SYSTEMS, METHODS, AND COMPUTER MEDIA FOR GENERATING AN AVATAR REFLECTING A PLAYER'S CURRENT APPEARANCE

Abstract

Systems, methods, and computer media for generating an avatar reflecting a player’s current appearance. Data describing the player’s current appearance is received. The data includes a visible spectrum image of the player, a depth image including both the player and a current background, and skeletal data for the player. The skeletal data indicates an outline of the player’s skeleton. Based at least in part on the received data, one or more of the following are captured: a facial appearance of the player; a hair appearance of the player; a clothing appearance of the player; and a skin color of the player. A 3D avatar resembling the player is generated by combining the captured facial appearance, hair appearance, clothing appearance, and/or skin color with predetermined avatar features.
FIG. 1

MEMORY 112

PROCESSOR(S) 114

PRESENTATION COMPONENT(S) 116

I/O PORT(S) 118

I/O COMPONENTS 120

POWER SUPPLY 122
500 RECEIVE DATA DESCRIBING PLAYER'S CURRENT APPEARANCE

502 CAPTURE FACIAL APPEARANCE

504 CAPTURE HAIR APPEARANCE

506 CAPTURE CLOTHING APPEARANCE

508 DETECT SLEEVES, PANT LEGS, OR HEMS

510 CAPTURE SKIN COLOR

512 GENERATE 3D AVATAR RESEMBLING PLAYER

FIG. 5
RECEIVE DATA DESCRIBING PLAYER'S CURRENT APPEARANCE

VISIBLE SPECTRUM IMAGE

DEPTH IMAGE

SKELETAL DATA

CAPTURE FACIAL APPEARANCE OF PLAYER BASED ON RECEIVED DATA

IDENTIFY PLAYER'S FACE

IDENTIFY FACIAL FEATURES

WARP IMAGE PORTION INTO TEXTURE MAP

CAPTURE HAIR APPEARANCE OF PLAYER BASED ON RECEIVED DATA

IDENTIFY EXTENT OF PLAYER'S HAIR

MATCH EXTENT TO HAIRSTYLE TEMPLATE

IDENTIFY HAIR COLOR

CAPTURE CLOTHING APPEARANCE OF PLAYER BASED ON RECEIVED DATA

DETERMINE OUTER CLOTHING BOUNDARY USING EDGE DETECTION

DETERMINE INNER CLOTHING BOUNDARY BASED ON COLOR GRADIENT

IDENTIFY CLOTHING

CAPTURE SKIN COLOR

SAMPLE COLOR OF AREAS IN PLAYER'S FACE

BLEND SAMPLED COLORS TO DETERMINE SKIN COLOR

GENERATE 3D AVATAR RESEMBLING PLAYER

FIG. 7
GENERATION OF AVATAR REFLECTING PLAYER APPEARANCE

CROSS-REFERENCE TO RELATED APPLICATIONS


BACKGROUND

[0002] Video games have advanced substantially in recent years. Improvements in processing power, image processing, and graphics, among other areas, have enabled games of increasing conceptual and graphical complexity. Conventionally, players selected a default character or selected one of a group of simple, predefined characters. As video games have evolved, the use of player avatars rather than simple characters has become increasingly common. Some games allow a player to select among a group of predefined avatars or customize various features of an avatar. Players can attempt to create an avatar resembling themselves by manually customizing features such as hair color, gender, height, glasses, etc. Such manual customization is time consuming and still does not result in an avatar that actually resembles the appearance of the player.

SUMMARY

[0003] Embodiments of the present invention relate to systems, methods, and computer media for generating an avatar reflecting a player’s current appearance. Data describing the player’s current appearance is received. The data includes a visible spectrum image of the player, a depth image including both the player and a current background, and skeletal data for the player. The skeletal data indicates an outline of the player’s skeleton. Based at least in part on the received data, one or more of the following are captured: a facial appearance of the player, a hair appearance of the player, a clothing appearance of the player, and a skin color of the player. A shirt sleeve, pants or shorts leg, or skirt or dress hem may also be detected. A 3D avatar resembling the player is generated by combining the captured facial appearance, hair appearance, clothing appearance, skin color, and/or detected shirt sleeve, pants or shorts leg, or skirt or dress hem with predetermined avatar features.

[0004] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] The present invention is described in detail below with reference to the attached drawing figures, wherein:

[0006] FIG. 1 is a block diagram of an exemplary computing environment suitable for use in implementing embodiments of the present invention;

[0007] FIG. 2 is a perspective view of a visible spectrum image of a player’s head in accordance with embodiments of the present invention;

[0008] FIG. 3 is a perspective view of a visible spectrum image of a player in accordance with embodiments of the present invention;

[0009] FIG. 4 is a perspective view of a 3D avatar resembling the player in FIGS. 2 and 3 in accordance with embodiments of the present invention;

[0010] FIG. 5 is a flow chart of an exemplary method for generating an avatar reflecting a player’s current appearance in accordance with an embodiment of the present invention;

[0011] FIG. 6 is an exemplary avatar generation system in accordance with an embodiment of the present invention;

[0012] FIG. 7 is a flow chart of an exemplary method for generating an avatar reflecting a player’s current appearance in accordance with an embodiment of the present invention in which sub-steps are shown for facial appearance capture, hair appearance capture, clothing capture, and skin color capture;

[0013] FIG. 8 is a perspective view of a visible spectrum image of a player in accordance with embodiments of the present invention;

[0014] FIG. 9 is a perspective view of a visible spectrum image of a player’s head in accordance with embodiments of the present invention;

