with the array of video cameras to capture the event.

30 In some embodiments, the transmitting to the production suite and the transmitting to the venue server are done using an internet link, preferably using a cloud service. In some embodiments it is done using a cellular link.

In some embodiments, the scene is a sports scene.

In some embodiments, a client module is selecting the stream of desired frames. The client module includes a displaying interface and a desired frame selector and controller. Preferably, it also includes a processing and communication unit. Exemplary displaying interfaces is a screen showing a panoramic view of a portion of a captured dynamic scene, and having a chosen field of view represented as a rectangle superimposed on the displayed scene. Another example is a monocular or binocular unit directly viewing the scene which includes a pan and tilt unit, frame size selection means and devices to measure the pan and tilt angles and the frame size for defining and controlling a desired field of view and means for graphical representation of the desired field of view.

Another possible displaying interface is a set of virtual reality goggles, which includes a display of a portion of the stream of panoramic views, and means for fitting the portion to head movement in similarity to a view seen by a real spectator of the scene having a similar head movement. Preferably, each goggle is fed by a different panoramic view such that a three dimensional scene is sensed by an operator wearing the goggles.

In some embodiments, a cloud server receives a sequence of panoramic views from the venue server, and streams the sequence of panoramic views to the client module using the world wide web.

In some embodiments, the method further includes preliminary calibrating the video cameras of the array based on mechanical design and lens specifications, and accurately calibrating of the video cameras based on the scene to get calibration parameters.

In some embodiments, the method includes further steps conducted at the venue server. The steps include decomposing frames captured by the video cameras into pixel tiles, the tiles are rectangular arrays of pixels whose size is smaller than a single camera frame size, associating the pixel tiles with time codes, and recording the pixel tiles. Preferably, the method also includes retrieving recorded pixel tiles associated with a selected frame, transforming stitching lines from world coordinates to camera coordinates and specifying the lines in the selected frame or close thereof, and geometrically aligning pixels belonging to neighboring video cameras using the calibration parameters. The method also includes blending pixel pairs in stitching zones belonging to neighboring video cameras in accordance with a distance from a stitching line, correcting color intensities of pixels belonging to different video

cameras, and adjusting hue, saturation, intensity and gamma value of the entire frame.

Additional steps conducted at the venue server prior to transmitting to the remote production suite include low quality blending and aligning of neighboring
5 frames captured by the array of video cameras, generating of low resolution panoramic view from the blended and aligned frames, and compressing and encoding the panoramic views for transmitting to the remote production suite.

In some embodiments, the method includes steps conducted at the remote production suite, like generating and distributing to operators partial frames of the
10 panoramic view in accordance with respective region of interest, the operators moving across and zooming in and out the partial frames to get actual frames, presenting to a director the actual frames, and the director selecting an actual frame to go on air.

In some embodiments, the method includes encoding the stream of high
15 resolution frames into a required output video format, and broadcasting or webcasting the formatted stream of high resolution frames.

It is provided according to preferred embodiment of the present application, a method for a production suite located remotely from a venue proximate to a scene. The production suite is operating in a coordinated way with a venue server in the
20 venue for producing a stream of desired image frames derived from a panoramic capturing of the scene by an array of video cameras. The production suite method includes receiving from the venue server a stream of low-resolution panoramic views derived from a stream of raw video images captured simultaneously by the video cameras, selecting a stream of desired frames including respective regions of interest.
25 The selected frames are portions of the low resolution panoramic views. The method also includes transmitting data on the stream of desired frames to the venue server.

In some embodiments, the method includes generating and distributing to operators in the production suite partial frames of the panoramic view in accordance with respective region of interest, the operators moving across and zooming in and
30 out the partial frames to get actual frames, presenting to a director the actual frames, and the director selecting an actual frame to go on air.

It is provided according to preferred embodiment of the present application, a method for a venue server producing, by a coordinated operation with a production

suite, a stream of desired image frames derived from a panoramic capturing of a scene by an array of video cameras. The method includes recording a stream of raw video images captured simultaneously by the video cameras, and converting the stream of raw video images to a stream of panoramic views at a reduced data rate, whereas the data rate reduction is by a factor of at least three. The method further includes transmitting the stream of panoramic views to the production suite for selecting thereof a stream of desired frames including respective regions of interest, whereas the selected frames being portions of the panoramic views, receiving from the production suite data on the stream of desired frames, and producing out of the recorded stream of raw video images, a stream of high resolution frames in accordance with the stream of desired frames. The high resolution frames are characterized by non perceptible boundaries between neighboring images derived from respective neighboring raw video images.

It is provided according to preferred embodiment of the present application, a method for a venue server producing a stream of desired image frames derived from a panoramic capturing of a scene by an array of video cameras. The method includes recording a stream of raw video images captured simultaneously by the video cameras, and receiving data on a stream of desired frames including respective regions of interest in the panoramic capturing of the scene. The method further includes producing out of the recorded stream of raw video images, a stream of high resolution frames in accordance with the stream of desired frames. Producing a frame includes actions restricted to only previously recorded images associated with the frame, like retrieving the only images associated with the frame from the recorded video images, and processing only the associated previously recorded images to get high resolution frame having non-perceptible boundaries between neighboring images.

In some embodiments, the data on a stream of desired frames is received from a client module comprising a displaying interface and a desired frame selector and controller. Preferably, the client module further comprises a processing and communication unit. Exemplary displaying interfaces include an high resolution screen showing a panoramic view of at least a portion of a captured dynamic scene having a chosen frame represented as a rectangle superimposed on the displayed scene, an head up display having head pose sensors, a partially reflecting transparent surface for direct real viewing of the scene and for reflecting a graphical

representation of a chosen field of view displayed on a miniature screen, and a monocular or binocular unit. That ocular unit directly views the scene and includes pan and tilt unit, frame size control means, pan and tilt sensors and frame size measuring means for controlling and measuring the size and angles of a desired field of view and means for graphical representation of the desired field of view.

BRIEF DESCRIPTION OF THE DRAWINGS

The subject matter regarded as the invention is particularly pointed out and distinctly claimed in the concluding portion of the specification. The invention, however, both as to system organization and method of operation, together with features and advantages thereof, may best be understood by reference to the following detailed description when read with the accompanied drawings in which:

Fig. 1a illustrates a television production system divided between two locations connected by a low bandwidth communication link.

Fig. 1b is a block diagram of a client module

Fig. 2 is a block diagram of a production suite including director screen and an operators team.

Fig. 3 is a flow chart of a method for video pre-processing by a venue server.

Fig. 4 is a flow chart of a method for a remote production suite.

Fig. 5 is a flow chart of a method for post processing a selected frame by the venue server.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will now be described in terms of specific example embodiments. It is to be understood that the invention is not limited to the example embodiments disclosed. It should also be understood that not every feature of the methods and systems is necessary to implement the invention as claimed in any particular one of the appended claims. Various elements and features of devices are described to fully enable the invention. It should also be understood that throughout this disclosure, where a method is shown or described, the steps of the method may be performed in any order or simultaneously, unless it is clear from the context that one step depends on another being performed first.

Before explaining several embodiments of the invention in detail, it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of the components set forth in the following description or illustrated in the drawings. The invention is capable of other
 5 embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which
 10 this invention belongs. The systems, methods, and examples provided herein are illustrative only and not intended to be limiting.

In the description and claims of the present application, each of the verbs "comprise", "include" and "have", and conjugates thereof, are used to indicate that the object or objects of the verb are not necessarily a complete listing of members,
 15 components, elements or parts of the subject or subjects of the verb.

Definitions

- The terms **field of view**, abbreviated as **FOV**, **desired frame**, and **requested frame** are used throughout this disclosure for describing a desired region in the scene or its image which a person would like to see, broadcast, etc.
- 20 - A client module is a device, system or piece of apparatus used by an operator or a home user to view a panoramic scene or its image and communicate a desired field of view and specific data. In some embodiments the operator of a client module receives the desired field of view for personal viewing. In other embodiments he just specifies a field of view to be broadcast or webcast to others.
- 25 -a virtual cameraman is in few embodiments a human operator located away from a scene whose task is to define a desired frame reflecting a region of interest. He/she is virtual in the sense that that he/she does not operate a real camera and only provides a service that helps a whole system to provide the service provided by a real cameraman located in the venue. In other embodiments a virtual cameraman might be
 30 a software agent or a specific program that provides he mentioned selection of a desired frame.

