



- (51) **International Patent Classification:**
G09G 5/00 (2006.01) *G06F 3/01* (2006.01)
- (21) **International Application Number:**
PCT/US2012/033505
- (22) **International Filing Date:**
13 April 2012 (13.04.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
13/104,346 10 May 2011 (10.05.2011) US
- (71) **Applicant (for all designated States except US):** APPLE INC. [US/US]; 1 Infinite Loop, Cupertino, California 95014 (US).
- (72) **Inventor; and**
- (75) **Inventor/Applicant (for US only):** DJAVAHERIAN, Amir [US/US]; 1 Infinite Loop, MS 60-1 ME, Cupertino, California 95014 (US).
- (74) **Agents:** NICHOLLES, Christian A. et al.; Kilpatrick Townsend & Stockton LLP, Two Embarcadero Center, 8th Floor, San Francisco, California 94111-3834 (US).

- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

(54) **Title:** SCALING OF VISUAL CONTENT BASED UPON USER PROXIMITY

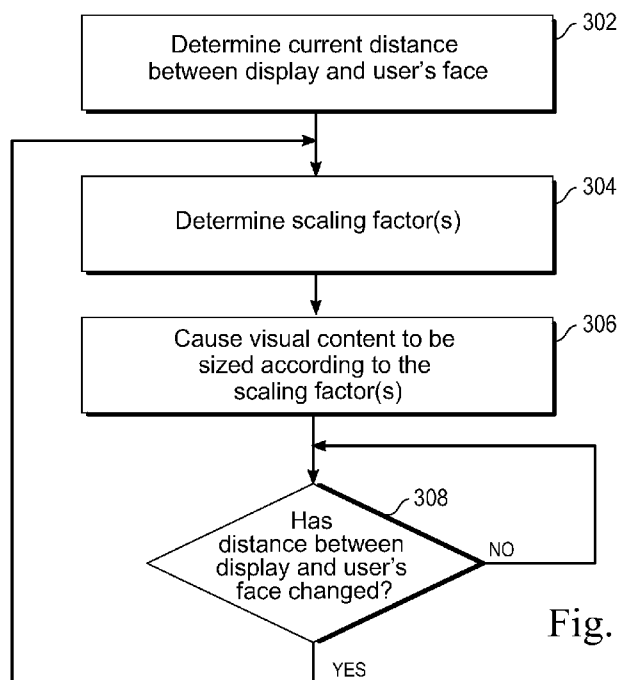


Fig. 3

(57) **Abstract:** A mechanism is disclosed for automatically scaling the size of a set of visual content based upon how close a user's face is to a display. In one implementation, the mechanism initially causes a set of visual content on a display to be sized according to a first scaling factor when the user's face is at a first distance from the display. The mechanism then determines that the user's face has moved relative to the display such that the user's face is no longer at the first distance from the display. In response, the mechanism causes the set of visual content on the display to be sized according to a second and different scaling factor. By doing so, the mechanism effectively causes the display size of the visual content to automatically change as the distance between the user's face and the display changes.

SCALING OF VISUAL CONTENT BASED UPON USER PROXIMITY

Background

[0001] Many of today's computing devices allow a user to scale the visual content that is being displayed to a size of the user's liking. For example, some smart phones and tablet computing devices allow a user to put two fingers on a touch sensitive display and either pinch the fingers together or spread them apart. Pinching the fingers together causes the display size of the visual content to be reduced, while spreading the fingers apart causes the display size of the visual content to be enlarged. By adjusting the scale of the visual content, the user can set the visual content to a size that is comfortable for him/her.

[0002] Often, during the course of using a computing device, especially one that is portable such as a smart phone or a tablet, a user may position the display of the computing device at different distances from the user's face at different times. For example, when the user starts using a computing device, the user may hold the display of the computing device at a relatively close distance X from the user's face. As the user's arm becomes fatigued, the user may set the computing device down on a table or on the user's lap, which is at a farther distance Y from the user's face. If the difference between the distances X and Y is significant, the scale of the visual content that was comfortable for the user at distance X may no longer be comfortable for the user at distance Y (e.g. the font size that was comfortable at distance X may be too small at distance Y). As a result, the user may have to manually readjust the scale of the visual content to make it comfortable at distance Y. If the user moves the display to different distances many times, the user may need to manually readjust the scale of the visual content many times. This can become inconvenient and tedious.

Brief Description of the Drawing(s)

[0003] Fig. 1 shows a block diagram of a sample computing device in which one embodiment of the present invention may be implemented.

[0004] Fig. 2 shows a flow diagram for a calibration procedure involving a distance determining component, in accordance with one embodiment of the present invention.

[0005] Fig. 3 shows a flow diagram for an automatic scaling procedure involving a distance determining component, in accordance with one embodiment of the present invention.

[0006] Fig. 4 shows a flow diagram for a calibration procedure involving a user-facing camera, in accordance with one embodiment of the present invention.

[0007] Fig. 5 shows a flow diagram for an automatic scaling procedure involving a user-facing camera, in accordance with one embodiment of the present invention.

Detailed Description of Embodiment(s)

Overview

[0008] In accordance with one embodiment of the present invention, a mechanism is provided for automatically scaling the size of a set of visual content based, at least in part, upon how close a user's face is to a display. By doing so, the mechanism relieves the user from having to manually readjust the scale of the visual content each time the user moves the display to a different distance from his/her face. In the following description, the term visual content will be used broadly to encompass any type of content that may be displayed on a display device, including but not limited to text, graphics (e.g. still images, motion pictures, etc.), webpages, graphical user interface components (e.g. buttons, menus, icons, etc.), and any other type of visual information.

[0009] According to one embodiment, the mechanism automatically rescales a set of visual content in the following manner. Initially, the mechanism causes a set of visual content on a display to be sized according to a first scaling factor when the user's face is at a first distance from the display. The mechanism then determines that the user's face has moved relative to the display such that the user's face is no longer at the first distance from the display. This determination may be made, for example, based upon sensor information received from one or more sensors. In response to a determination that the user's face has moved relative to the display, the mechanism causes the set of visual content on the display to be sized according to a second and different scaling factor. By doing so, the mechanism effectively causes the display size of the visual content to automatically change as the distance between the user's face and the display changes.

[0010] As used herein, the term scaling factor refers generally to any one or more factors that affect the display size of a set of visual content. For example, in the case where the visual content includes text, the scaling factor may include a font size for the text. In the case where the visual content includes graphics, the scaling factor may include a magnification or zoom factor for the graphics.

[0011] In one embodiment, as the user's face gets closer to the display, the scaling factor, and hence, the display size of the visual content is made smaller (down to a certain minimum limit), and as the user's face gets farther from the display, the scaling factor, and hence, the

display size of the visual content is made larger (up to a certain maximum limit). In terms of text, this may mean that as the user's face gets closer to the display, the font size is made smaller, and as the user's face gets farther away from the display, the font size is made larger. In terms of graphics, this may mean that as the user's face gets closer to the display, the magnification factor is decreased, and as the user's face gets farther from the display, the magnification factor is increased. In this embodiment, the mechanism attempts to maintain the visual content at a comfortable size for the user regardless of how far the display is from the user's face. Thus, this mode of operation is referred to as comfort mode.

[0012] In an alternative embodiment, as the user's face gets closer to the display, the scaling factor, and hence, the display size of the visual content is made larger (thereby giving the impression of "zooming in" on the visual content), and as the user's face gets farther from the display, the scaling factor, and hence, the display size of the visual content is made smaller (thereby giving the impression of "panning out" from the visual content). Such an embodiment may be useful in various applications, such as in games with graphics, image/video editing applications, mapping applications, etc. By moving his/her face closer to the display, the user is in effect sending an implicit signal to the application to "zoom in" (e.g. to increase the magnification factor) on a scene or a map, and by moving his/her face farther from the display, the user is sending an implicit signal to the application to "pan out" (e.g. to decrease the magnification factor) from a scene or a map. Because this mode of operation provides a convenient way for the user to zoom in and out of a set of visual content, it is referred to herein as zoom mode.

[0013] The above modes of operation may be used advantageously to improve a user's experience in viewing a set of visual content on a display.