[0015] FIG. 10 is a partial perspective view of a visible spectrum image of a player’s head in accordance with embodiments of the present invention, the image showing the face and other portions of the head;

[0016] FIG. 11 the image of FIG. 10 shown with alignment points in accordance with embodiments of the present invention;

[0017] FIG. 12 is a template texture map with destination points in accordance with embodiments of the present invention;

[0018] FIG. 13 is a face texture map resulting from warping the image in FIG. 11 by matching the alignment points to the destination points in FIG. 12;

[0019] FIG. 14 is a 3D avatar reflecting the current appearance of the player in the image in FIGS. 8-11 and 13;

[0020] FIG. 15 is a perspective view of a visible spectrum image of a player in accordance with embodiments of the present invention; and

[0021] FIG. 16 is a depth image of the player shown in FIG. 15 in which a simplified skeleton has been identified.

DETAILED DESCRIPTION

[0022] Embodiments of the present invention are described with specificity herein to meet statutory requirements. However, the description itself is not intended to limit the scope of this patent. Rather, the inventors have contemplated that the claimed subject matter might also be embodied in other ways, to include different steps or combinations of steps similar to the ones described in this document, in conjunction with other present or future technologies. Moreover, although the terms “step” and/or “block” or “module” etc. might be used herein to connote different components of methods or systems employed, the terms should not be interpreted as implying any particular order among or between various steps herein disclosed unless and except when the order of individual steps is explicitly described.

[0023] Embodiments of the present invention relate to systems, methods, and computer media for generating an avatar reflecting a player’s current appearance. In accordance with
embodiments of the present invention, an avatar is generated from data describing the player’s current appearance. Various aspects of the player’s appearance are captured based on the data and combined with predetermined avatar features to generate a 3D avatar resembling the player.

[0024] As discussed above, players can attempt to create an avatar resembling themselves by manually customizing certain features, but the resulting avatar still does not actually reflect the appearance of the player.

[0025] Various sensors can be used, however, to gather data describing a player’s current appearance. These sensors provide visible spectrum images, depth images, and skeletal data, among other data, that contain detailed information describing the player’s appearance and can be used to generate an avatar that accurately reflects the player’s appearance. The combination of visible spectrum data, depth data, and skeletal data allows capture of detailed aspects of a player’s appearance not previously possible. Additionally, an avatar resembling the player can be automatically generated without player selection of various features after sensor data is acquired.

[0026] In one embodiment of the present invention, data describing the player’s current appearance is received. The data includes: a visible spectrum image of the player, a depth image including both the player and a current background, and skeletal data for the player. The skeletal data indicates an outline of the player’s skeleton. A facial appearance of the player is captured based at least in part on the received data. The facial appearance is captured by identifying the player’s face; identifying facial features of the player’s face; and warping a portion of the received visible spectrum image that includes the player’s face into a face texture map based on the identified facial features. A hair appearance of the player is captured based at least in part on the received data. The hair appearance is captured by identifying an extent of the player’s hair; matching the identified extent of the player’s hair to a predetermined hairstyle template; and identifying a color of the player’s hair. A 3D avatar reflecting the captured facial appearance and hair appearance of the player is generated such that the 3D avatar resembles the player. The 3D avatar is generated by combining the face texture map, hairstyle template that matches the identified extent of the player’s hair, and the identified player hair color with predetermined avatar features.

[0027] In another embodiment, a data acquisition component receives data describing the player’s current appearance. The data includes: a visible spectrum image of the player, a depth image including both the player and a current background, and skeletal data for the player. The skeletal data indicates an outline of the player’s skeleton. A face capture component captures a facial appearance of the player based at least in part on the received data. A hair capture component captures a hair appearance of the player based at least in part on the received data. A clothing capture component captures a clothing appearance of the player based at least in part on the received data. An avatar generation component generates a 3D avatar resembling the player by combining the captured facial appearance, hair appearance, and clothing appearance with predetermined avatar features.

[0028] In still another embodiment, data describing the player’s current appearance is received. The data includes: a visible spectrum image of the player; a depth image including both the player and a current background, and skeletal data for the player. The skeletal data indicates an outline of the player’s skeleton. A facial appearance of the player is captured based at least in part on the received data. The facial appearance is captured by identifying the player’s face; identifying facial features of the player’s face; and warping a portion of the received visible spectrum image that includes the player’s face into a face texture map based on the identified facial features. A hair appearance of the player is captured based at least in part on the received data. The hair appearance is captured by identifying an extent of the player’s hair; matching the identified extent of the player’s hair to a predetermined hairstyle template; and identifying a color of the player’s hair.

[0029] A clothing appearance of the player is captured based at least in part on the received data. The clothing appearance is captured, using the received depth image including both the player and the current background and the received visible spectrum image of the player, by determining an outer clothing boundary using edge detection. The clothing appearance is also captured by, using the received visible spectrum image of the player, determining at least one inner clothing boundary by distinguishing clothing from skin based on a determined color gradient. The clothing appearance is further captured by identifying player clothing as the area bounded at least in part by the outer clothing boundary and the at least one inner clothing boundary.

[0030] A skin color of the player is captured based at least in part on the received data. The skin color is captured by sampling the color of one or more areas of the received visible spectrum image in areas corresponding to the player’s face and blending the sampled colors of the one or more areas to determine a player skin color. A processor of a computing device is used to generate a 3D avatar reflecting the captured facial appearance, hair appearance, clothing appearance, and skin color of the player such that the 3D avatar resembles the player. The 3D avatar is generated by combining the face texture map, hairstyle template that matches the identified extent of the player’s hair, and identified player hair color with predetermined avatar features.