Referring now to Fig. 1a, it describes a system **100** for low cost television production, which includes a data server **115** in a venue **105** and a production suite **160**. In the following, "in venue" production is described first, followed by a

description of a coordinated production carried on by the venue server 105 and by the remote production suite 160. Then, interactive and immersive virtual reality embodiments for home users is presented. Finally, methods for these embodiments are presented.

5 In venue live production embodiment(Figs. 1a,1b)

As mentioned, it is suggested a method and system for low cost efficient television production. The proposed method is supposed to replace current outdoor productions of general outdoors and especially sports scenes that are expensive in equipment and manpower. The system proposed herein is composed of a fixed camera array 110, a
10 venue data server 115 and a client module 195. It is noted that the client module 195 is optional. The proposed system may be used only for post production applications, and for such applications neither a client module nor a production suite is required during the live capturing of the scene.

The cameras of array 110 are used for video acquisition. If there is more than
15 one camera, the cameras are frame synchronized. Each camera is capturing video images of a certain zone of the scene. A lens is adjusted for each camera to cover the zone. The camera array 110 preferably covers a continuous zone of the venue. Optionally, a method of optimal coverage of the venue composed of lenses with a variety of focal lengths and cameras with different number and size of horizontal and
20 vertical pixels may be used by packing simulation techniques. Typically, neighboring cameras have overlapping coverage zones.

The cameras of array 110 may be CMOS or CCD cameras. Each camera may be composed of a single or of three imagers. Preferably, the cameras are of industrial grade. The data output format of the cameras can be GigE, USB 3.0, Camera Link,
25 HD-SDI or any other format available for such cameras.

The data from all cameras is transferred to the venue server 115, located besides the camera array, in another location in the venue or in a remote location like a broadcast studio or in a "cloud server farm". Data transmission from the camera array module to the data server may be done via optical cable or any other wired or wireless
30 communication system. In a preferred embodiment of the current invention, the cameras have GigE output and a network switch with fiber optical cable used to transfer the data to the data server 115. The cameras data can also be transferred using the worldwide web.

The data server 115 includes the following modules:

-A data acquisition and grabbing module, or pre-processor 120, to grab and initially process the camera data. This initial processing may be composed of cameras photometric and geometrical calibration, video compression process using methods such as JPEG2000 and MPEG2, image transformation and image enhancement processes such as high pass filtering and color correction, and generation of low resolution panoramic views. The photometric calibration process is preferably frequently repeated during the event. Typically, it is repeated every few seconds but in principle it can be done every frame or not done at all during an entire event. Such initial processing may also be done in the camera array 110 or in the client modules 10 195 or in the selected frame filling module 130.

-A stitching and aligning module, or selected frame filling module 130, whose function is to adjust the color and brightness of the cameras involved in the desired frame, to transform the stitching lines from world to camera coordinates, to blend the pixels in the stitching zones and to geometrically align the outputs of the cameras composing the frame using one of the methods known in prior art to calibrate and align images captured by camera array. In another embodiment of the current invention, the stitching and alignment module may be located in a client module 195.

-A data storage module, or recorder 125, whose function is to store at least a portion of the data acquired by the data acquisition module 120. The data storage module 125 is composed of either hard disks, flash memory modules or any other means for massive fast data storage. Module 125 stores the data for the live production, for replays, for post production as well as for other applications. In a preferred embodiment of the current invention, the video data recorded onto the recorder module 125 is raw or just slightly compressed video data wherein no pixels blending, alignment or other pixels processing tasks have been applied. In this preferred embodiment, all pixels blending, alignment and other processing tasks are done in the selected frame filling module 130 and only on pixels belonging to the single desired frame going on air and thus a lot of processing time and resources are saved.

- The selected frame filling module 130, or a field of view(FOV) filling module also receives a requested frame FOV specification from a client. The client may be a human "virtual cameraman", or a computer program doing automatic event production. The FOV filling module 130 retrieves the required data from the data storage module and performs certain pixels operations on the acquired data. The operations of the FOV filling module 130 include geometrical image alignment,

aberrations correction, vignetting correction, alignment, stitching zones blending, “minimum difference” stitching line drawing and a variety of image filter operations. The FOV filling module 130 transmits the matrix of pixels filling the frame as requested by the client. This process is done for each frame request.

5 The client module 195 can be located at the venue 105, remotely in a broadcast studio, remotely at a server farm, or remotely at a user home. The client module 195 is typically composed of a display means 196, a FOV controller 197 and optionally a processing and communicating unit 198, as shown by Fig. 1b.

10 The most common embodiment of the client display 196 is a high resolution LCD screen showing a panoramic view of at least a portion of the captured dynamic scene. The current FOV or desired frame as chosen by the client serving as a “virtual cameraman” is represented, for example, as a graphical rectangle superimposed on the captured scene.

15 Alternative client display 196 is a Head Up Display (HUD) module equipped with head pose measuring sensors wherein the graphical representation of the current FOV is generated by means of a miniature LCD screen and a collimating optics and is reflected from a partially reflecting transparent surface mounted in front of the HUD. The user directly views the scenery through the transparent optical element. The desired FOV graphical representation is projected to infinity and is superimposed on
20 the natural view of the virtual cameraman wearing the HUD. With such an HUD display module the virtual cameraman watches the real scenery and not a captured video.

25 Another optional client display 196 is a mono or binoculars unit mounted on top of a manually driven pan and tilt unit and frame size control means 197, and preferably having also means for pan and tilt angles and frame size measurement. The graphical FOV representation is injected to at least one eyepiece of this telescope (ocular) and is reflected from a partially reflecting beam splitter located in front, behind or within the eyepiece. The FOV graphical representation is superimposed with the operator’s magnified natural view. The graphical FOV representation is
30 generated by means of a miniature LCD screen and a collimating optics module. The advantage of this embodiment with respect to the former embodiment is the fact the virtual cameraman watches a magnified view of the natural scenery. The telescope’s magnification is preferably adjustable.

Alternatively, standard camera handles are used as a frame controller 197 so that the cameraman is using its native operational environment, for defining and presenting the desired FOV or frame.

5 A set of virtual reality goggles may be used as client module 195, which includes a display of a portion of a view seen by the stream of panoramic views, and means for fitting the portion to head movement in similarity to a view seen by a real spectator of the scene having a similar head movement. Preferably, each goggle is fed by a different panoramic view such that a three dimensional scene is sensed by an operator wearing the goggles.

10 Another option for a FOV controller 197 is a software that automatically determines the FOV location and size on the panoramic venue view using image understanding and objects tracking techniques. For example, if the captured event is a football match, the proposed software automatically detects and tracks the players and the ball, automatically recognizes game events like free kick, corner kick etc and
15 determines the appropriate size and location of the instantaneous FOV based on the nature of the current captured play event.

The FOV controller 197 may also be partially automatic and partially manual. In this case, either the automatic software or the human operator may have the override decision authority.

20 Remote live production embodiment(Figs. 1a,1b and 2)

Referring now to Figs. 1a and 2, they illustrate a coordinated production by the venue server 115 and the production suite 160. The production suite 160, including team and equipment, is located in a remote production suite and not in the venue. The suite equipment includes a suite processor 170, a director split screen 165 and several
25 client modules 190-n used by respective operators(virtual cameramen) 180-1,180-2, 180-n. The camera array 110 and the data server 115 are still located in the venue.

A stream of low resolution panoramic images or views is prepared as detailed below and transmitted to the remote production suite 160. The transmission can be done via the world wide web but can also be using any other wired or wireless
30 transmission link 155. Also, a cloud service may be used for transmission. A low resolution representation is used to ensure real time high fidelity transfer of the data over low bandwidth communication links. Thus, in case a broad bandwidth data link is available, a full resolution or just higher resolution data may be transmitted.