Sample Computing Device

[0014] With reference to Fig. 1, there is shown a block diagram of a sample computing device 100 in which one embodiment of the present invention may be implemented. As shown, device 100 includes a bus 102 for facilitating information exchange, and one or more processors 104 coupled to bus 102 for executing instructions and processing information. Device 100 also includes one or more storages 106 (also referred to herein as computer readable storage media) coupled to the bus 102. Storage(s) 106 may be used to store executable programs, permanent data, temporary data that is generated during program execution, and any other information needed to carry out computer processing.

[0015] Storage(s) 106 may include any and all types of storages that may be used to carry out computer processing. For example, storage(s) 106 may include main memory (e.g. random access memory (RAM) or other dynamic storage device), cache memory, read only memory (ROM), permanent storage (e.g. one or more magnetic disks or optical disks, flash storage, etc.), as well as other types of storage. The various storages 106 may be volatile or non-volatile. Common forms of computer readable storage media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, DVD, or any other optical storage medium, punchcards, papertape, or any other physical medium with patterns of holes, a RAM, a PROM, an EPROM, a FLASH-EPROM or any other type of flash memory, any memory chip or cartridge, and any other storage medium from which a computer can read.

[0016] As shown in Fig. 1, storage(s) 106 store at least several sets of executable instructions, including an operating system 114 and one or more applications 112. The processor(s) 102 execute the operating system 114 to provide a platform on which other sets of software may operate, and execute one or more of the applications 112 to provide additional, specific functionality. For purposes of the present invention, the applications 112 may be any type of application that generates visual content that can be scaled to different sizes. In one embodiment, the automatic scaling functionality described herein is provided by the operating system 114 as a service to the applications 112. Thus, when an application 112 has a set of visual content that it wants to render to a user, it calls to the operating system 114 and asks for a scaling factor. It then uses the scaling factor to scale the visual content. As an alternative, the application 112 may provide the visual content to the operating system 114, and ask the operating system 114 to scale the visual content according to a scaling factor determined by the operating system 114. As an alternative to having the operating system 114 provide the automatic scaling functionality, the automatic scaling functionality may instead be provided by the applications 112 themselves. As a further alternative, the automatic scaling functionality may be provided by a combination of or cooperation between the operating system 114 and one or more of the applications 112. All such possible divisions of functionality are within the scope of the present invention.

[0017] The device 100 further comprises one or more user interface components 108 coupled to the bus 102. These components 108 enable the device 100 to receive input from and provide output to a user. On the input side, the user interface components 108 may include, for example, a keyboard/keypad having alphanumeric keys, a cursor control device (e.g. mouse, trackball, touchpad, etc.), a touch sensitive screen, a microphone for receiving

audio input, etc. On the output side, the components 108 may include a graphical interface (e.g. a graphics card) and an audio interface (e.g. sound card) for providing visual and audio content. The user interface components 108 may further include a display 116, a set of speakers, etc., for presenting the audio and visual content to a user. In one embodiment, the operating system 114 and the one or more applications 112 executed by the processor(s) 104 may provide a software user interface that takes advantage of and interacts with the user interface components 108 to receive input from and provide output to a user. This software user interface may, for example, provide a menu that the user can navigate using one of the user input devices mentioned above.

[0018] The user interface components 108 further include one or more distance indicating components 118. These components 118, which in one embodiment are situated on or near the display 116, provide information indicating how far a user's face is from the display 116. Examples of distance indicating components 118 include but are not limited to: an infrared (IR) sensor (which includes an IR emitter and an IR receiver that detects the IR signal reflected from a surface); a laser sensor (which includes a laser emitter and a laser sensor that detects the laser signal reflected from a surface); a SONAR sensor (which includes an audio emitter and an audio sensor that detects the audio signal reflected from a surface); and a user-facing camera. With an IR sensor, the distance between the IR sensor and a surface (e.g. a user's face) may be calculated based upon the intensity of the IR signal that is reflected back from the surface and detected by the IR sensor. With a laser sensor and a SONAR sensor, the distance between the sensor and a surface may be calculated based upon how long it takes for a signal to bounce back from the surface. With a user-facing camera, distance may be determined based upon the dimensions of a certain feature of a user's face (e.g. the distance between the user's eyes). Specifically, the closer a user is to the camera, the larger the dimensions of the feature would be. In one embodiment, the one or more distance indicating components 118 provide the sensor information needed to determine how close a user's face is to the display 116.

[0019] In addition to the components set forth above, the device 100 further comprises one or more communication interfaces 110 coupled to the bus 102. These interfaces 110 enable the device 100 to communicate with other components. The communication interfaces 110 may include, for example, a network interface (wired or wireless) for enabling the device 100 to send messages to and receive messages from a network. The communications interfaces 110 may further include a wireless interface (e.g. Bluetooth) for communicating wirelessly with nearby devices, and a wired interface for direct coupling with

a compatible local device. Furthermore, the communications interfaces 110 may include a 3G interface for enabling the device to access the Internet without using a local network. These and other interfaces may be included in the device 100.

Sample Operation

[0020] With the above description in mind, and with reference to Figs. 1-5, the operation of device 100 in accordance with several embodiments of the present invention will now be described. In the following description, it will be assumed for the sake of illustration that the automatic scaling functionality is provided by the operating system 114. However, as noted above, this is just one possible implementation. Other implementations where the automatic scaling functionality is provided by the applications 112 themselves or by a combination of or cooperation between the operating system 114 and one or more of the applications 112 are also possible. All such implementations are within the scope of the present invention.

[0021] As mentioned above, the device 100 includes one or more distance indicating components 118. In one embodiment, a distance indicating component 118 may be one of two types of components: (1) a distance determining component such as an IR sensor, a laser sensor, a SONAR sensor, etc.; or (2) a user-facing camera. Because automatic scaling is carried out slightly differently depending upon whether component 118 is a distance determining component or a user-facing camera, the automatic scaling functionality will be described separately for each type of component. For the sake of simplicity, the following description will assume that there is only one distance indicating component 118 in the device 100. However, it should be noted that more distance indicating components 118 may be included and used if so desired.

Operation Using a Distance Determining Component

Calibration

[0022] In one embodiment, before automatic scaling is carried out using a distance determining component, a calibration procedure is performed. This calibration procedure allows the operating system 114 to tailor the automatic scaling to a user's particular preference. A flow diagram showing the calibration procedure in accordance with one embodiment of the present invention is provided in Fig. 2.

[0023] In performing the calibration procedure, the operating system 114 initially displays (block 202) a set of visual content (which in one embodiment includes both text and a graphics image) on the display 116 of device 100. The operating system 114 then prompts

(block 204) the user to hold the display 116 at a first distance from the user's face and to adjust the visual content to a size that is comfortable for the user at that distance. In one embodiment, the first distance may be the closest distance that the user would expect to have his/her face to the display 116. In response to this prompt, the user uses the user interface components 108 of device 100 to scale the visual content to a size that is comfortable for him/her at the first distance. The user may do this, for example, using keys on a keyboard, a mouse, a touch sensitive screen (e.g. by pinching or spreading two fingers), or some other input mechanism. By doing so, the user is in effect providing input indicating the scaling factor(s) that the user would like the operating system 114 to use to scale visual content at this first distance. In one embodiment, the scaling factor(s) may include a preferred font size for the text and a preferred magnification factor for the graphics image.

[0024] The operating system 114 receives (block 206) this user input. In addition, the operating system 114 receives some sensor information from the distance determining component (e.g. the IR sensor, the laser sensor, the SONAR sensor, etc.), and uses this information to determine (block 208) the current distance between the user's face and the display 116. In the case of an IR sensor, the operating system 114 receives an intensity value (indicating the intensity of the IR signal sensed by the IR sensor). Based upon this value and perhaps a table of intensity-to-distance values (not shown), the operating system 114 determines a current distance between the user's face and the display 116. In the case of a laser or SONAR sensor, the operating system 114 receives a time value (indicating how long it took for the laser or SONAR signal to bounce back from the user's face). Based upon this value and perhaps a table of timing-to-distance values (not shown), the operating system 114 determines a current distance between the user's face and the display 116. After the current distance is determined, it is stored (block 210) along with the scaling factors; thus, at this point, the operating system 114 knows the first distance and the scaling factor(s) that should be applied at that distance.