[0031] Having briefly described an overview of some embodiments of the present invention, an exemplary operating environment in which embodiments of the present invention may be implemented is described below in order to provide a general context for various aspects of the present invention. Referring initially to FIG. 1 in particular, an exemplary operating environment for implementing embodiments of the present invention is shown and designated generally as computing device 100. Computing device 100 is but one example of a suitable computing environment and is not intended to suggest any limitation as to the scope of use or functionality of embodiments of the present invention. Neither should the computing device 100 be interpreted as having any dependency or requirement relating to any one or combination of components illustrated.

[0032] Embodiments of the present invention may be described in the general context of computer code or machine-readable instructions, including computer-executable instructions such as program modules, being executed by a computer or other machine, such as a personal data assistant or other handheld device. Generally, program modules including routines, programs, objects, components, data structures, etc., refer to code that perform particular tasks or implement particular abstract data types. Embodiments of the present invention may be practiced in a variety of system configurations, including hand-held devices, consumer elec-
tronics, general-purpose computers, more specialty computing devices, etc. Embodiments of the present invention may also be practiced in distributed computing environments where tasks are performed by remote-processing devices that are linked through a communications network.

[0033] With reference to FIG. 1, computing device 100 includes a bus 110 that directly or indirectly couples the following devices: memory 112, one or more processors 114, one or more presentation components 116, input/output ports 118, input/output components 120, and an illustrative power supply 122. Bus 110 represents what may be one or more busses (such as an address bus, data bus, or combination thereof). Although the various blocks of FIG. 1 are shown with lines for the sake of clarity, in reality, delineating various components is not so clear, and metaphorically, the lines would more accurately be grey and fuzzy. For example, one may consider a presentation component such as a display device to be an I/O component. Also, processors have memory. We recognize that such is the nature of the art, and reiterate that the diagram of FIG. 1 is merely illustrative of an exemplary computing device that can be used in connection with one or more embodiments of the present invention. Distinction is not made between such categories as “workstation,” “server,” “laptop,” “hand-held device,” etc., as all are contemplated within the scope of FIG. 1 and reference to “computing device.”

[0034] Computing device 100 typically includes a variety of computer-readable media. Computer-readable media can be any available media that can be accessed by computing device 100 and includes both volatile and nonvolatile media, removable and non-removable media. By way of example, and not limitation, computer-readable media may comprise computer storage media and communication media. Computer storage media includes both volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer-readable instructions, data structures, program modules, or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by computing device 100.

[0035] Communication media typically embodies computer-readable instructions, data structures, program modules, or other data in a modulated data signal such as a carrier wave. The term “modulated data signal” refers to a propagated signal that has one or more of its characteristics set or changed to encode information in the signal. By way of example, and not limitation, communication media includes wired media, such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared, radio, microwave, spread-spectrum, and other wireless media. Combinations of the above are included within the scope of computer-readable media.

[0036] Memory 112 includes computer storage media in the form of volatile and/or nonvolatile memory. The memory may be removable, nonremovable, or a combination thereof. Exemplary hardware devices include solid-state memory, hard drives, optical-disc drives, etc. Computing device 100 includes one or more processors that read data from various entities such as memory 112 or I/O components 120. Presentation component(s) 116 present data indications to a user or other device. Exemplary presentation components include a display device, speaker, printing component, vibrating component, etc.

[0037] I/O ports 118 allow computing device 100 to be logically coupled to other devices including I/O components 120, some of which may be built in. Illustrative components include a microphone, joystick, game pad, satellite dish, scanner, printer, wireless device, etc.

[0038] As discussed previously, embodiments of the present invention relate to systems, methods, and computer media for generating an avatar resembling a player’s current appearance. Embodiments of the present invention will be discussed with reference to FIGS. 2-16.

[0039] FIGS. 2-4 illustrate an implementation of an embodiment of the present invention to acquire data describing a player and generate an avatar resembling the player’s current appearance. FIG. 2 illustrates a visible spectrum image 200 of player 201. Image 200 includes face 202, hair 204, and clothes 206 of player 201. In one embodiment, player 200 positions herself in front of a camera according to guide markers 208 to facilitate capture of image 201. FIG. 3 illustrates another visible spectrum image 300 of player 201. Substantially all of the body of player 201 is included in image 300. FIG. 3 also includes guide markers 308 to facilitate capture of image 308.

[0040] In addition to visible spectrum images 200 and 300, sensors gather one or more depth images of player 201. A depth image indicates the depth of each pixel in the image, which allows the background of an image to be distinguished from a person or object in the image and also allows identification of various body features of a person, such as facial features, based on the depth. The person can be recognized as including the pixels that are of a similar depth in the foreground of the image, whereas the background is determined to be the pixels having a larger depth.

[0041] Depth data also allows creation of skeletal data as known in the art. In one embodiment, the portion of an image representing a person can be distinguished from the background of the image using depth information. The portion of the image representing the person can then be compared to a database of poses to identify a match to the pose of the identified person. Once a match is identified, a simplified skeletal structure corresponding to the match can be identified. The simplified skeletal structure can indicate the outline of a player’s skeleton and can include points that represent joints and lines connecting the points to represent appendages or other body features. Other methods for identifying a simplified player skeleton from depth data are also envisioned.