A FOV specification process is done manually by a “virtual cameraman” or operators 180-1, 180-2, 180 and 180-n. As an example, the operator 180-1 may use a joystick to move the FOV rectangle across the displayed dynamic scene and a button to setup the FOV size (magnification). More specifically, the suite processor
 5 170 gets the panoramic views and distribute the whole panoramic view or portions thereof to displays used by operators 180-1, 180-2 ad 180-n. The director can define in advance preset fixed frames 225-1, ...225-n capturing fixed regions of interest, for example the bench in a basketball game or the referee stand in tennis. Such a preset frame may be selected by the director as the frame to go on air. An operator 180-1 can
 10 use this or another preset frame as a starting frame, he can then use a joystick to move across the panoramic view (or the portion of the panoramic view allocated to him) and another control to zoom in and out. The operator can also use a starting frame that is a portion of the panoramic view (or of the portion of the panoramic view allocated to him) which he is manually selecting. The operator 180-1 generates a
 15 frame 220-1 displayed for the director choice. As an example, operator 180-1 represents a leading camera “virtual cameraman” and operator 180-2 generating frame 220-2 is a closeups “virtual cameraman”.

Other options for a desired frame to go on air which are displayed for the director’s choice are a moving virtual camera 230-1 performing a pre-defined virtual
 20 motion or tracking an object in the scene, a frame 235-1 determined by an algorithm for automatic detection of a predefined event, and a frame 240-1 selected by replay operator 210-1 associated with an event which occurred in a certain location in the scene previously for a time/location replay. Similarly, camera operator 180-n generates a frame 220-n using the frame controller 197, replay operator 210-n
 25 generates a replay frame 240-n while preset fixed frame 225-n, moving camera 230-n, and automatic algorithm 235-n are predefined by the director 175.

Replay operators 210-1, 210-2 and 210-n use the traditional switches and controls of a sports replay machine or a more specific frame controller 197 allowing location change to get a time/location replay 210-1 or 210-n.

30 Moving the desired frame in pan and tilt directions and changing the zoom may be conducted using client modules like a mouse or a system based on hand gestures recognition. Operator 180-n, for example, uses a set of virtual reality goggles as client module 190-n. Set 190-n includes a display of a portion of a view seen by the stream of panoramic views, means to adjust the visual magnification and means for

fitting the portion to head movement in similarity to a view seen by a real spectator of the scene having a similar head movement. Preferably, each goggle is fed by a different panoramic view such that a three dimensional scene is sensed by an operator wearing the goggles.

5 An event director 175 accepts several frames coming from several operators 180-1, 180-2, 180-n, 210-1, 210, 210-n. These views are displayed for him on a split screen 165 or on plurality of video monitors. Other views constantly displayed on the director's split screen are the predefined preset frames 225-1 and 225-n, the frames 235-1 and 235-n generated by automatic algorithms and icons 230-1 and 230-n
10 representing the predefined moving virtual cameras. The director selects the view he/she wants to go on air at a given time. This process is done continuously throughout the entire production, where usually at a given time only one frame is going on air. The transition between desired frames selections occurs in arbitrary points of time and once an operator output is selected it stays on air until further
15 selection.

Sometimes, transition occurs gradually such that two selections of desired frames are presented simultaneously with the old one fading out while the new one takes over. Other transition effects like wipe are also possible. The director can select the transition he would like to have when the switch to another frame is launched.

20 The director 175 also selects the audio channel he would like to go on air and the audio signal specifications. For example he can choose the commentator 262 or announcer 260 feeds as well as the ambient venue noise as well as the sound volume using his audio controls 265. He can also mix several sound sources together. In addition, the director can select graphical objects to be superimposed on the frame
25 going on air using the graphics controls 255 on his screen. The templates displayed on the screen can also represent animations. In addition a stats operator 250 loads statistical data into the suite processor 170 using the graphics controls 255. The director selects the timing and shape for this data to be presented to the viewers. The director 175 decides which of these audio, graphics, transition and stats data items
30 should accompany the selected stream of frames, and data and instructions regarding these data items are submitted to the venue server 115 with the data on the selected frames.

The participation of the event director 175 is just optional. In case there is a single virtual cameraman 180-1 or an automatic software 235-1 determining the FOV, there is just a single feed and no need for an event director.

The specification of the current FOV to go on air, represented for example by four angular coordinates, is sent back to the venue server 115 by means of the world wide web or any other wired or wireless transmission link 155 or 157. In the venue 105, the selected frame filling module 130 renders the specified FOV with the high resolution cameras data that have been previously recorded onto the recorder 125. In addition, the audio, stats and graphical data items are embedded within the produced stream of high resolution frames.

An output formatting module 135 prepares the stream of frames to the required video standard and transmits it to the TV station studio 145 or directly to the TV viewers using satellite, cable or any other broadcast transmission method. Additional destination with increasing popularity nowadays is a webcast media 140 allowing users to view the streamed presentation over the internet. The signal can also be directed to an individual web client, a home user, directly or via a cloud server, according to his specified frames sequence request. The generated high resolution sequence of frames may also be transmitted back to the production suite using link 157.

Sometimes, a venue cameraman with a shoulder mounted or another moving real camera is capturing the scene in addition to the fixed camera array 110. The stream of frames produced by the moving camera is transmitted to the production suite 160, possibly after some reduction of its data rate and compression, while raw images are recorded on recorder 125 alongside with the fixed cameras footage. The moving real camera transmitted video stream is displayed to the director together with the other "virtual" frames. In case the director 175 is selecting this stream of images to go on air, an appropriate instruction is transmitted to the venue server 115, and the stream of previously recorded high resolution images are broadcasted, possibly after some processing.

In some embodiments, several arrays 110 of video cameras are capturing the scene from several venue locations. In these embodiments or in other embodiments the streams of panoramic views are transmitted to one or more production suites 160.

Virtual reality embodiment for home users (Figs. 1a, 1b and 2)

In this embodiment of the current invention the client is a home user using a virtual reality headset 195 with wide FOV and built in sensors to measure the instantaneous head position. One example of such a headset is manufactured by Oculus VR of Irvine, California, having a FOV of 110 degrees.

5 The virtual reality headset, originally designed for gamers, is absolutely opaque to the outside world and offers the users a totally immersive experience. This experience can also be stereoscopic when different left and right eye views are projected correspondingly in front of the left and right eyes of the headset user. Typically such commercial headsets have also built in speakers to enhance the visual
10 experience with an immersive audio environment.

 In this embodiment, a sequence of panoramic frames is continuously sent from the venue server 115 to the client processor 198 using the world wide web preferably via a cloud service or using any other transmission link. The user can dynamically determine the zoom level he wants using a zoom control. When turning his head the
15 head position sensors read the new pose. The head position parameters are transmitted to the processor 198 and the new FOV is thus generated and immediately presented to the user.

 It is also possible to generate a stereoscopic representation of the captured event. One preferred way to do that is to use the Stergen's system as described in US
20 patent Publication 2012/0013711 to Tamir et al, to convert in real time the 2D representation to 3D stereoscopic one. There are other ways to generate a stereoscopic visual representation including using at least two camera arrays with the appropriate lateral separation and convergence, using a single array with at least partially overlapping zones or using a generic 2D to 3D convertor like IF-2D3D1
25 manufactured by JVC.

 Alongside the pixels data, audio data from a microphone located in the venue, preferably two microphones, maybe transmitted to the processor 198 alongside the visual data and synchronized with the video stream. The audio/visual data is transmitted to the projection devices built into the headset 195 worn by the home user
30 and into the built in speakers. In case there are no such built in speakers, the data can be routed into an external speaker located nearby the home user.

 In case there is a 3D stereoscopic representation of the event, two streams of data are separately directed into the left and right eyes projection means in the headset 195. The data streams going to the left and right eyes, as well as the audio data, are

slaved to the headset's built in sensors measuring the head position. When a head position parameter (pan, tilt and roll angles, three translation axes) changes, the new head position parameters are sent to the processor 198. A corresponding new FOV is then calculated using the new parameters of the headset position.