[0025] To continue the calibration procedure, the operating system 114 prompts (block 212) the user to hold the display 116 at a second distance from the user's face and to adjust the visual content to a size that is comfortable for the user at that distance. In one embodiment, the second distance may be the farthest distance that the user would expect to have his/her face from the display 116. In response to this prompt, the user uses the user interface components 108 to scale the visual content on the display to a size that is comfortable for him/her at the second distance. The user may do this in a manner similar to that described above. By doing so, the user is in effect providing input indicating the scaling

factor(s) that the user would like the operating system 114 to use to scale visual content at the second distance. Again, the scaling factor(s) may include a preferred font size for the text and a preferred magnification factor for the graphics image.

[0026] The operating system 114 receives (block 214) this user input. In addition, the operating system 114 receives some sensor information from the distance determining component, and uses this information to determine (block 216) the current distance between the user's face and the display 116. This distance determination may be performed in the manner described above. After the current distance is determined, it is stored (stored 218) along with the scaling factor(s); thus, at this point, in addition to knowing the first distance and its associated scaling factor(s), the operating system 114 also knows the second distance and its associated scaling factor(s). With these two sets of data, the operating system 114 can use interpolation to determine the scaling factor(s) that should be applied for any distance between the first and second distances.

[0027] The above calibration procedure may be used to perform calibration for both the comfort mode and the zoom mode. The difference will mainly be that the scaling factor(s) specified by the user will be different for the two modes. That is, for comfort mode, the user will specify a smaller scaling factor(s) at the first (shorter) distance than at the second (longer) distance, but for zoom mode, the user will specify a larger scaling factor(s) at the first distance than at the second distance. Other than that, the overall procedure is generally similar. In one embodiment, the calibration procedure is performed twice: once for comfort mode and once for zoom mode.

[0028] After calibration is performed, the operating system 114, in one embodiment, generates (block 220) one or more lookup tables for subsequent use. Such a lookup table may contain multiple entries, and each entry may include a distance value and an associated set of scaling factor value(s). One entry may contain the first distance and the set of scaling factor value(s) specified by the user for the first distance. Another entry may contain the second distance and the set of scaling factor value(s) specified by the user for the second distance. The lookup table may further include other entries that have distances and scaling factor value(s) that are generated based upon these two entries. For example, using linear interpolation, the operating system 114 can generate multiple entries with distance and scaling factor value(s) that are between the distances and scaling factor value(s) of the first and second distances. For example, if the first distance is A and the second distance is B, and if a first scaling factor associated with distance A is X and a second scaling factor associated

with distance B is Y, then for a distance C that is between A and B, the scaling factor can be computed using linear interpolation as follows:

$$Z = X + (Y - X) * (C - A) / (B - A)$$

[0029] where Z is the scaling factor associated with distance C.

[0030] Using this methodology, the operating system 114 can populate the lookup table with many entries, with each entry containing a distance and an associated set of scaling factor value(s). Such a lookup table may thereafter be used during regular operation to determine a scaling factor(s) for any given distance. In one embodiment, the operating system 114 generates two lookup tables: one for comfort mode and another for zoom mode. Once generated, the lookup tables are ready to be used during regular operation.

[0031] In the above example, the lookup tables are generated using linear interpolation. It should be noted that this is not required. If so desired, other types of interpolation (e.g. non-linear, exponential, geometric, etc.) may be used instead. Also, the operating system 114 may choose not to generate any lookup tables at all. Instead, the operating system 114 may calculate scaling factors on the fly. These and other alternative implementations are within the scope of the present invention.

Regular Operation

[0032] After the calibration procedure is performed, the operating system 114 is ready to implement automatic scaling during regular operation. A flow diagram illustrating regular operation in accordance with one embodiment of the present invention is shown in Fig. 3.

[0033] Initially, the operating system 114 receives a request from one of the applications 112 to provide the automatic scaling service. In one embodiment, the request specifies whether comfort mode or zoom mode is desired. In response to the request, the operating system 114 determines (block 302) a current distance between the user's face and the display 116. This may be done by receiving sensor information from the distance determining component (e.g. the IR sensor, laser sensor, SONAR sensor, etc.) and using the sensor information to determine (in the manner described previously) how far the user's face currently is from the display 116.

[0034] Based at least in part upon this current distance, the operating system 114 determines (block 304) a set of scaling factor(s). In one embodiment, the set of scaling factor(s) is determined by accessing an appropriate lookup table (e.g. the comfort mode table

or the zoom mode table) generated during the calibration process, and accessing the appropriate entry in the lookup table using the current distance as a key. In many instances, there may not be an exact match between the current distance and a distance in the table. In such a case, the operating system 114 may select the entry with the closest distance value. From that entry, the operating system 114 obtains a set of scaling factor(s). As an alternative to accessing a lookup table, the operating system 114 may calculate the set of scaling factor(s) on the fly. In one embodiment, if the current distance is shorter than the first (closest) distance determined during calibration, the operating system 114 will use the scaling factor(s) provided by the user in association with the first distance. If the current distance is longer than the second (farthest) distance determined during calibration, the operating system 114 will use the scaling factor(s) provided by the user in association with the second distance.

[0035] After the set of scaling factor(s) is determined, the operating system 114 causes (block 306) a set of visual content to be sized in accordance with the set of scaling factor(s). In one embodiment, the operating system 114 may do this by: (1) providing the set of scaling factor(s) to the calling application and having the calling application scale the visual content in accordance with the set of scaling factor(s); or (2) receiving the visual content from the calling application, and scaling the visual content for the calling application in accordance with the set of scaling factor(s). Either way, when the visual content is rendered on the display 116, it will have a scale appropriate for the current distance between the user's face and the display 116.

[0036] Thereafter, the operating system 114 periodically checks (block 308) to determine whether the distance between the user's face and the display 116 has changed. The operating system 114 may do this by periodically receiving sensor information from the distance determining component and using that information to determine a current distance between the user's face and the display 116. This current distance is compared against the distance that was used to determine the set of scaling factor(s). If the distances are different, then the operating system 114 may proceed to rescale the visual content. In one embodiment, the operating system 114 will initiate a rescaling of the visual content only if the difference in distances is greater than a certain threshold. If the difference is below the threshold, the operating system 114 will leave the scaling factor(s) the same. Implementing this threshold prevents the scaling factor(s), and hence the size of the visual content, from constantly changing in response to small changes in distance, which may be distracting and uncomfortable for the user.

[0037] In block 308, if the operating system 114 determines that the difference between the current distance and the distance that was used to determine the set of scaling factor(s) is less than the threshold, then the operating system 114 loops back and continues to check (block 308) to see if the distance between the user's face and the display 116 has changed. On the other hand, if the operating system 114 determines that the difference between the current distance and the distance that was used to determine the set of scaling factor(s) is greater than the threshold, then the operating system 114 proceeds to rescale the visual content.

[0038] In one embodiment, the operating system 114 rescales the visual content by looping back to block 304 and determining a new set of scaling factor(s) based at least in part upon the new current distance. In one embodiment, the new set of scaling factor(s) is determined by accessing the appropriate lookup table (e.g. the comfort mode table or the zoom mode table), and accessing the appropriate entry in that lookup table using the new current distance as a key. As an alternative, the operating system 114 may calculate the new set of scaling factor(s) on the fly.

[0039] After the new set of scaling factor(s) is determined, the operating system 114 causes (block 306) the visual content to be resized in accordance with the new set of scaling factor(s). In one embodiment, the operating system 114 may do this by providing the new set of scaling factor(s) to the calling application and having the calling application rescale the visual content in accordance with the new set of scaling factor(s), or by receiving the visual content from the calling application and rescaling the visual content for the calling application in accordance with the new set of scaling factor(s). Either way, when the visual content is rendered on the display 116, it will have a new scale appropriate for the new current distance between the user's face and the display 116.

[0040] After the visual content is rescaled, the operating system 114 proceeds to block 308 to once again determine whether the distance between the user's face and the display 116 has changed. If so, the operating system 114 may rescale the visual content again. In the manner described, the device 100 automatically scales the size of a set of visual content in response to the distance between a user's face and the display 116.