[0042] Visible spectrum images 200 and 300, along with skeletal data and a depth image, are combined with predetermined avatar features to generate an avatar 400 shown in FIG. 4 resembling player 201. In some embodiments, the generated avatar has certain predetermined proportions, such as height and width, and certain predetermined features such as hands, feet, eyes, arm and leg width, or other features. Avatar 400 reflects face 202, hair 204, and clothes 206 of player 201, achieving an avatar that much more accurately represents player 201 that manual selection of various features can accomplish.

[0043] FIG. 5 illustrates a method of generating an avatar reflecting a player’s current appearance. In step 502, data describing the player’s current appearance is received. The data includes: a visible spectrum image of the player, a depth
image including both the player and a current background, and skeletal data for the player. The skeletal data indicates an outline of the player’s skeleton. In step 504, a facial appearance of the player is captured based at least in part on the received data: a hair appearance and clothing appearance are captured in steps 506 and 508; sleeves, pant legs, or hems are detected in step 510; and skin color is captured in step 512. In step 514, a 3D avatar resembling the player is generated by combining the captured facial appearance, hair appearance, clothing appearance, sleeves/leg/hems, and skin color with predetermined avatar features.

**[0044]** FIG. 6 illustrates a system 600 in accordance with an embodiment of the present invention. Data acquisition component 602 receives data describing the player’s current appearance from sensors 604. The data includes: a visible spectrum image of the player, a depth image including both the player and a current background, and skeletal data for the player. The skeletal data indicates an outline of the player’s skeleton. Sensors 604 may include an RGB camera and an infrared or other depth sensor. In one embodiment, a player stands in the path of sensors 604 in a variety of preferred positions to facilitate acquisition of sensor data. Data acquisition component 602 provides received data to face capture component 606, hair capture component 608, clothing capture component 610, sleeve detection component 612; and skin color capture component 614. Components 606, 608, 610, 612, and 614 capture or detect information and provide this information to avatar generation component 616, which generates a 3D avatar resembling the player using the information.

**[0045]** In various embodiments, any combination of components 606, 608, 610, 612, and 614 are included. Face capture component 606 captures a facial appearance of the player based at least in part on the received data. Hair capture component 608 captures a hair appearance of the player based at least in part on the received data. Clothing capture component 610 captures a clothing appearance of the player based at least in part on the received data. Sleeve detection component 612 identifies, based at least in part on the received data, at least one of: (1) a shirt sleeve, (2) a pants or shorts leg, and (3) a skirt or dress hem. Skin color capture component 614 identifies a skin color of the player based at least in part on the received data. Avatar generation component 616 generates a 3D avatar resembling the player by combining the captured facial appearance, hair appearance, and clothing appearance with predetermined avatar features. It is envisioned that the functionality of components 602, 606, 608, 610, 612, 614, and 616 may reside on separate physical components or devices or may be implemented together.

**[0046]** FIG. 7 illustrates a method 700 of generating an avatar reflecting a player’s current appearance. The steps of method 700 may be implemented by the components of system 600 of FIG. 6. In step 702, data describing a player’s current appearance is received. This data includes at least one visible spectrum image 702A of the player, at least one depth image 702B including both the player and a current background, and skeletal data 702C for the player. Skeletal data 702C indicates an outline of the player’s skeleton and may be derived from the at least one depth image 702B.

**[0047]** In step 704, a facial appearance of the player is captured based at least in part on the received data. Step 704 can be implemented by sub-steps 704A-704C. In sub-step 704A, the player’s face is identified. In sub-step 704B, a portion of the received visible spectrum image that includes the player’s face is warped into a face texture map based on the identified facial features. The face texture map is a UV space map that identifies how to map 2D image points to a 3D model.

**[0048]** In step 706, a hair appearance of the player is captured based at least in part on the received data. Step 706 can be implemented by sub-steps 706A-706C. In sub-step 706A, the extent of the player’s hair is identified. In sub-step 706B, the identified extent of the player’s hair is matched to a predetermined hairstyle template. In sub-step 706C, a color of the player’s hair is identified.

**[0049]** In step 708, a clothing appearance of the player is captured based at least in part on the received data. Step 708 can be implemented by sub-steps 708A-708C. In sub-step 708A, using the received depth image including both the player and the current background and the received visible spectrum image of the player, an outer clothing boundary is determined using edge detection. In sub-step 708B, using the received visible spectrum image of the player, at least one inner clothing boundary is determined by distinguishing clothing from skin based on a determined color gradient. In sub-step 708C, a clothing appearance is determined for the area bounded at least in part by the outer clothing boundary and the at least one inner clothing boundary.

**[0050]** A skin color of the player is captured in step 710, the capture based at least in part on the received data. Step 710 can be implemented by sub-steps 710A-710B. In sub-step 710A, the color of one or more areas of the received visible spectrum image is sampled in areas corresponding to the player’s face. In sub-step 710B, the sampled colors of the one or more areas are blended to determine a player skin color.

**[0051]** In step 712, a 3D avatar is generated reflecting the captured facial appearance, hair appearance, clothing appearance, and skin color of the player such that the 3D avatar resembles the player by combining the face texture map, hairstyle template that matches the identified extent of the player’s hair, and identified player hair color with predetermined avatar features.

**[0052]** Facial appearance capture, hair appearance capture, clothing appearance capture, and sleeve/leg/hem detection, discussed with reference to FIGS. 5-7 will now be discussed in greater detail.