5

Methods for producing a stream of desired image frames (Figs. 3-5)

It is provided a method for producing a stream of desired image frames derived from a panoramic capturing of a scene by an array of video cameras 110. The method is presented in Figs. 3-5 which includes a flow chart of a method 300 for pre-processing
10 in a venue server 115, a coordinated method 400 carried by the venue server 115 proximate to the scene and at a production suite 160 located remotely from the venue 105, and a method 500 for post processing by the venue server 115.

Method 300, carried by the venue server 115, includes a step 310 of preliminary calibrating of the video cameras of the array 110 based on mechanical
15 design and lens specifications, a step 315 of accurately calibrating of the video cameras based on the scene to get calibration parameters, and a step 320 of capturing an event by the rig cameras of array 110 to get a stream of raw video images.

Method 300 includes two steps of pre-processing the raw images which are conducted continuously in parallel. The first step is recording a stream of raw or just
20 slightly compressed video images captured simultaneously by the video cameras. The second step is converting the stream of video images at a first data rate to a stream of panoramic views at a second data rate. The second data rate is less than one third of the first data rate.

Within the framework of recording, method 300 includes a step 325 of
25 decomposing frames captured by the video cameras into pixel tiles, a step 330 of associating the pixel tiles with time codes, and a step 335 of recording the pixel tiles in recorder 125.

Within the framework of converting the raw images, method 300 includes a step 360 of low quality blending and aligning of neighboring frames (generated by
30 neighboring cameras) captured by the array 110 of video cameras, a step 365 of generating low resolution panoramic view from the blended and aligned frames, and a step 370 of compressing and encoding the panoramic views for transmitting to the remote production suite 160. Step 370 is followed by step 375 of transmitting the

stream of panoramic views to the remote production suite 160 through a low bandwidth link 155.

Coordinated method 400 includes a step 405 of determining several preset fixed frames and virtual camera preprogrammed motions for presenting respective fixed regions of interest constituting portions of the panoramic views and desired camera motions across the scene. A stream of panoramic views is continuously transmitted from the venue server 115 as represented in step 410. Method 400 further includes a step 415 of generating and distributing to operators partial frames of the panoramic views in accordance with respective regions of interest. For example, the leading “virtual cameraman” receives the panoramic view of the whole scene while “virtual cameraman” 2, supposed to produce a tighter image, receives the frame of the leading cameraman as a background view to move across. Method 400 also includes a step 420 of the operators moving across and zooming in and out the panoramic frames to get actual frames, and a step 425 of presenting to a director 175 the actual “virtual cameramen” and replay, preset, moving cameras and “automatic algorithm” frames.

Method 400 also includes a step 430 of the director 175 selecting an actual frame to go on air, and a step 435 of the director 175 selecting a graphics template, an audio source, transition and stats data to go on air.

In addition, method 400 includes a step 440 of transmitting data on the stream of desired frames to the venue server 115 together with instructions regarding the graphics, audio, transition and stats data using a narrow bandwidth link 155, a step 445 of generating well stitched high resolution content for the selected frames, a step 450 of rendering the graphics template, switching to the audio source and adjusting other data instructions, and a step 455 of broadcasting or webcasting the graphics/audio embedded video stream. The high resolution frames are characterized by non perceptible or seamless boundaries between neighboring images derived from respective neighboring raw video images.

Referring now to Fig. 5, it presents a flow chart of a method 500 for post processing and rendering the desired sequence of frames in the venue server 115. The method 500 includes a step 505 of receiving from the production suite data on the stream of desired frames, graphics template, audio source and specifications and other data, a step 510 of retrieving recorded pixel tiles associated with a selected frame, a step 515 of transforming stitching lines from world coordinates to camera coordinates and specifying the lines in the selected frame or close thereof, and a step 520 of

geometrically aligning pixels belonging to neighboring video cameras using the calibration parameters. The method 500 also includes blending pixel pairs in stitching zones belonging to neighboring video cameras in accordance with a distance from a stitching line, a step 530 of correcting color intensities of pixels belonging to different
5 video cameras, and a step 535 of adjusting hue, saturation, intensity and gamma value of the entire frame.

Method 500 further includes a step 540 of encoding the stream of high resolution frames into a required output video format, and a step 455 of broadcasting or webcasting the formatted stream of high resolution frames.

10 Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. In particular, the present invention is not limited in any
15 way by the examples described.

CLAIMS:

1. A method for producing, by a coordinated operation conducted at a venue proximate to a scene and at a production suite located remotely from the venue, a stream of desired image frames derived from a panoramic capturing of the scene by an array of video cameras, the method comprising:
 - (a) a venue server recording a stream of raw video images associated with images captured simultaneously by the video cameras;
 - (b) converting said stream of raw video images at a first data rate to a stream of panoramic views at a second data rate, said second data rate being less than one third of said first data rate;
 - (b) transmitting said stream of panoramic views to at least one remote production suite;
 - (c) in the remote production suite, selecting a stream of desired frames including respective regions of interest, the selected frames being portions of respective panoramic views;
 - (d) transmitting data on said stream of desired frames to said venue server; and
 - (e) said venue server producing out of the recorded stream of raw video images, a stream of high resolution frames in accordance with said stream of desired frames, said high resolution frames being characterized by non perceptible boundaries between neighboring images derived from respective neighboring raw video images.
2. The method of claim 1 wherein the method includes determining a plurality of preset fixed frames for presenting a respective plurality of fixed regions of interest.
3. The method of claim 1 wherein the method includes selecting a frame from a group of frames consisting of:
 - (i) a preset fixed frame determined in advance for capturing a fixed region of interest;
 - (ii) a virtual moving camera following an object in the scene;
 - (iii) a virtual moving camera performing a pre-defined motion;
 - (iv) a frame determined by an algorithm for automatic detection and inclusion of a predefined event; and
 - (v) a frame associated with an event which occurred in a certain location in the scene previously.

4. The method of claim 1 wherein the method includes transmitting to the venue server for appending to a desired frame instructions regarding at least one data item of a group of data items consisting of:
 - (a) a graphical item;
 - 5 (b) a desired artistic transition between frames;
 - (b) an audio item; and
 - (c) statistical data regarding identified events occurring in the scene.
5. The method of claim 1 wherein two or more arrays of video cameras are capturing the scene from two or more venue locations, and respective streams of panoramic
 10 views are transmitted to one or more production suites.
6. The method of claim 1 wherein at least one moving video camera is used in conjunction with the array of video cameras to capture the event.
7. The method of claim 1 wherein at least one of said transmitting to said production suite and said transmitting to said venue server is done using an internet link.
- 15 8. The method of claim 7 wherein at least one of said transmitting to said production suite and said transmitting to said venue server is done using a cloud service.
9. The method of claim 1 wherein said scene is a sports scene.
10. The method of claim 1 wherein a client module is selecting said stream of desired frames, said client module comprising a displaying interface and a desired frame
 20 selector and controller.
11. The method of claim 10 wherein said client module serves a home user.
12. The method of claim 10 wherein said client module further comprises a processing and communication unit.
13. The method of claim 10 wherein said displaying interface is selected from a
 25 group of interfaces consisting of:
 - (i) a screen showing a panoramic view of at least a portion of a dynamically captured scene having a chosen field of view as a rectangle superimposed on the displayed scene; and
 - (ii) a monocular or binocular unit directly viewing at least a portion of the
 30 dynamic scene, comprising:
 - (A) a pan and tilt unit, means for zoom control and means for measuring the pan and tilt angles and the frame size for controlling a desired field of view; and
 - (B) means for graphical representation of said desired field of view.