Operation Using a User-Facing Camera

Calibration

[0041] The above discussion describes how automatic scaling may be carried out using a distance determining component. In one embodiment, automatic scaling may also be

performed using a user-facing camera. The following discussion describes how this may be done, in accordance with one embodiment of the present invention.

[0042] In one embodiment, before automatic scaling is carried out using a user-facing camera, a calibration procedure is performed. This calibration procedure allows the operating system 114 to tailor the automatic scaling to a user's particular preference. A flow diagram showing the calibration procedure in accordance with one embodiment of the present invention is provided in Fig. 4.

[0043] In performing the calibration procedure, the operating system 114 initially displays (block 402) a set of visual content (which in one embodiment includes both text and a graphics image) on the display 116 of device 100. The operating system 114 then prompts (block 404) the user to hold the display 116 at a first distance from the user's face and to adjust the visual content to a size that is comfortable for the user at that distance. In one embodiment, the first distance may be the closest distance that the user would expect to have his/her face to the display 116. In response to this prompt, the user uses the user interface components 108 of device 100 to scale the visual content to a size that is comfortable for him/her at the first distance. The user may do this, for example, using keys on a keyboard, a mouse, a touch sensitive screen (e.g. by pinching or spreading two fingers), or some other input mechanism. By doing so, the user is in effect providing input indicating the scaling factor(s) that the user would like the operating system 114 to use to scale visual content at this first distance. In one embodiment, the scaling factor(s) may include a preferred font size for the text and a preferred magnification factor for the graphics image.

[0044] The operating system 114 receives (block 406) this user input. In addition, the operating system 114 causes the user-facing camera to capture a current image of the user's face, and receives (block 408) this captured image from the camera. Using the captured image, the operating system 114 determines (block 410) the current size or dimensions of a certain feature of the user's face. For purposes of the present invention, any feature of the user's face may be used for this purpose, including but not limited to the distance between the user's eyes, the distance from one side of the user's head to the other, etc. In the following example, it will be assumed that the distance between the user's eyes is the feature that is measured.

[0045] In one embodiment, this distance may be measured using facial recognition techniques. More specifically, the operating system 114 implements, or invokes a routine (not shown) that implements, a facial recognition technique to analyze the captured image to locate the user's eyes. The user's eyes may be found, for example, by looking for two

relatively round dark areas (the pupils) surrounded by white areas (the whites of the eyes). Facial recognition techniques capable of performing this type of operation is relatively well known (see, for example, W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face Recognition: A Literature Survey, ACM Computing Surveys, 2003, pp. 399-458, a portion of which is included herein). Once the eyes are found, the distance between the eyes (which in one embodiment is measured from the center of one pupil to the center of the other pupil) is measured. In one embodiment, this measurement may be expressed in terms of the number of pixels between the centers of the pupils. This measurement provides an indication of how far the user's face is from the display 116. That is, when the number of pixels between the user's eyes is this value, the user's face is at the first distance from the display 116.

[0046] After the number of pixels between the user's eyes is measured, it is stored (block 412) along with the scaling factors; thus, at this point, the operating system 114 knows the number of pixels between the user's eyes when the user's face is at the first distance, and it knows the scaling factor(s) that should be applied when the number of pixels between the user's eyes is at this value.

[0047] To continue the calibration procedure, the operating system 114 prompts (block 414) the user to hold the display 116 at a second distance from the user's face and to adjust the visual content to a size that is comfortable for the user at that distance. In one embodiment, the second distance may be the farthest distance that the user would expect to have his/her face from the display 116. In response to this prompt, the user uses the user interface components 108 to scale the visual content on the display to a size that is comfortable for him/her at the second distance. The user may do this in a manner similar to that described above. By doing so, the user is in effect providing input indicating the scaling factor(s) that the user would like the operating system 114 to use to scale visual content at the second distance. Again, the scaling factor(s) may include a preferred font size for the text and a preferred magnification factor for the graphics image.

[0048] The operating system 114 receives (block 416) this user input. In addition, the operating system 114 causes the user-facing camera to capture a second image of the user's face, and receives (block 418) this captured image from the camera. Using the second captured image, the operating system 114 determines (block 420) the number of pixels between the user's eyes when the user's face is at the second distance from the display 116. This may be done in the manner described above. Since, in the second image, the user's face is farther from the display 116, the number of pixels between the user's eyes in the second image should be fewer than in the first image. After the number of pixels between the user's

eyes is determined, it is stored (stored 422) along with the scaling factor(s). Thus, at this point, the operating system 114 has two sets of data: (1) a first set that includes the number of pixels between the user's eyes at the first distance and the scaling factor(s) to be applied at the first distance; and (2) a second set that includes the number of pixels between the user's eyes at the second distance and the scaling factor(s) to be applied at the second distance. With these two sets of data, the operating system 114 can use interpolation to determine the scaling factor(s) that should be applied for any distance between the first and second distances. For the sake of convenience, the number of pixels between the user's eyes at the first distance will be referred to below as the "first number of pixels", and the number of pixels between the user's eyes at the second distance will be referred to below as the "second number of pixels".

[0049] The above calibration procedure may be used to perform calibration for both the comfort mode and the zoom mode. The difference will mainly be that the scaling factor(s) specified by the user will be different for the two modes. That is, for comfort mode, the user will specify a smaller scaling factor(s) at the first (shorter) distance than at the second (longer) distance, but for zoom mode, the user will specify a larger scaling factor(s) at the first distance than at the second distance. Other than that, the overall procedure is generally similar. In one embodiment, the calibration procedure is performed twice: once for comfort mode and once for zoom mode.

[0050] After calibration is performed, the operating system 114, in one embodiment, generates (block 424) one or more lookup tables for subsequent use. Such a lookup table may contain multiple entries, and each entry may include a "number of pixels" value and an associated set of scaling factor(s) value(s). One entry may contain the "first number of pixels" and the set of scaling factor value(s) specified by the user for the first distance. Another entry may contain the "second number of pixels" and the set of scaling factor value(s) specified by the user for the second distance. The lookup table may further include other entries that have "number of pixels" values and scaling factor value(s) that are generated based upon these two entries. For example, using linear interpolation, the operating system 114 can generate multiple entries with "number of pixels" values that are between the "first number of pixels" and the "second number of pixels" and scaling factor value(s) that are between the first and second sets of associated scaling factor values(s). For example, if the "first number of pixels" is A and the "second number of pixels" is B, and if a first scaling factor associated with the first distance is X and a second scaling factor

associated with the second distance is Y, then for a "number of pixels" C that is between A and B, the scaling factor can be computed using linear interpolation as follows:

$$Z = X + (Y - X) * (C - A) / (B - A)$$

[0051] where Z is the scaling factor associated with the "number of pixels" C.

[0052] Using this methodology, the operating system 114 can populate the lookup table with many entries, with each entry containing a "number of pixels" value (which provides an indication of how far the user's face is from the display 116) and an associated set of scaling factor value(s). Such a lookup table may thereafter be used during regular operation to determine a scaling factor(s) for any given "number of pixels" value. In one embodiment, the operating system 114 generates two lookup tables: one for comfort mode and another for zoom mode. Once generated, the lookup tables are ready to be used during regular operation.

[0053] In the above example, the lookup tables are generated using linear interpolation. It should be noted that this is not required. If so desired, other types of interpolation (e.g. non-linear, exponential, geometric, etc.) may be used instead. Also, the operating system 114 may choose not to generate any lookup tables at all. Instead, the operating system 114 may calculate scaling factors on the fly. These and other alternative implementations are within the scope of the present invention.

Regular Operation

[0054] After the calibration procedure is performed, the operating system 114 is ready to implement automatic scaling during regular operation. A flow diagram illustrating regular operation in accordance with one embodiment of the present invention is shown in Fig. 5.