**[0053]** Facial Appearance Capture

**[0054]** As discussed above, a visible spectrum image, depth image, and skeletal data can be used to capture a player’s facial appearance. The captured facial appearance can then be incorporated into a 3D avatar. In one embodiment, a players head can be identified by analyzing the received skeletal data for the player. A head portion of the received visible spectrum image that corresponds to the player’s head can then be identified. Thus, a rough location of the player’s head and face can first be identified with skeletal data, and the portion of the visible spectrum image corresponding to this location can be analyzed further. In one embodiment, a face detection algorithm is applied to the head portion of the received visible spectrum image. In one particular algorithm, machine learning techniques are used that compare the face to a database of known faces.

**[0055]** From an identified face, a variety of facial features can be identified. In some embodiments, face alignment points corresponding to the identified facial features are determined. The identified face alignment points are matched
to destination points on a template texture map. The portion of the received visible spectrum image that includes the player’s face is warped into a face texture map such that the face texture map includes the face alignment points corresponding to the identified facial features of the player’s face mapped to the destination points of the template texture map. By doing this, an image of a player’s face is slightly distorted to fit a destination template. Texture maps as used in this application are in the UV space and contain information allowing them to be mapped to a 3D model. Thus, by identifying facial features and corresponding alignment points and mapping the alignment points to destination points of a template texture map, a 3D model of the player’s face can now be created.

[0056] This is illustrated in FIGS. 8-13. FIG. 8 shows visible spectrum image 800 of player 802. FIG. 9 shows a second visible spectrum image 900 of player 802. The guide markers shown in FIGS. 8 and 9 are shown to facilitate feature detection—i.e., player 802 positions his body in locations the system is expecting, feature detection becomes more accurate. FIG. 10 illustrates the face 1000 of player 802. Face 1000 is shown here along with a hair and a portion of the player’s clothing. FIG. 11 illustrates exemplary face alignment points 1102 corresponding to identified facial features. Facial features can include the eyes, chin, ears, nose, mouth, eyebrows, jaw, and other features. In one embodiment, 99 face alignment points are used as follows: 0-7 left eye, 8-15 right eye, 16-25 left eyebrow, 26-35 right eyebrow, 36-47 nose, 48-59 outer mouth, 60-67 inner mouth, 68-86 jawline, 87 left eye center, 88 right eye center, 89 nose center, and 90-99 above eyebrows.

[0057] FIG. 12 illustrates a template texture map 1200 that defines a relationship between the destination points 1202 and a 3D model. Texture maps in the UV space are well known in the art. Face alignment points 1102 serve as source points to destination points 1202. Template texture map 1200 may be an artist-created map to map to a particular stylized 3D head model. Various template texture maps may be used. In some embodiments, not all alignment points need to be aligned with destination points. For example, eye and mouth alignment points may not be matched to better preserve a player’s expression. FIG. 13 illustrates the resulting face texture map 1300. Face texture map 1300 is distorted slightly as compared to image 1000 of FIG. 10. Face texture map 1300 can be morphed into a 3D head model based representing player 802’s head that is included as part of a 3D avatar reflecting player 802’s current appearance. FIG. 14 illustrates avatar 1400 having a face and head 1402 resembling player 802.

[0058] In some embodiments, data describing the player’s current appearance is only received for the front of the player. Data for the sides and back of the player is inferred. In other embodiments, images of the player from multiple angles or sides are received, and feature recognition is performed for each side of a player’s head and body.

[0059] Various filtering and processing steps may also be undertaken. In one embodiment, the following filters and values are applied: bilateral filter to an RGB image (kernel radius 3 pixels, edge threshold (0-255) of 20; color noise reduction (bilateral filter; H/S/V color space), kernel radius 5 pixels, edge threshold (0-255) of 42; auto-contrast (H/S/V space), re-center value (0-1) of 0.54, re-Standard Deviation value (0-1) of 0.15, and blend with original (0-1) of 0.5; adjust brightness curve (H/S/V space), curve point 1 of 0.6, 1.0, blend with original (0-1) of 0.5.

[0060] In some embodiments, prior to morphing the face texture map into a 3D head model, the alignment points are analyzed, and the size of the corresponding facial features are calculated. Based on the face texture map and the calculated size of the plurality of identified facial features, a plurality of predetermined head models are combined to create a morphed 3D head model.

[0061] In one embodiment, the measurements are normalized to consider the actual size of the player’s head in the visible spectrum image to account for the image being of a head close up or far away. A morph configuration is then determined for each feature by comparing the normalized values to mean and standard deviation values. This effectively provides information regarding how large/small/wide/narrow, etc. a particular feature is as compared to the average. A variety of artist-created 3D head models exist, and appropriate features from various models are combined according to the morph configurations. Thus, normalized face dimensions are converted into morph configurations, and the morph configurations specify how to blend the various head models to reflect the player’s features.

[0062] In one particular embodiment, morph configurations are categorized as follows: jaw/chin as wide, narrow, long, short; and nose as narrow, wide, short, long; eye region as wide spacing, narrow spacing, placed high or placed low.

[0063] Hair Appearance Capture

[0064] In addition to capturing a player’s face, a player’s hair appearance can be captured. Hair appearance includes hair color as well as an approximate hair style. In one embodiment, various aspects of a player’s hair such as bangs, amount on top, amount on each side, amount beyond chin, forehead coverage, etc. are analyzed and combined to match predefined hairstyle templates.

[0065] In one embodiment, the extent of the player’s hair is identified by determining an inner and outer hair boundary. Using the received depth image including both the player and the current background, an outer hair boundary is determined by distinguishing the depth of the background from the depth of the player’s hair. Using the received visible spectrum image of the player, an inner hair boundary is determined by distinguishing the player’s skin and/or clothing from the player’s hair. In one embodiment, the inner hair boundary is determined by distinguishing the player’s hair from clothing worn by the player and by determining the extent of the player’s forehead using the facial features identified in capturing the facial appearance of the player. Knowing the location of the player’s forehead helps narrow the locations in which to analyze the visible spectrum image for hair.