14. The method of claim 10 wherein:
- (A) a cloud server receives a sequence of panoramic views from said venue server; and
 - (B) said cloud server streams said sequence of panoramic views to said client module using the world wide web.
15. The method of claim 10 wherein said displaying interface is a set of virtual reality goggles comprising:
- (A) a display of at least a portion of a view seen by said stream of panoramic views; and
 - (B) means for fitting said at least a portion to head movement in similarity to a view seen by a real spectator of the scene having a similar head movement.
16. The method of claim 15 wherein each goggle of said set of virtual reality goggles is fed by a different panoramic view such that a three dimensional scene is sensed by an operator wearing the goggles.
17. The method of claim 1 wherein the method further includes a step selected from a group of steps consisting of:
- (i) preliminary calibrating the video cameras of the array based on mechanical design and lens specifications; and
 - (ii) accurately calibrating of said video cameras based on said scene to get calibration parameters.
18. The method of claim 1 wherein the method includes steps conducted at the venue server, the steps include:
- (i) decomposing frames captured by the video cameras into pixel tiles, said tiles are rectangular arrays of pixels whose size is smaller than a single camera frame size;
 - (ii) associating said pixel tiles with time codes; and
 - (iii) recording said pixel tiles.
19. The method of claim 18 wherein the method includes at least one action selected from a group of actions consisting of:
- (i) retrieving recorded pixel tiles associated with a selected frame;
 - (ii) transforming stitching lines from world coordinates to camera coordinates and specifying the lines in the selected frame or close thereof;
 - (iii) geometrically aligning pixels belonging to neighboring video cameras using said calibration parameters;

- (iv) blending pixel pairs in stitching zones belonging to neighboring video cameras in accordance with a distance from a stitching line;
 - (v) correcting color intensities of pixels belonging to different video cameras; and
 - (vi) adjusting hue, saturation, intensity and gamma value of the entire frame.
- 5 20. The method of claim 1 wherein the method includes at least one step conducted at the venue server and selected from a group of steps consisting of:
- (i) low quality blending and aligning of neighboring frames captured by said array of video cameras;
 - (ii) generating of low resolution panoramic view from the blended and aligned
 - 10 frames; and
 - (iii) compressing and encoding the panoramic views for transmitting to the remote production suite.
21. The method of claim 1 wherein the method includes at least one step conducted at the remote production suite and selected from a group of steps consisting of :
- 15 (i) generating and distributing to operators partial frames of the panoramic view in accordance with respective region of interest;
 - (ii) said operators moving across and zooming in and out said partial frames to get actual frames;
 - (iii) presenting to a director said actual frames; and
 - 20 (iv) said director selecting an actual frame to go on air.
22. The method of claim 1 wherein the method includes:
- (i) encoding said stream of high resolution frames into a required output video format; and
 - (ii) broadcasting or webcasting the formatted stream of high resolution frames.
- 25 23. A method for a production suite located remotely from a venue proximate to a scene, the production suite operating in a coordinated way with a venue server in the venue for producing a stream of desired image frames derived from a panoramic capturing of the scene by an array of video cameras, the production suite method comprising:
- 30 (a) receiving from the venue server a stream of low-resolution panoramic views derived from a stream of raw video images captured simultaneously by the video cameras;
 - (b) selecting a stream of desired frames including respective regions of interest, the selected frames being portions of said low resolution panoramic views; and

(c) transmitting data on said stream of desired frames to said venue server.

24. The method of claim 23 wherein the method includes at least one step selected from a group of steps consisting of:

- (i) generating and distributing to operators in the production suite partial frames
5 of the panoramic view in accordance with respective region of interest;
- (ii) said operators moving across and zooming in and out said partial frames to get actual frames;
- (iii) presenting to a director said actual frames; and
- (iv) said director selecting an actual frame to go on air.

10 25. A method for a venue server producing, by a coordinated operation with a production suite, a stream of desired image frames derived from a panoramic capturing of a scene by an array of video cameras, the method comprising:

- (a) recording a stream of raw video images captured simultaneously by the video cameras;
- 15 (b) converting said stream of raw video images to a stream of panoramic views at a reduced data rate, the data rate reduction being by a factor of at least three;
- (c) transmitting said stream of panoramic views to the production suite for selecting thereof a stream of desired frames including respective regions of interest, the selected frames being portions of the panoramic views;
- 20 (d) receiving from said production suite data on said stream of desired frames; and
- (e) producing out of the recorded stream of raw video images, a stream of high resolution frames in accordance with said stream of desired frames, said high resolution frames being characterized by non perceptible boundaries between neighboring images derived from respective neighboring raw video images.

25 26. The method of claim 25 wherein the method includes:

- (i) decomposing frames captured by the video cameras into pixel tiles;
- (ii) associating said pixel tiles with time codes; and
- (iii) recording said pixel tiles.

30 27. The method of claim 26 wherein the method includes at least one action selected from a group of actions consisting of:

- (i) retrieving recorded pixel tiles associated with a selected frame;
- (ii) transforming stitching lines from world coordinates to camera coordinates and specifying the lines in the selected frame or close thereof;

- (iii) geometrically aligning pixels belonging to neighboring video cameras using said calibration parameters;
- (iv) blending pixel pairs in stitching zones belonging to neighboring video cameras in accordance with a distance from a stitching line;
- 5 (v) correcting color intensities of pixels belonging to different video cameras; and
- (vi) adjusting hue, saturation, intensity and gamma value of the entire frame.

28. A method for a venue server producing a stream of desired image frames derived from a panoramic capturing of a scene by an array of video cameras, the method comprising:

- (a) recording a stream of video images associated with a stream of raw video images captured simultaneously by the video cameras;
- (b) receiving data on a stream of desired frames including respective regions of interest in the panoramic capturing of the scene; and
- 15 (c) producing out of the previously recorded stream of video images, a stream of high resolution frames in accordance with said stream of desired frames, producing a frame of said stream of high resolution frames including actions restricted to only previously recorded images associated with said frame, the restricted actions including:
 - 20 (i) retrieving said only previously recorded images associated with said frame from the previously recorded video images; and
 - (ii) processing only the associated images to get high resolution frame having non perceptible boundaries between neighboring images.

25 29. The method of claim 28 wherein said data on a stream of desired frames is received from a client module comprising a displaying interface and a desired frame controller.

30 30. The method of claim 29 wherein said client module further comprises a processing and communication unit.

31. The method of claim 29 wherein said displaying interface is selected from a group of interfaces consisting of:

5 (i) an high resolution screen showing a panoramic view of at least a portion of a captured dynamic scene having a chosen frame as a rectangle superimposed on the displayed scene;

10 (ii) an head up display having a partially reflecting transparent surface for direct real viewing of the scene and for reflecting a graphical representation of a chosen frame displayed on a miniature screen, and having means to measure the head pose;

(iii) an ocular unit directly viewing the scene, comprising:

(A) a pan and tilt unit and zoom control means and pan and tilt angles and frame size measuring devices for controlling a desired frame; and

(B) means for graphical representation of said desired frame;

15 and

(iv) a set of virtual reality goggles, comprising:

- a display of a portion of the stream of panoramic views; and

- means for fitting said portion to head movement in similarity to a view seen by a real spectator of the scene having a similar head movement.