[0055] Initially, the operating system 114 receives a request from one of the applications 112 to provide the automatic scaling service. In one embodiment, the request specifies whether comfort mode or zoom mode is desired. In response to the request, the operating system 114 determines (block 502) a current size of a facial feature of the user. In one embodiment, this entails measuring the number of pixels between the eyes of the user. This may be done by causing the user-facing camera to capture a current image of the user, and receiving this captured image from the camera. Using the captured image, the operating system 114 measures (in the manner described above) how many pixels are between the pupils of the user's eyes. This current "number of pixels" value provides an indication of how far the user's face currently is from the display 116.

[0056] Based at least in part upon this current "number of pixels" value, the operating system 114 determines (block 504) a set of scaling factor(s). In one embodiment, the set of scaling factor(s) is determined by accessing an appropriate lookup table (e.g. the comfort mode table or the zoom mode table) generated during the calibration process, and accessing the appropriate entry in the lookup table using the current "number of pixels" value as a key. In many instances, there may not be an exact match between the current "number of pixels" value and a "number of pixels" value in the table. In such a case, the operating system 114 may select the entry with the closest "number of pixels" value. From that entry, the operating system 114 obtains a set of scaling factor(s). As an alternative to accessing a lookup table, the operating system 114 may calculate the set of scaling factor(s) on the fly. In one embodiment, if the current "number of pixels" value is smaller than the "first number of pixels" determined during calibration, the operating system 114 will use the scaling factor(s) associated with the "first number of pixels". If the current "number of pixels" value is larger than the "second number of pixels" determined during calibration, the operating system 114 will use the scaling factor(s) associated with the "second number of pixels".

[0057] After the set of scaling factor(s) is determined, the operating system 114 causes (block 506) a set of visual content to be sized in accordance with the set of scaling factor(s). In one embodiment, the operating system 114 may do this by: (1) providing the set of scaling factor(s) to the calling application and having the calling application scale the visual content in accordance with the set of scaling factor(s); or (2) receiving the visual content from the calling application, and scaling the visual content for the calling application in accordance with the set of scaling factor(s). Either way, when the visual content is rendered on the display 116, it will have a scale appropriate for the current number of pixels between the user's eyes (and hence, for the current distance between the user's face and the display 116).

[0058] Thereafter, the operating system 114 periodically checks (block 508) to determine whether the number of pixels between the user's eyes has changed. The operating system 114 may do this by periodically receiving captured images of the user's face from the user-facing camera, and measuring the current number of pixels between the user's eyes. This current number of pixels is compared against the number of pixels that was used to determine the set of scaling factor(s). If the numbers of pixels are different, then the operating system 114 may proceed to rescale the visual content. In one embodiment, the operating system 114 will initiate a rescaling of the visual content only if the difference in numbers of pixels is greater than a certain threshold. If the difference is below the threshold, the operating system 114 will leave the scaling factor(s) the same. Implementing this threshold prevents the scaling

factor(s), and hence the size of the visual content, from constantly changing in response to small changes in the numbers of pixels, which may be distracting and uncomfortable for the user.

[0059] In block 508, if the operating system 114 determines that the difference between the current number of pixels and the number of pixels that was used to determine the set of scaling factor(s) is less than the threshold, the operating system 114 loops back and continues to check (block 508) to see if the number of pixels between the user's eyes has changed. On the other hand, if the operating system 114 determines that the difference between the current number of pixels and the number of pixels that was used to determine the set of scaling factor(s) is greater than the threshold, then the operating system 114 proceeds to rescale the visual content.

[0060] In one embodiment, the operating system 114 rescales the visual content by looping back to block 504 and determining a new set of scaling factor(s) based at least in part upon the new current number of pixels. In one embodiment, the new set of scaling factor(s) is determined by accessing the appropriate lookup table (e.g. the comfort mode table or the zoom mode table), and accessing the appropriate entry in that lookup table using the new current number of pixels as a key. As an alternative, the operating system 114 may calculate the new set of scaling factor(s) on the fly.

[0061] After the new set of scaling factor(s) is determined, the operating system 114 causes (block 506) the visual content to be resized in accordance with the new set of scaling factor(s). In one embodiment, the operating system 114 may do this by providing the new set of scaling factor(s) to the calling application and having the calling application rescale the visual content in accordance with the new set of scaling factor(s), or by receiving the visual content from the calling application and rescaling the visual content for the calling application in accordance with the new set of scaling factor(s). Either way, when the visual content is rendered on the display 116, it will have a new scale appropriate for the new current number of pixels between the user's eyes (and hence, appropriate for the current distance between the user's face and the display 116).

[0062] After the visual content is rescaled, the operating system 114 proceeds to block 508 to once again determine whether the distance between the user's eyes has changed. If so, the operating system 114 may rescale the visual content again. In the manner described, the device 100 automatically scales the size of a set of visual content in response to how close a user's face is to a display.

[0063] In the foregoing specification, embodiments of the present invention have been described with reference to numerous specific details that may vary from implementation to implementation. Thus, the sole and exclusive indicator of what is the invention, and is intended by the Applicants to be the invention, is the set of claims that issue from this application, in the specific form in which such claims issue, including any subsequent correction. Any definitions expressly set forth herein for terms contained in such claims shall govern the meaning of such terms as used in the claims. Hence, no limitation, element, property, feature, advantage or attribute that is not expressly recited in a claim should limit the scope of such claim in any way. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.

Face Recognition: A Literature Survey¹

W. Zhao²

Sarnoff Corporation

R. Chellappa and A. Rosenfeld³

University of Maryland

P.J. Phillips⁴

National Institute of Standards and Technology

Abstract

As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention, especially during the past several years. This is evidenced by the emergence of face recognition conferences such as AFGR [1] and AVBPA [2], and systematic empirical evaluations of face recognition techniques, including the FERET [3, 4, 5, 6] and XM2VTS [7] protocols. There are at least two reasons for this trend; the first is the wide range of commercial and law enforcement applications, and the second is the availability of feasible technologies after 30 years of research. This paper provides an up-to-date critical survey of still- and video-based face recognition research.

¹The support of the Office of Naval Research under Grants N00014-95-1-0521 and N00014-00-1-0908 is gratefully acknowledged.

²Vision Technologies Lab, Sarnoff Corporation, Princeton, NJ 08543-5300.

³Center for Automation Research, University of Maryland, College Park, MD 20742-3275.

⁴National Institute of Standards and Technology, Gaithersburg, MD 20899.

1 Introduction

As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention, especially during the past few years. This is evidenced by the emergence of face recognition conferences such as AFGR [1] and AVBPA [2], and systematic empirical evaluations of face recognition techniques (FRT), including the FERET [3, 4, 5, 6] and XM2VTS [7] protocols. There are at least two reasons for this trend; the first is the wide range of commercial and law enforcement applications and the second is the availability of feasible technologies after 30 years of research.

The strong need for user-friendly systems that can secure our assets and protect our privacy without losing our identity in a sea of numbers is obvious. At present, one needs a PIN to get cash from an ATM, a password for a computer, a dozen others to access the internet, and so on. Although extremely reliable methods of biometric personal identification exist, e.g., fingerprint analysis and retinal or iris scans, these methods rely on the cooperation of the participants, whereas a personal identification system based on analysis of frontal or profile images of the face is often effective without the participant's cooperation or knowledge. The advantages/disadvantages of different biometrics are described in [8]. Table 1 lists some of the applications of face recognition.

Areas	Specific Applications
Biometrics	Drivers' Licenses, Entitlement Programs
	Immigration, National ID, Passports, Voter Registration
	Welfare Fraud
Information Security	Desktop Logon (Windows NT, Windows 95)
	Application Security, Database Security, File Encryption
	Intranet Security, Internet Access, Medical Records
	Secure Trading Terminals
Law Enforcement and Surveillance	Advanced Video Surveillance, CCTV Control
	Portal Control, Post-Event Analysis
	Shoplifting and Suspect Tracking and Investigation
Smart Cards	Stored Value Security, User Authentication
Access Control	Facility Access, Vehicular Access

Table 1: Typical applications of face recognition.

A general statement of the problem can be formulated as follows: Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. Available collateral information such as race, age, gender, facial expression and speech may be used in narrowing the search (enhancing recognition). The solution to the problem involves segmentation of faces (face detection) from cluttered scenes, feature extraction from the face region, recognition or verification. In identification problems, the input to the system is an unknown face, and the system reports back the determined identity from a database of known individuals, whereas in verification problems, the system needs to confirm or reject the claimed identity of the input face.