[0066] The player’s hair is identified as the area bounded at least in part by the outer and inner hair boundaries. In some embodiments, a plurality of hair attributes of the player’s hair are identified, each of the plurality of hair attributes describing a portion or characteristic of the player’s hair. Hair attributes include but are not limited to bangs, amount on top, amount on each side, amount beyond chin, forehead coverage, etc. The identified plurality of hair attributes can then be compared to a library of predetermined hairstyle templates to find a best match. In some embodiments, the library contains between 10 and 20 pre-determined hairstyles. In other embodiments, the distance away from the player’s head affects how likely a particular pixel is to be classified as hair or not—the closer to the player’s head, the more likely the pixel is to be considered hair.
The details of hair appearance capture can be implemented in a variety of ways. Several specific examples of various aspects of hair appearance capture are presented below. Hair can be segmented from a color visible spectrum image using depth-informed background removal, forehead estimation, and face removal. Depth-informed background removal can be accomplished by: deciding on a background threshold (e.g., face distance > 20 cm); initializing color image to all-background (tag with special color or alpha); marking as background all depth pixels when they have either value = background or value = 0 (depth hole); marking as background all depth pixels whose immediate neighbors were previously marked as background (don’t spread from new ones, however)—this erodes depth edges inwards; mapping color pixels to depth pixels; copying all source color pixels which correspond to non-background depth pixels into the target color image; and filling the color image back out along background edges using range-limited flood-fill—for any given non-background color pixel which borders on a background pixel, moving outward away from the edge and copy pixels from the source image to the target image until either a certain distance has been travelled or the source image pixel color deviates too much from the original edge pixel color.

Forehead estimation can be accomplished by: sampling central face color reference using face-detection bounds and seed the color average with this color; sampling across the width of the forehead region starting just above the eyebrows and walk up the forehead, keeping a running average of the color; when one line of sample colors differs significantly from the average, resetting the running average and record a transition point; and stopping walking upwards when only background pixels are sampled and record a transition point. The first recorded transition point is likely the hairline level, and the last transition point is the top of the hair. Eyebrow level is the bottom of the forehead.

Face removal can be accomplished by: computing a 2D convex hull of face-detection feature points that gives a good outline of the lower face but not the forehead; reflecting some of the jawline points over the line running through the eyes and add those to the convex hull, using the forehead hairline estimate to avoid placing the reflected points above the hairline; marking as background all color pixels within the convex hull; and marking as background all color pixels within the vertical region between the mouth and eyes, excluding the ears.

In another embodiment, hair color is sampled by: constraining sampling to a rectangle around the face, high enough to get tall hair and low enough to get shoulder-length long hair; sampling all color pixels not marked as background for an overall average; and sampling all the non-background color pixels again, excluding any pixels which are significantly different from the average color in custom-scaled HCV color space. If hair style estimation is complete, the color gathered from the sides of the players head can be emphasized over the color gathered from the top, because the sides tend to be less influenced by lighting conditions.

In one embodiment, hairstyle is estimated by dividing the hair sampling area into five regions: top center, upper left/right, and lower left/right. The horizontal threshold between top and upper left/right is at the edge of the eyebrows, and the threshold between upper & lower is the middle of the face (which should be blank from removing the ears). Hair color classification is used to sample the number of hair pixels in each region, and the forehead estimation is used to measure the forehead height and top-hair height. Five values have now been obtained: forehead height, top-hair height, top-hair area, upper-sides area, lower-sides area. The top-hair height and area values can be blended into a single value, resulting in four values. The values are normalized to the estimated face size in pixels based on face detection. Each value can then be classified into bins such as “none”, “some”, and “lots”, where the threshold values are empirically derived through human or machine learning. The binned values are matched against an asset matrix to select a visual hairstyle to display as feedback for a particular player.

Hue-chroma-value (HCV) color space can be used for perceptual color-difference comparisons. This color space is cone-shaped, with black as a single point at the bottom and white as a single point at the top, and saturated colors forming the widest part of the cone at the top. HSV has almost-black colors spread widely across the bottom of a cylinder, which causes problems when image noise can induce wide fluctuations in hue & saturation. The distance between two values is measured using polar coordinate distance on the HC plane and adding Euclidean distance for V. The radius and height of the cone can be scaled to emphasize different characteristics of the color. The color-space distance threshold is another variable to manipulate, but 1 of these 3 variables can be fixed. A different scaling approach holds the top of the cone at a fixed radius of 1 and instead spreads out the bottom point of the cone into a disc by blending chroma and saturation, with a radius of distance threshold / 2. This approach guarantees that all dark colors are within the threshold distance of one another without affecting the saturated colors at the top of the pseudo-cone. In one embodiment, specific tuning values for hair include: distance threshold = 0.1; H = H / C = lerpf(S, C, distance threshold / 2) (threshold = 0 produces pure HCV); and V = V * 0.5.

Clothing Appearance Capture

A clothing appearance can also be captured from the received data describing the player’s current appearance. The generated 3D avatar can then include clothing similar to that which the player is wearing.