20

25

1/5

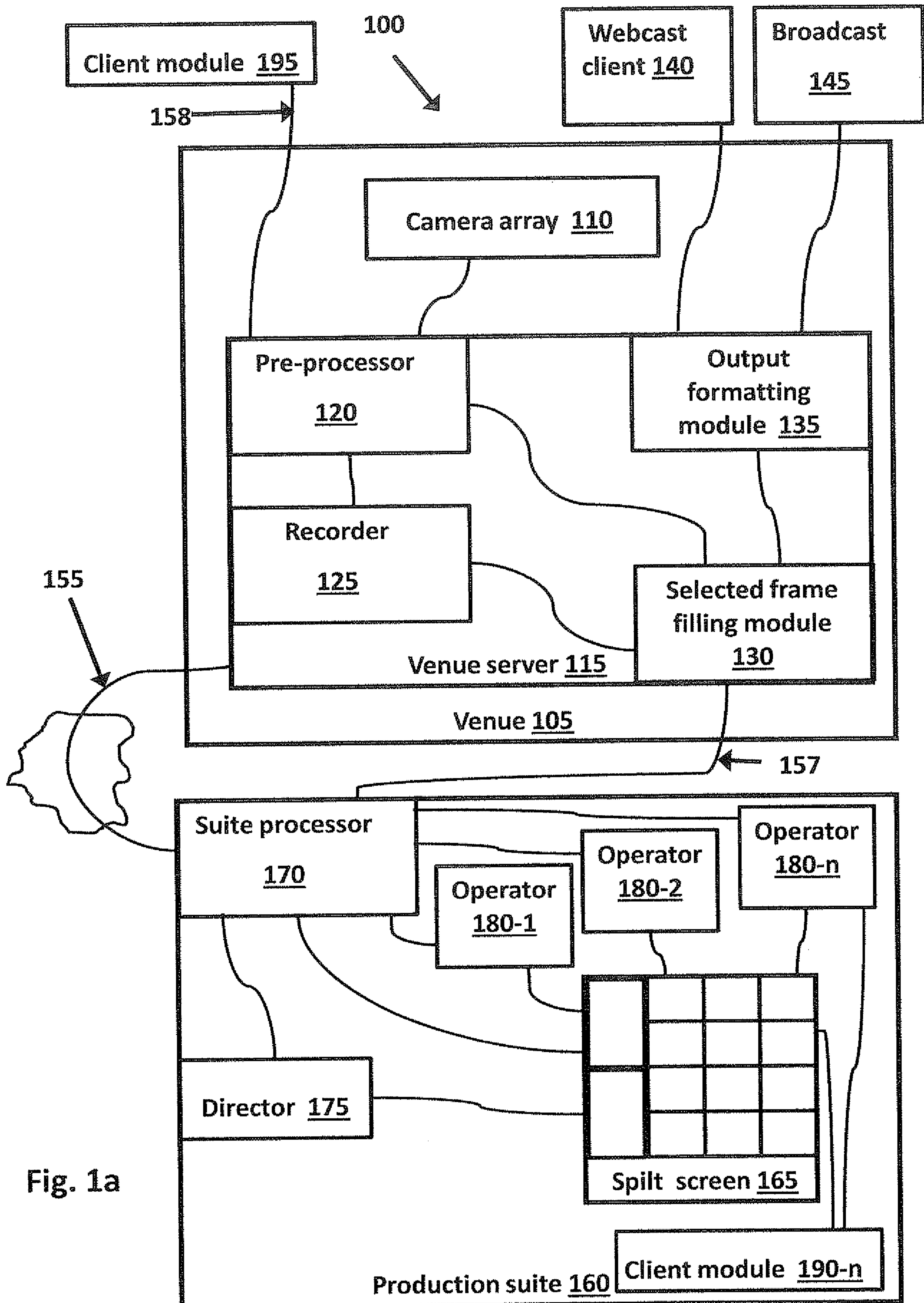


Fig. 1a

2/5

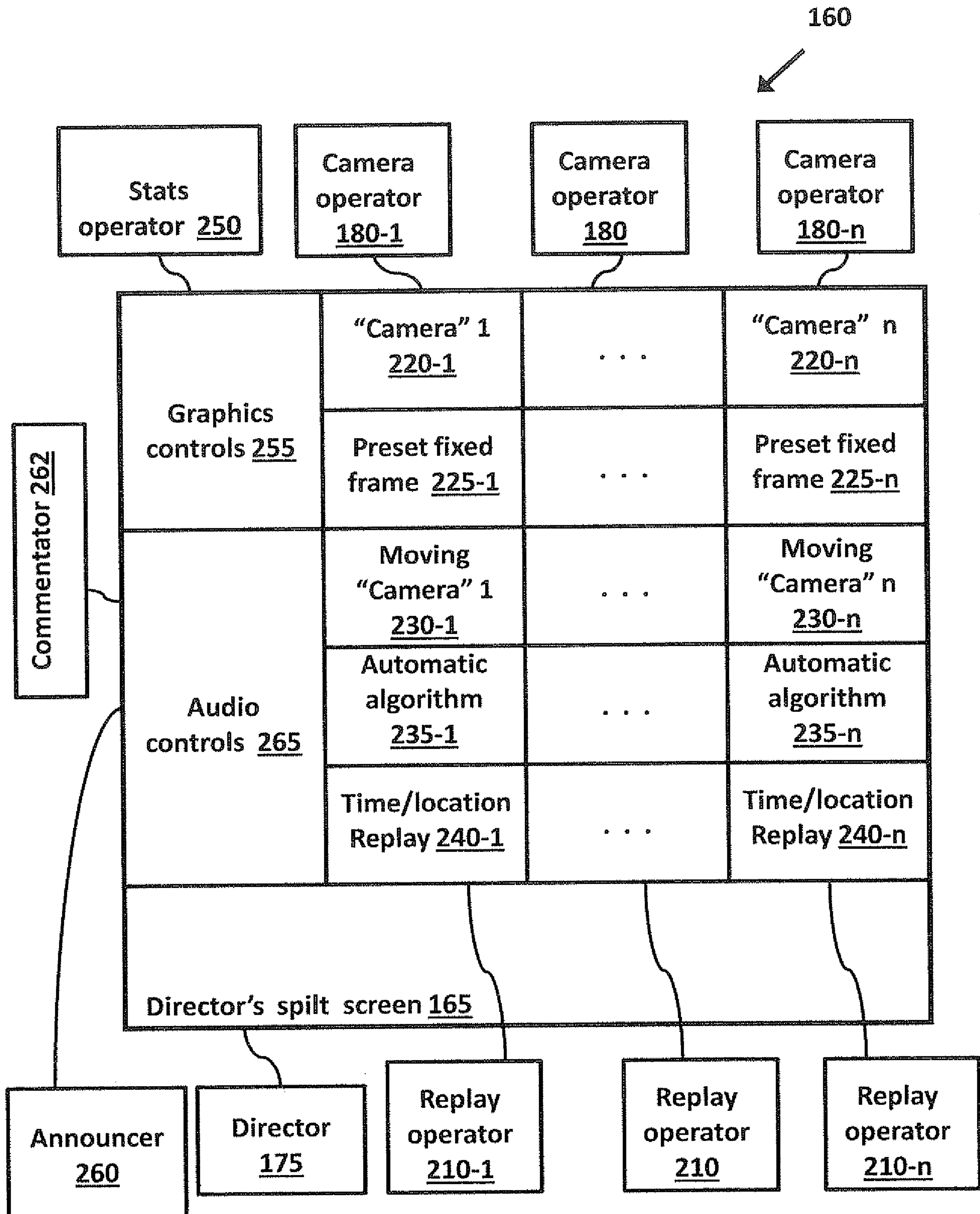