Commercial and law enforcement applications of FRT range from static, controlled format photographs to uncontrolled video images, posing a wide range of different technical challenges and requiring an equally wide range of techniques from image processing, analysis, understanding and pattern recognition. One can broadly classify the challenges and techniques into two groups: static and dynamic/video matching. Within these groups, significant differences exist, depending on the specific application. The differences are in terms of image quality, amount of background clutter (posing challenges to segmentation algorithms), availability of a well-defined matching criterion, and the nature, type and amount of input from a user. In some applications, such as computerized aging, one is only concerned with defining a set of transformations so that the images created by the system are similar to what humans expect based on their recollections.

In 1995, a review paper by Chellappa et al. [9] gave a thorough survey of FRT at that time. (An earlier survey [10] appeared in 1992.) At that time, video-based face recognition was still in a nascent stage. During the past five years, face recognition has received increased attention and has advanced technically. Many commercial systems using face recognition are now available. Significant research efforts have been focused on video-based face modeling, processing and recognition. It is not an overstatement to say that face recognition has become one of the most successful applications of pattern recognition, image analysis and understanding.

In this paper we provide a critical review of the most recent developments in face recognition. This paper is organized as follows: In Section 2 we briefly review issues that are relevant from the psychophysical point of view. Section 3 provides a detailed review of recent developments in face recognition techniques using grayscale, range and other images. In Section 4 face recognition techniques based on video are reviewed, including face tracking, modeling, and non-face/face based recognition. Data collection and performance evaluation of face recognition algorithms are addressed in Section 5 with detailed descriptions of two representative protocols: FERET and XM2VTS. Finally, in Section 6 we discuss two difficult technical problems common to all the algorithms: lack of robustness to illumination and pose variations, and suggest possible ways to overcome these limitations.

2 Psychophysics/Neuroscience Issues Relevant to Face Recognition

In general, the human face recognition system utilizes a broad spectrum of stimuli, obtained from many, if not all, of the senses (visual, auditory, olfactory, tactile, etc.). These stimuli are used either individually or collectively for storage and retrieval of face images. In many cases contextual knowledge is also used, i.e. the surroundings play an important role in recognizing faces in relation to where they are supposed to be located. It is futile (using existing technology) to even attempt to develop a system that can mimic all these remarkable capabilities of humans. However, the human brain has its limitations in the total number of persons that it can accurately "remember". A key potential advantage of a computer system is its capacity to handle large datasets of face images. In most applications the images are single or multiple views of 2-D intensity data, which forces the inputs to computer algorithms to be visual only. For this reason, the literature reviewed in this section is related to aspects of human visual perception.

Many studies and findings in psychology and neuroscience have direct relevance to engineers interested in designing algorithms or systems for machine recognition of faces. On the other hand, better machine systems can provide better tools for conducting studies in psychology and neuroscience [11]. For example, a possible engineering explanation of the lighting effect illustrated in [12] is as follows: for familiar faces a 3D model is usually built in memory; when the actual lighting direction is opposite to the usually assumed direction, a shape-from-shading algorithm recovers incorrect structural information and hence makes recognition of faces harder.

A complete review of relevant studies in psychophysics and neuroscience is beyond the scope of this paper. We only summarize findings that are potentially relevant to the design of face recognition systems. For details the reader is referred to the papers cited below. The issues that are of potential interest to designers are:

- **Is face recognition a dedicated process?** [13, 14]: Evidence for the existence of a dedicated face processing system comes from three sources [13]. A) Faces are more easily remembered by humans than other objects when presented in an upright orientation. B) Prosopagnosia patients are unable to recognize previously familiar faces, but usually have no other profound agnosia. They recognize people by their voices, hair color, dress, etc. Although they can perceive eyes, nose, mouth, hair, etc., they are unable to put these features together for the purpose of identification. It should be noted that prosopagnosia patients recognize whether the given object is a face or not, but then have difficulty in identifying the face. C) It is argued that infants come into the world prewired to be attracted by faces. Neonates seem to prefer to look at moving stimuli that have face-like patterns in preference to those containing no patterns or jumbled facial features. Some recent studies on this subject further confirm that face recognition is a dedicated process which is different from general object recognition [14]. Seven differences between face recognition and object recognition can be listed based on empirical results: 1) Configural effects (related to the choice of different types of machine recognition systems), 2) expertise, 3) differences verbalizable, 4) sensitivity to contrast polarity and illumination direction (related to the illumination problem in machine recognition systems), 5) metric variation, 6) rotation in depth (related to the pose variation problem in machine recognition systems), and 7) rotation in plane/inverted face.
- **Is face perception the result of wholistic or feature analysis?** [15] Both wholistic and feature information are crucial for the perception and recognition of faces. Studies suggest the possibility of global descriptions serving as a front end for finer, feature-based perception. If dominant features are present, wholistic descriptions may not be used. For example, in face recall studies, humans quickly focus on odd features such as big ears, a crooked nose, a staring eye, etc. One of the strongest pieces of evidence to support the view that face recognition involves more configural/wholistic processing than other object recognition tasks has been the face inversion effect, where an inverted face is much harder to recognize than a normal face. An excellent example is given in [16] using the "Thatcher illusion" [17]. In this illusion, the eyes and mouth of a face are inverted. The result looks grotesque

in an upright face; however, when shown inverted, the face looks fairly normal, and the inversion of the features is not readily noticed.

- **Ranking of significance of facial features:** Hair, face outline, eyes and mouth (not necessarily in that order) have been determined to be important for perceiving and remembering faces. Several studies have shown that the nose plays an insignificant role. In face recognition using profiles (which may be important in mugshot matching applications, where profiles can be extracted from side views), several fiducial points (“features”) are in or near the nose region. Another outcome of some of the studies is that both external and internal features are important in the recognition of previously presented but otherwise unfamiliar faces, and internal features are more dominant in the recognition of familiar faces. It has also been found that the upper part of the face is more useful for face recognition than the lower part. The role of aesthetic attributes such as beauty, attractiveness and/or pleasantness has also been studied, with the conclusion that the more attractive the faces are, the better is their recognition rate; the least attractive faces come next, followed by the mid-range faces, in terms of ease of being recognized.
- **Caricatures [18]:** Perkins [19] formally defines a caricature as “a symbol that exaggerates measurements relative to any measure which varies from one person to another”. Thus the length of a nose is a measure that varies from person to person, and may be useful as a symbol in caricaturing someone, but not the number of ears. Caricatures do not contain as much information as photographs, but they manage to capture the important characteristics of a face; experiments comparing the usefulness of caricatures and line drawings decidedly favor the former.
- **Distinctiveness:** Studies show that distinctive faces are better retained in memory and are recognized better and faster than typical faces. However, if a decision has to be made as to whether an object is a face or not, it takes longer to recognize an atypical face than a typical face. This may be explained by different mechanisms being used for detection and identification.
- **The role of spatial frequency analysis:** Earlier studies [20, 21] concluded that information in low spatial frequency bands plays a dominant role in face recognition. Later studies [22] showed that, depending on the recognition task, the low-, bandpass and high-frequency components may play different roles. For example the sex judgment task can be successfully accomplished using low-frequency components only, while the identification task requires the use of high-frequency components. The low-frequency components contribute to the global description, while the high-frequency components contribute to the finer details required in the identification task.
- **Viewpoint-invariant recognition?[23, 24]:** Much work in visual object recognition (e.g., [24]) has been cast within a theoretical framework introduced by Marr [25] in which different views of objects are analyzed in a way which allows access to (largely) viewpoint-invariant descriptions. Recently, there has been some debate

about whether object recognition is viewpoint-invariant. In face recognition it seems clear that memory is highly viewpoint-dependent. Hill et al. [26] show that generalization even from one profile viewpoint to another is poor, though generalization from one 3/4 view to the other is very good.