In one embodiment, the clothing appearance is captured by: using the received depth image including both the player and the current background and the received visible spectrum image of the player, determining an outer clothing boundary using edge detection; using the received visible spectrum image of the player, determining at least one inner clothing boundary by distinguishing clothing from skin based on a color analysis; and identifying player clothing as the area bounded at least in part by the outer clothing boundary and the at least one inner clothing boundary. In another embodiment, the edge detection distinguishes the depth of the background from the depth of the player in the received depth image and distinguishes the color of clothes from the color of the background in the received visible spectrum image.

In some embodiments, body features of the player are identified using the received skeletal data. The outer clothing boundary and the at least one inner clothing boundary are determined by analyzing areas of the received visible spectrum image and areas of the received depth image corresponding to identified body features of the player where clothes are typically worn. For example, wrists, elbows, shoulders, knees, hips, etc can be identified from skeletal data, providing the system a narrowed analysis area for recognizing clothing.
In other embodiments, capturing the clothing appearance of the player further comprises: rescaling the skeletal data for the player to a predetermined skeleton size to form a scaled player skeleton; reposing the scaled player skeleton to a predetermined pose; and creating a clothing texture map based on the reposed, scaled player skeleton and the identified player clothing. A 3D body model generated from the clothing texture map is then used in generating the 3D avatar.

In some embodiments, particular articles of clothing are identified and captured. In other embodiments, and entire body clothing capture is performed. In still other embodiments, when clothing appearance capture is unable to detect particular clothing to a minimum threshold, a short-sleeve shirt and pants or shorts are used as the default. Similarly, if no clothing is detected on the player’s torso, a short-sleeve shirt may be included in the avatar.

In one embodiment, skinning of the provided captured depth map positions onto the tracked skeleton is performed. A mesh is then generated. Skin removal is then performed by detecting skin pixels and filling intelligently into areas that may not register as skin but are attached to skin pixels. The background is removed using a Canny edge detect and eroding the player mask by detecting edges in the visible spectrum image close to the edge of the area identified as the player. A bone scale calculation is then performed to try to stretch bones lengthwise and width-wise to fill the mask. The skeletal data is used to pose the skeleton into the provided 3D skeletal positions. The visible spectrum image is then applied onto the mask using the newly retargeted mesh. A per-pixel stretch and/or simple RGB flood fill is used to fill in the remaining white space of the mask.

Sleeve/Leg/Hem Detection

Detection of a shirt sleeve, pants or shorts leg, or skirt or dress hem may be thought of as either separate functionality or as part of clothing appearance capture. In one embodiment, areas corresponding to the arms and legs of the player are identified using the received skeletal data. The color of the received visible spectrum image of the player is then analyzed along the identified areas corresponding to the arms and legs of the player to detect at least one of: (1) a shirt sleeve, (2) a pants or shorts leg, and (3) a skirt or dress hem. The analysis may be, for example, color classification or gradient analysis. In one embodiment, the skeletal data can be used to identify the path along which to analyze the color of the visible spectrum image. For example, because the arm portion of the skeleton is known, the color of pixels can be analyzed moving down the arm from the shoulder to see where a sleeve occurs.

In another embodiment, a skirt or dress hem can be detected. Using the received skeletal data, areas corresponding to the legs of the player are identified. Using the received depth image including both the player and the current background, the area between the identified areas corresponding to the legs of the player is analyzed to detect a skirt or dress edge by distinguishing background depth from player depth and identifying a substantially horizontal edge separating the background depth from the player depth.

Skeletal data has been discussed at various points in the above disclosure. FIG. 15 illustrates a visible spectrum image 1500 of a player 1502, and FIG. 16 illustrates a corresponding depth image 1600 of player 1502. Depth image 1600, which may be produced by an infrared sensor measuring time-of-flight from a sensor, to a sensed object, and back, shows background portion 1604 at a first depth and player portion 1602 at a second depth closer to the front. A simplified skeleton 1606 can be determined based on the pose of player portion 1602. In one embodiment, player portion 1602 is compared to various player poses in a database, and upon identifying a match, skeleton 1606 is determined to be the skeleton of the matching pose in the database.

The present invention has been described in relation to particular embodiments, which are intended in all respects to be illustrative rather than restrictive. Alternative embodiments will become apparent to those of ordinary skill in the art to which the present invention pertains without departing from its scope.

From the foregoing, it will be seen that this invention is one well adapted to attain all the ends and objects set forth above, together with other advantages which are obvious and inherent to the system and method. It will be understood that certain features and sub-combinations are of utility and may be employed without reference to other features and sub-combinations. This is contemplated by and is within the scope of the claims.

Having thus described the invention, what is claimed is:

1. A system for generating an avatar reflecting a user’s current appearance comprising:
   a data acquisition component configured to generate data corresponding to the user’s current appearance, wherein the data includes a visible spectrum image of the user, a depth image associated with the user and a background, and skeletal data indicating an outline of the user’s skeleton;
   a face capture component that receives the data and generates a face texture map comprising a plurality of face alignment points by:
   identifying a head portion of the visible spectrum image corresponding to the user’s head and face using the skeletal data;
   identifying facial features in the identified head portion;
   determining face alignment points corresponding to identified facial features; and
   mapping face alignment points to identified facial features to associated destination points of a template texture map;
   and
   an avatar generation component that generates the avatar using the face texture map.

2. The system of claim 1, further comprising:
   a clothing capture component that captures an appearance of the user’s clothing, wherein the appearance of the user’s clothing comprises an outer clothing boundary identified using the depth image and at least one inner clothing boundary identified using the visible spectrum image.