Fig. 2

3/5

Fig. 1b

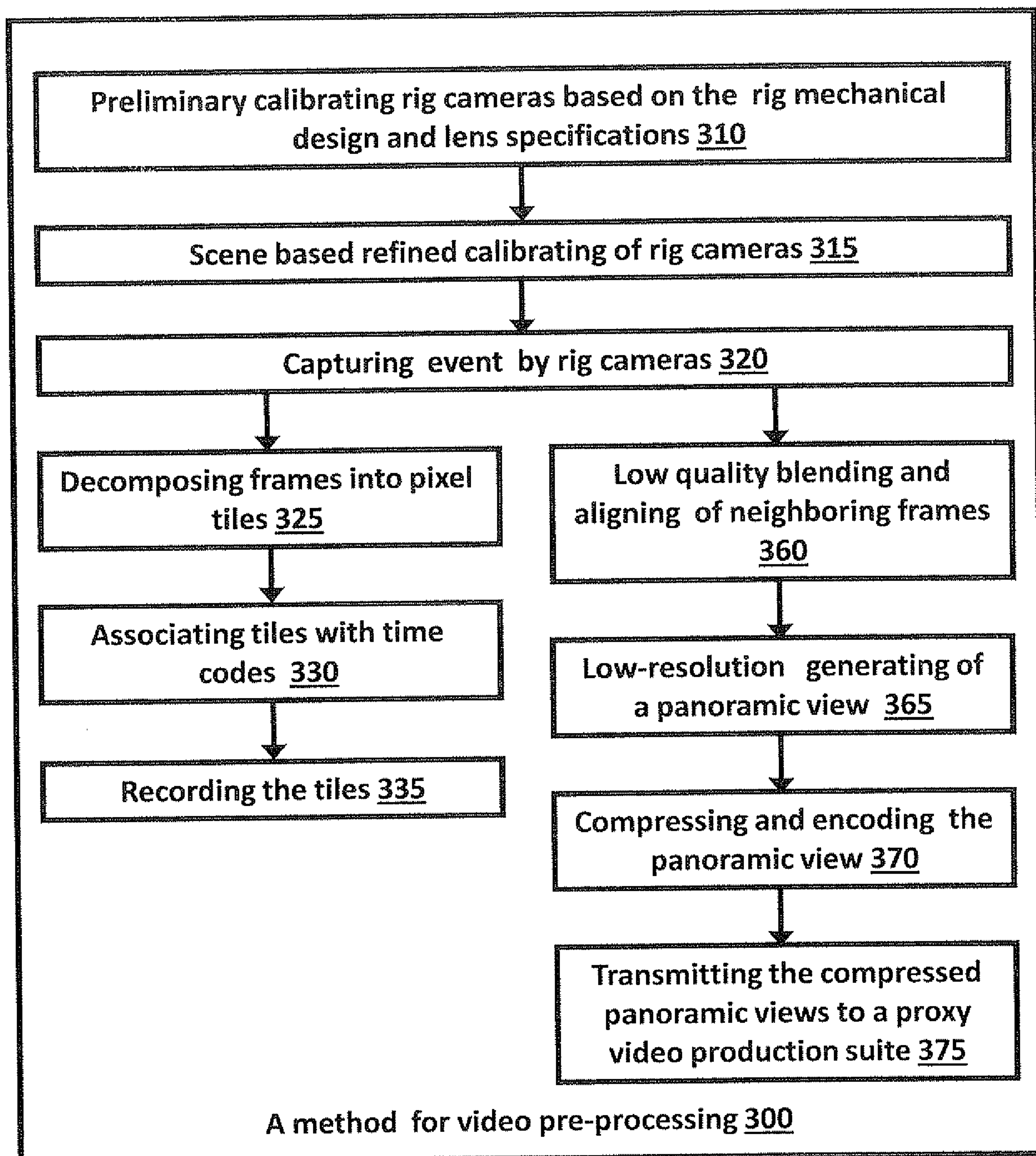
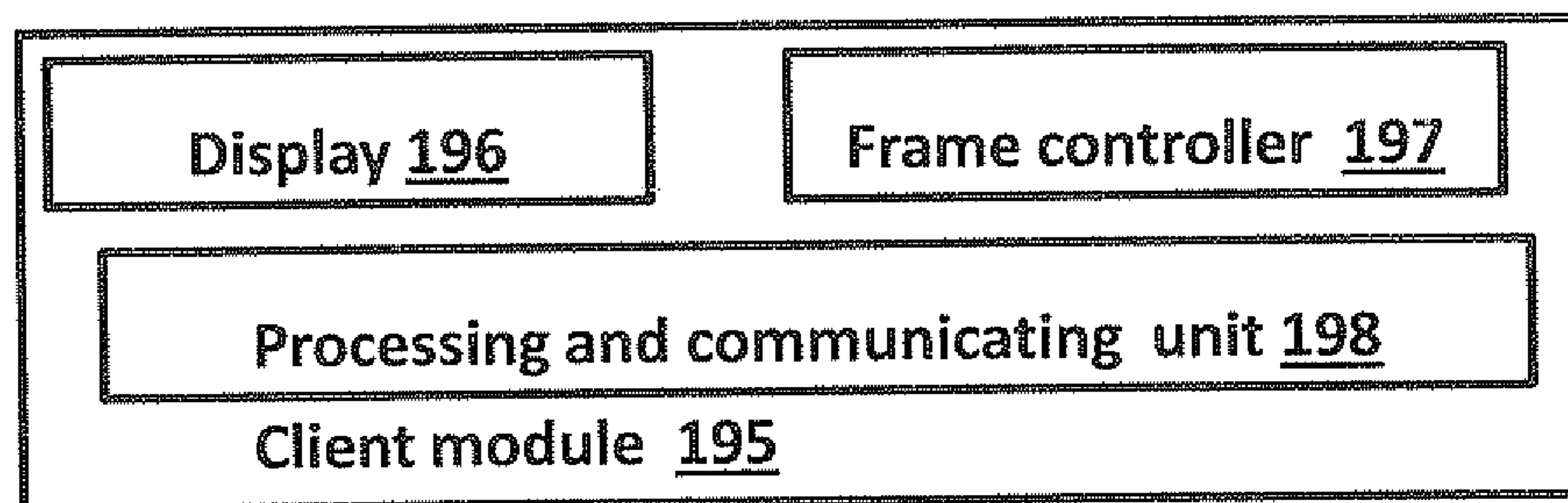