- **Effect of lighting change**[12, 15, 27]: It has long been informally observed that photographic negatives of faces are difficult to recognize. However, relatively little work has explored why it is so difficult to recognize negative images of faces. In [12], experiments were conducted to explore whether difficulties with negative images of faces, and inverted images of faces, arise because each of these manipulations reverses the apparent direction of lighting, rendering a top-lit image of a face as if lit from below. This work demonstrated that bottom lighting does indeed make it harder to identify familiar faces. In [27], the importance of top lighting for face recognition, using the task of matching surface images of faces for identity, is demonstrated.
- **Movement and face recognition**[15, 28]: A recent intriguing study [28] shows that famous faces are easier to recognize when shown in moving sequences than in still photographs. This observation has been extended to show that movement helps in the recognition of familiar faces under a range of different types of degradations — negated, inverted, or thresholded (shown as black-and-white images) [15]. Even more interesting is that movement seems to provide a benefit even if the information content is equated in dynamic and static conditions. On the other hand, experiments with unfamiliar faces suggest no additional benefit from viewing animated rather than static sequences.
- **Facial expression**[29]: Based on neurophysiological studies, it seems that analysis of facial expressions is accomplished in parallel to face recognition. Some prosopagnosic patients, who have difficulties in identifying familiar faces, nevertheless seem to recognize facial expressions due to emotions. Patients who suffer from “organic brain syndrome” do poorly at expression analysis but perform face recognition quite well. Normal humans also exhibit parallel capabilities for facial expression analysis and face recognition. Similarly, separation of face recognition and “focused visual processing” tasks (look for someone with a thick mustache) has been claimed.

3 Face Recognition from Single Intensity or Other Images

In this section we survey the state of the art in face recognition in the engineering literature. Extraction of features such as the eyes and mouth, and face segmentation/detection are reviewed in Section 3.1. Sections 3.2 and 3.3 are detailed reviews of recent work in face recognition, including statistical and neural approaches.

3.1 Segmentation/detection and feature extraction

3.1.1 Segmentation/detection

Up to the middle 90's, most of the work in this area was focused on single-face segmentation from a simple or complex background. The approaches included using a face template, a deformable feature-based template, skin color, and a neural network. During the past five years, more reliable face detection methods have been developed to cope with multiple face detection in a complex background, where the face images may be partly occluded, rotated in plane, or rotated in depth. For technical details, please refer to [30, 31, 32, 33, 34, 35, 36, 37, 38, 39]. Some of these methods were tested on relatively large databases, e.g. [30, 38]. A recent survey paper on face detection is [40]. Here we review two well-known approaches: The neural network approach of Kanade et al. [38, 39] and the example-based learning approach of Sung and Poggio [30]. A recent approach using a Support Vector Machine (SVM) is also briefly reviewed [37].

In [30], an example-based learning approach to locating vertical frontal views of human faces in complex scenes is presented. This technique models the distribution of human face patterns by means of a few view-based "face" and "non-face" prototype clusters. At each image location, a difference feature vector is computed between the local image pattern and the distribution-based model. This difference vector is then fed into a trained classifier to determine whether or not a human face is present at the current image location. The system detects faces of different sizes by exhaustively scanning an image for face-like local image patterns at all possible scales. More specifically, the system performs the following steps:

1. The input sub-images are all rescaled to size 19×19 , and a mask is applied to eliminate near-boundary pixels. Normalization in intensity is done by first subtracting a best-fit brightness plane from the un-masked widow pixels and then applying histogram equalization.
2. A distribution-based model of canonical face- and non-face-patterns is constructed from samples. The model consists of 12 multi-dimensional Gaussian clusters; six of them represent face- and six represent non-face-pattern prototypes. The clusters are constructed by an elliptical k -means clustering algorithm which uses an adaptively varying normalized Mahalanobis distance metric.
3. A vector of matching measurements is computed for each pattern. This is a vector of distances between the test window pattern and the canonical face model's 12 cluster centroids. Two metrics are used; one is a Mahalanobis-like distance defined on the subspace spanned by the 75 largest eigenvectors of the prototype cluster, and the other is Euclidean distance.
4. A MLP classifier is trained for face/non-face discrimination using the 24-dimensional matching measurement vectors. The training set consists of 47316 measurement vectors, 4150 of which are examples of face patterns.

To detect faces in an image, preprocessing is done as in step 1, followed by matching measurement computation (step 3), and finally the MLP is used for detection. Results

are reported on two large databases; the detection rate varied from 79.9% to 96.3% with a small number of false positives.

In [38], face knowledge is incorporated into a retinally connected neural network. The neural network uses image windows of size 20×20 , and has one hidden layer with 26 units, where 4 units cover 10×10 non-overlapping subregions, 16 units cover 5×5 subregions, and 6 units cover 20×5 overlapping horizontal stripes. The image windows are preprocessed as described in step 1 above. To deal with overlapping detections, two heuristics are used: 1) "thresholding", where the classification of a face depends on the number of detections in a neighborhood, 2) "overlap elimination", where when a region is classified as a face, overlapping detections are rejected.

To further improve system performance, multiple neural networks are trained and their outputs are combined using an arbitrary strategy including ANDing, ORing, voting, or a separate arbitration neural network. A detection rate on a dataset of 130 test images varying from 77.9% to 90.3%, with an acceptable number of false positives, was reported. To handle faces at different angles, in [39] the authors propose using a router neural net to detect the angles of the faces. After angle detection, the virtual face detection system can be applied. The router neural network is a fully connected MLP with one hidden layer and 36 output units (each unit represents 10°).

In [37], a face detection scheme based on SVMs is proposed. SVM is a learning technique developed by Vapnik et al. at AT&T [41]. It can be viewed as a way to train polynomial, neural network, or Radial Basis Function classifiers. While most of the techniques used to train these classifiers are based on the idea of minimizing the training error, the *empirical risk*, SVMs operate on another induction principle, called *structural risk minimization*, which minimizes the upper bound of the generalization error. From an implementation point of view, training an SVM is equivalent to solving a linearly constrained Quadratic Programming (QP) problem. The challenge in applying SVMs to face detection is the complexity of solving a large scale QP problem. The authors propose using a decomposition algorithm to replace the original problem with a sequence of smaller problems. Their system is very similar to that in [30] except that no matching measurements are computed and the classifier is a SVM. The authors reported comparable results on two databases.

3.1.2 Feature Extraction

Feature extraction is the key to both face segmentation and recognition, as it is to any pattern classification task. For a comprehensive review of this subject see [9]. Here we review only a few representative techniques.

There has been renewed interest in the use of the Karhunen-Loeve (KL) expansion for the representation [42, 43] and recognition [44, 45] of faces. [42] considered the problem of KL representation of cropped face images. Noting that the number of images M usually available for computing the covariance matrix of the data is much less than the row or column dimensionality of the covariance matrix, leading to singularity of the matrix, a standard method from linear algebra [46] is used that calculates only the M eigenvectors that do not belong to the null space of the degenerate matrix. Once the eigenvectors (referred to as eigenpictures) are obtained, any image in the ensemble can

be approximately reconstructed using a weighted combination of eigenpictures. By using an increasing number of eigenpictures, one gets an improved approximation to the given image. Examples of approximating an arbitrary image (not included in the calculation of the eigenvectors) by the eigenpictures are also given.

A generalized symmetry operator is used in [47] to find the eyes and mouth in a face. The motivation stems from the almost symmetric nature of the face about a vertical line through the nose. Subsequent symmetries lie within features such as the eyes, nose and mouth. The symmetry operator locates points in the image corresponding to high values of a symmetry measure discussed in detail in [47]. The procedure is claimed to be superior to other correlation-based schemes such as that of [48] in the sense that it is independent of scale or orientation. However, since no a priori knowledge of face location is used, the search for symmetry points is computationally intensive. A success rate of 95% is reported on a face image database, with the constraint that the faces occupy between 15-60% of the image.