3. The system of claim 2, wherein the clothing capture component also:
   identifies areas of the visible spectrum image of the user corresponding to an arm, a leg, or a combination thereof using the skeletal data; and
   analyzes identified areas of the visible spectrum image using either color classification or gradient analysis to detect one or more of: a shirt sleeve, a pants leg, a shorts leg, a skirt hem, a dress hem, or a combination thereof.

4. The system of claim 3, wherein the clothing capture component also:
   identifies areas of the depth image of the user corresponding to the user’s legs using the skeletal data; and
analyzes between the identified areas of the depth image to detect either a skirt hem or a dress hem by:

- distinguishing a background depth from a user depth; and
- identifying a substantially horizontal edge separating the background depth from the user depth.

5. The system of claim 1, further comprising:

- a hair capture component that captures attributes of the user’s hair appearance, in part, by identifying an extent of the user’s hair and matches the captured attributes to a predetermined hairstyle template.

6. One or more computer storage devices storing computer-useable instructions that, when used by one or more computing devices, cause the one or more computing devices to perform a method for generating an avatar reflecting a user’s current appearance, the method comprising:

- receiving data corresponding to the user’s current appearance, wherein the data includes a visible spectrum image of the user, a depth image associated with the user and a background, and skeletal data indicating an outline of the user’s skeleton;
- identifying a head portion of the visible spectrum image corresponding to the user’s head using the skeletal data;
- identifying a face portion within the head portion corresponding to the user’s face and facial features within the face portion;
- identifying face alignment points corresponding to identified facial features;
- mapping face alignment points to corresponding destination points of a template texture map;
- capturing a plurality of hairstyle attributes, in part, by identifying an extent of the user’s hair;
- matching the identified extent of the user’s hair to a predetermined hairstyle template; and
- generating the avatar using the face texture map and the captured plurality of hairstyle attributes.

7. The one or more computer storage devices of claim 6, wherein the face portion is identified by comparing the head portion to a database of known faces using machine learning techniques.

8. The one or more computer storage devices of claim 6, wherein the extent of the user’s hair comprises an outer hair boundary identified using the depth image.

9. The one or more computer storage devices of claim 6, wherein the extent of the user’s hair comprises an inner hair boundary identified by analyzing the visible spectrum image.

10. The one or more computer storage devices of claim 9, wherein the inner hair boundary is identified, in part, using depth-informed background removal.

11. The one or more computer storage devices of claim 9, wherein the method further comprises estimating the user’s forehead location within the face portion by sampling a central face color wherein the inner hair boundary is identified, in part, using the estimated forehead location.

12. The one or more computer storage devices of claim 6, wherein capturing the plurality of hairstyle attributes further comprises segmenting the user’s hair from the visible spectrum image using depth-informed background removal, forehead estimation, face removal, or a combination thereof.

13. The one or more computer storage devices of claim 6, wherein the captured plurality of hairstyle attributes comprise the identified extent of the user’s hair and one or more aspects of the user’s hair including: bangs, an amount of hair on top, an amount of hair on each side, an amount beyond the user’s chin, forehead coverage, or a combination thereof; and wherein a combination of the captured plurality of hairstyle attributes are matched to at least one predetermined hairstyle template.

14. The one or more computer storage devices of claim 6, wherein the method further comprises identifying the user’s hair color using the visible spectrum image, wherein the identified hair color is applied to the predetermined hairstyle template.

15. A computerized method comprising for generating an avatar reflecting a user’s current appearance, the method comprising:

- receiving data corresponding to the user’s current appearance, wherein the data includes a visible spectrum image of the user, a depth image associated with the user and a background, and skeletal data indicating an outline of the user’s skeleton;
- identifying a head portion of the visible spectrum image corresponding to the user’s head using the skeletal data;
- identifying a face portion within the head portion corresponding to the user’s face and facial features within the face portion;
- determining face alignment points corresponding to identified facial features;
- calculating sizes of a plurality of facial features using the determined face alignment points;
- determining morph configurations for each of the plurality of facial features using the calculated sizes;
- combining a plurality of predetermined three-dimensional (3D) head models based upon the determined morph configurations;
- mapping face alignment points corresponding to identified facial features to associated destination points of the morphed 3-D head model; and
- generating the avatar using the morphed 3-D head model.

16. The computerized method of claim 15, further comprising:

- capturing a plurality of hairstyle attributes, in part, by identifying an extent of the user’s hair.

17. The computerized method of claim 16, wherein identifying the extent of the user’s hair comprises identifying an outer hair boundary using the depth image and an inner hair boundary by analyzing the visible spectrum image.

18. The computerized method of claim 15, further comprising:

- capturing an appearance of the user’s clothing, wherein the appearance of the user’s clothing comprises an outer clothing boundary identified using the depth image and at least one inner clothing boundary identified using the visible spectrum image.

19. The computerized method of claim 18, wherein capturing the appearance of the user’s clothing further comprises:

- identifying areas of the visible spectrum image of the user corresponding to an arm, a leg, or a combination thereof using the skeletal data; and
- analyzing the identified areas of the visible spectrum image using either color classification or gradient analysis to detect one or more of: a shirt sleeve, a pant leg, a shorts leg, a skirt hem, a dress hem, or a combination thereof.

20. The computerized method of claim 18, wherein capturing the appearance of the user’s clothing further comprises:
identifying areas of the depth image of the user corresponding to the user’s legs using the skeletal data; and analyzing between the identified areas of the depth image to detect either a skirt hem or a dress hem by:
distinguishing a background depth from a user depth; and
identifying a substantially horizontal edge separating the background depth from the user depth.

* * * * *