Fig. 3

4/5

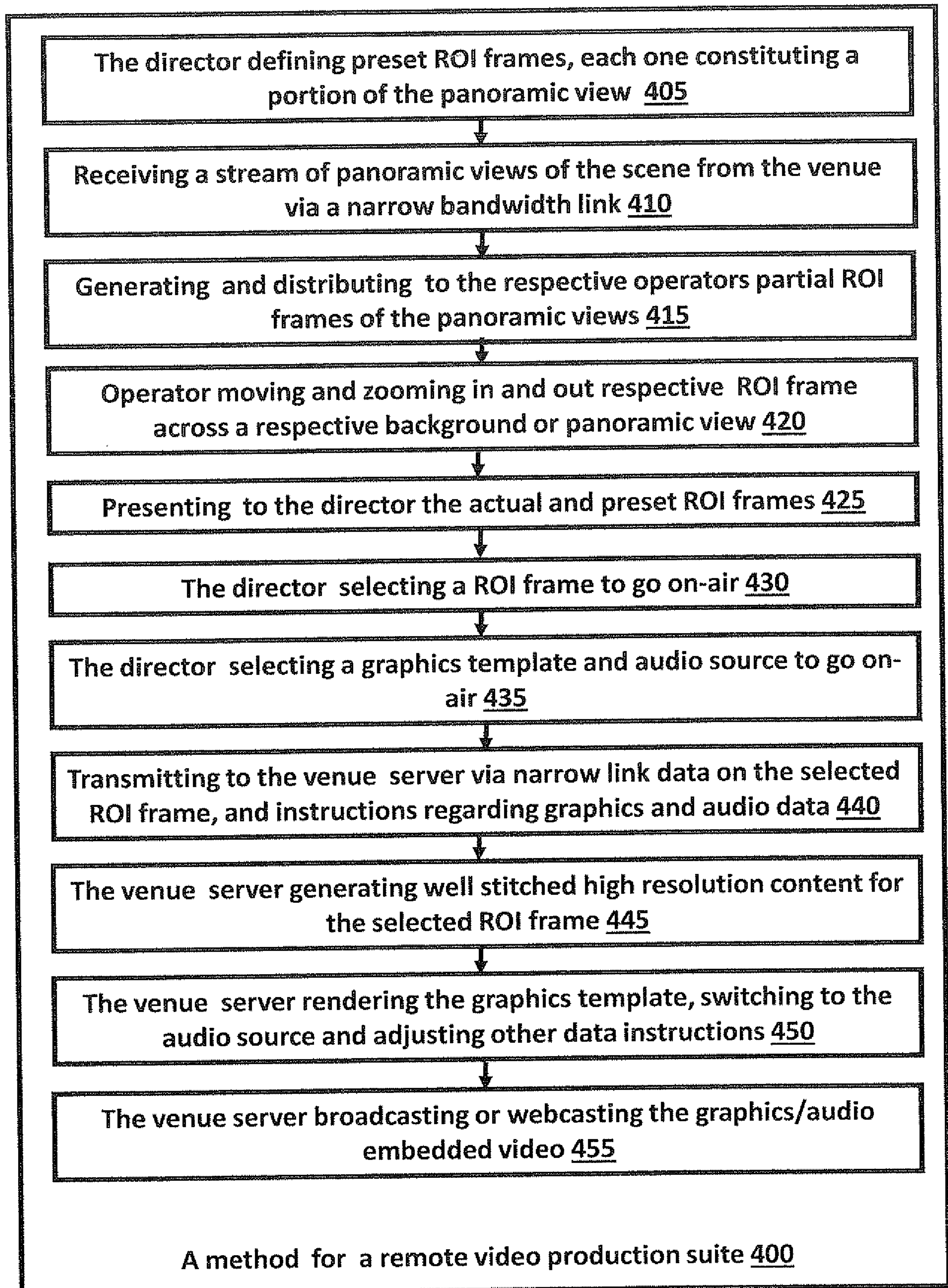


Fig. 4

5/5

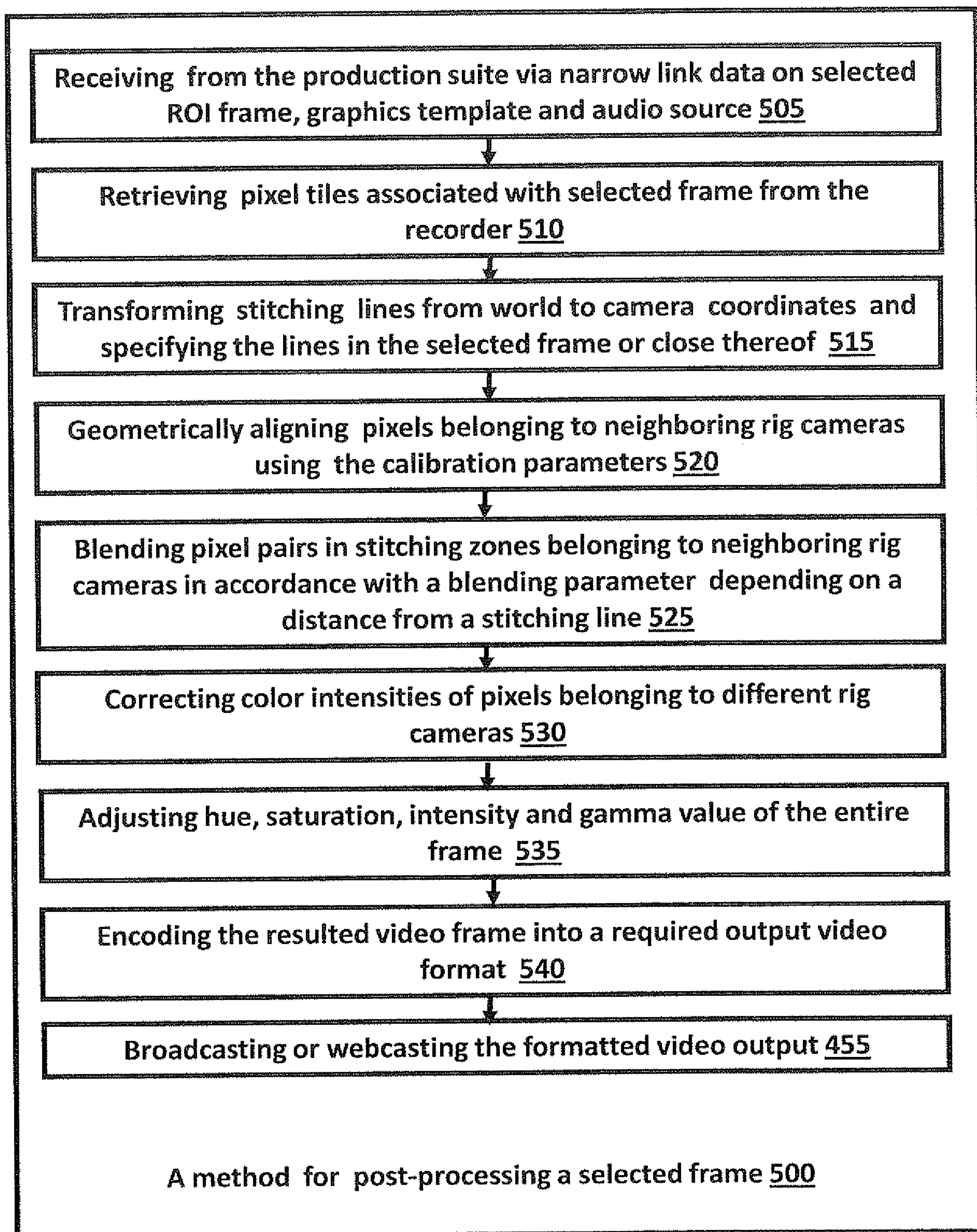


Fig. 5

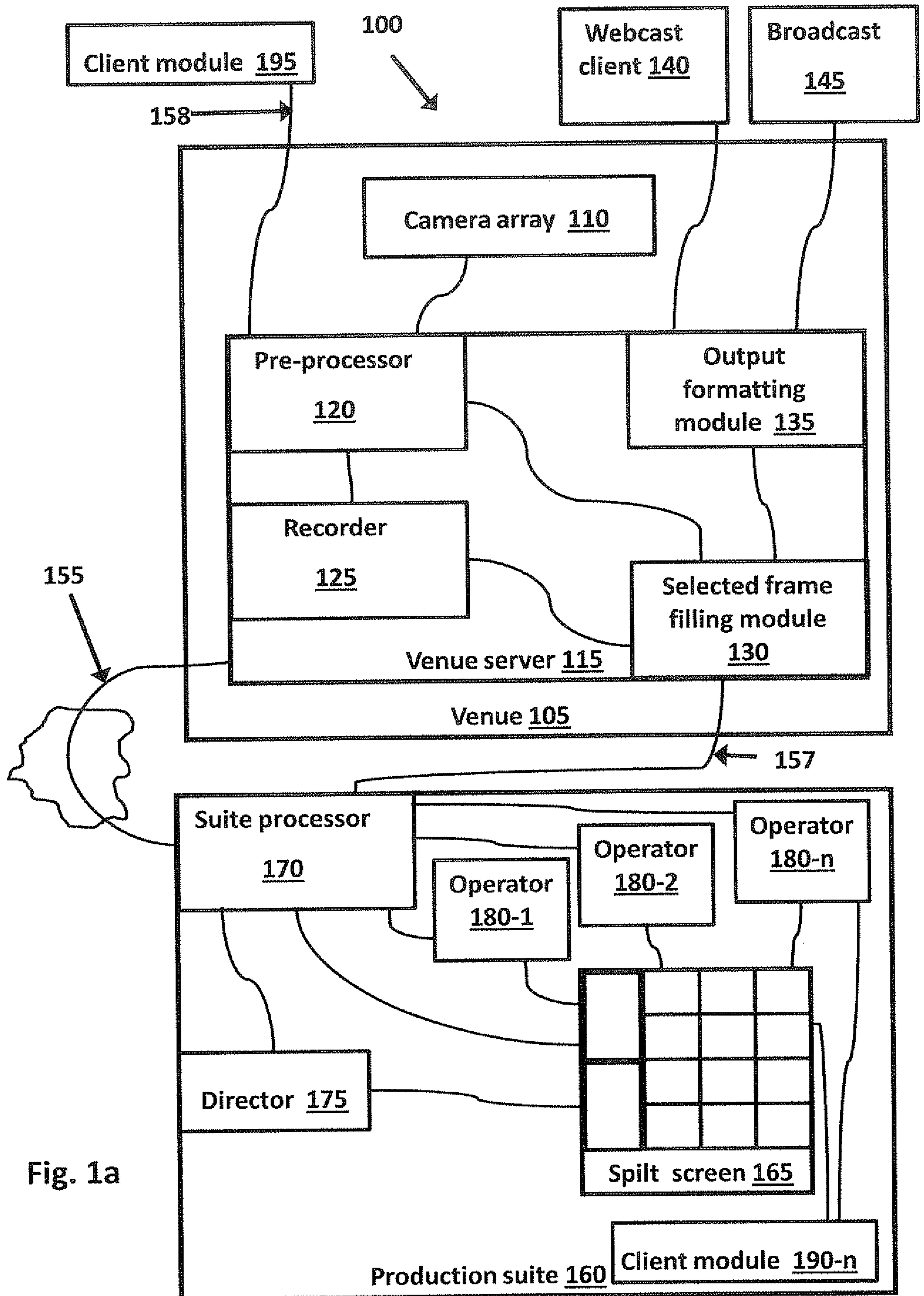


Fig. 1a