A statistically motivated approach to detecting and recognizing the human eye in an intensity image with a frontal face is described in [49], which uses a template-based approach to detect the eyes in an image. The template has two regions of uniform intensity; the first is the iris region and the other is the white region of the eye. The approach constructs an "archetypal" eye and models various distributions as variations of it. For the "ideal" eye a uniform intensity is chosen for both the iris and whites. In an actual eye discrepancies from this ideal are present; these discrepancies can be modeled as "noise" components added to the ideal image. An α -trimmed distribution is used for both the iris and the white, and the amount of degradation, which determines the value of α , is estimated. α is easily optimized since the percentage of trimming and the area of the trimmed template are in 1-1 correspondence. A "blob" detection system is developed to locate the intensity valley caused by the iris enclosed by the white. In the experiments three sets of data were used. One consisted of 25 images used as a testing set, another had 107 positive eyes, and the third consisted of images with most probably erroneous locations which could be chosen as candidate templates. For locating the valleys, as many as 60 false alarms for the first data set, 30 for the second, and 110 for the third were reported. A tabular representation of results for three sets of values for the α 's is presented. An increase in the hit rate is reported when using an α -trimmed distribution. The overall best hit rate reported was 80%.

[50] proposes an edge-based approach to accurately detecting two-dimensional shapes including faces. The motivations for proposing such a shape detection scheme are the following observations: 1) many two-dimensional shapes including faces can be well approximated by straight lines and rectangles, and 2) in practice it is more difficult to model the intensity values of an object and its background than to exploit the intensity differential along the object's boundary. Rather than looking for a shape from an edge map, edges are extracted directly from an image according to a given shape description. This approach is said to offer several advantages over previous methods of collecting edges into global shape description such as grouping and fitting. For example, it provides a tool for systematic analysis of edge-based shape detection. The computational complexity of this approach can be alleviated using multi-resolution processing.

To demonstrate the effectiveness of the proposed approach, results of face and facial



Figure 1: Face detection and facial feature detection in a group photo

feature detection are presented. One of these results is shown in Fig. 1 where the algorithm was applied to a group photo. For the detection of facial features, a small set of operators was designed. To limit the search space, the face center region is estimated using an ellipse-shaped operator, and is marked by a white dotted ellipse having the matched ellipse size. The face region detection is biased because only simple ellipses were fitted to the faces. Iris and eyelid detections are marked.

[51] presents a method of extracting pertinent feature points from a face image. It employs Gabor wavelet decomposition and local scale interaction to extract features at curvature maxima in the image. These feature points are then stored in a data base and subsequent target face images are matched using a graph matching technique. The 2-D Gabor function used and its Fourier transform are

$$g(x, y : u_0, v_0) = \exp(-[x^2/2\sigma_x^2 + y^2/2\sigma_y^2] + 2\pi i[u_0x + v_0y]) \quad (1)$$

$$G(u, v) = \exp(-2\pi^2(\sigma_x^2(u - u_0)^2 + \sigma_y^2(v - v_0)^2)) \quad (2)$$

where σ_x and σ_y represent the spatial widths of the Gaussian and (u_0, v_0) is the frequency of the complex sinusoid.

The Gabor functions form a complete, though non-orthogonal, basis set. As with Fourier series, a function $g(x, y)$ can easily be expanded using the Gabor functions:

$$\Phi_\lambda(x, y, \theta) = \exp[(-\lambda^2(x'^2 + y'^2)) + i\pi x'] \quad (3)$$

$$x' = x \cos \theta + y \sin \theta \quad (4)$$

$$y' = -x \sin \theta + y \cos \theta \quad (5)$$

where θ is the preferred spatial orientation and λ is the aspect ratio of the Gaussian.

The feature detection process uses a simple mechanism to model end-inhibition. It uses interscale interaction to group the responses of cells from different frequency channels. This results in the generation of the end-stop regions. The orientation parameter θ

determines the direction of the edges. Hypercomplex cells are sensitive to oriented lines and step edges of short lengths, and their response decreases if the lengths are increased. They can be modeled by

$$I_{m,n}(x, y) = \max_{\theta} g(\|W_m(x, y, \theta) - \gamma W_n(x, y, \theta)\|) \quad (6)$$

and

$$W_j(x, y, \theta) = f \otimes \Phi(\alpha^j x, \alpha^j y, \theta), \quad j = \{0, -1, -2, \dots\} \quad (7)$$

where f represents the input image, g is a sigmoid non-linearity, γ is a normalizing factor, and $n > m$. The final step is to localize the features at the local maxima of the feature responses.

Recently, the issue of feature detection accuracy has been addressed. In many systems, good recognition results are dependent on accurate feature (eyes, mouth) registration; and performance degradation is observed if the feature locations are not determined accurately enough [52]. [53] describes a robust and accurate feature localization method. In this method, images are pairwise registered using a robust form of correlation. The registration process is treated as an optimization problem in a search space defined by the set of all possible geometric and photometric transformations. At each point of the search space, a score function is evaluated and the optimum of this function is localized using a combined gradient-based and stochastic optimization technique. To meet real-time requirements and ensure high registration accuracy, a multiresolution scheme is used in both the image and parameter domains. After global registration, feature selection is based on minimizing the intra-class variance and at the same time maximizing the inter-class variance. Good results were obtained in experiments on a database (the extended M2VTS database [54]) containing 295 subjects.

What is claimed is:

1. A method comprising:
causing a set of visual content on a display to be sized according to a first scaling factor, wherein a user's face is currently at a first distance from the display;
determining that the user's face has moved relative to the display such that the user's face is no longer at the first distance from the display; and
in response to determining that the user's face has moved relative to the display, causing the set of visual content on the display to be sized according to a second and different scaling factor to cause a display size of the set of visual content to change.
2. The method of claim 1, wherein determining that the user's face has moved relative to the display comprises:
determining whether the user's face has moved closer to or farther from the display.
3. The method of claim 2, wherein causing the set of visual content to be sized according to a second scaling factor comprises:
in response to determining that the user's face has moved closer to the display, causing the set of visual content to be scaled to a second scaling factor that causes the display size of the set of visual content to be reduced; and
in response to determining that the user's face has moved farther from the display, causing the set of visual content to be scaled to a second scaling factor that causes the display size of the set of visual content to be enlarged.
4. The method of claim 2, wherein causing the set of visual content to be sized according to a second scaling factor comprises:
in response to determining that the user's face has moved closer to the display, causing the set of visual content to be scaled to a second scaling factor that causes the display size of the set of visual content to be enlarged; and
in response to determining that the user's face has moved farther from the display, causing the set of visual content to be scaled to a second scaling factor that causes the display size of the set of visual content to be reduced.

5. The method of claim 1, wherein the visual content includes text, and wherein the first and second scaling factors represent different font sizes.

6. The method of claim 1, wherein the visual content includes a graphic, and wherein the first and second scaling factors represent different magnification factors.

7. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 1.

8. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 2.

9. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 3.

10. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 4.

11. An apparatus, comprising:
one or more processors; and
one or more storages having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform the operations of:
causing a set of visual content on a display to be sized according to a first scaling factor, wherein a user's face is currently at a first distance from the display;
determining that the user's face has moved relative to the display such that the user's face is no longer at the first distance from the display; and
in response to determining that the user's face has moved relative to the display, causing the set of visual content on the display to be sized according to a second and different scaling factor to cause a display size of the set of visual content to change.

12. A method, comprising:
determining that a user's face is at a first distance from a display;
determining, based at least in part upon the first distance, a first scaling factor;
causing a set of visual content on the display to be sized according to the first scaling factor;
determining that the user's face has moved to a second distance from the display, wherein the second distance is different from the first distance;
determining, based at least in part upon the second distance, a second scaling factor, wherein the second scaling factor is different from the first scaling factor; and
causing the set of visual content to be sized according to the second scaling factor to cause a display size of the set of visual content to change.

13. The method of claim 12, further comprising:
performing a calibration procedure, wherein the calibration procedure comprises:
receiving input from the user indicating a first desired scaling factor when the user's face is at a first calibration distance from the display; and
receiving input from the user indicating a second desired scaling factor when the user's face is at a second and different calibration distance from the display.

14. The method of claim 12, wherein:
determining that a user's face is at a first distance from a display comprises:
receiving information from a distance indicating component indicating that the user's face is at the first distance from the display; and
determining that the user's face has moved to a second distance from the display comprises:
receiving information from the distance indicating component indicating that the user's face is at the second distance from the display.

15. The method of claim 12, wherein:
determining that a user's face is at a first distance from a display comprises:
receiving a first set of sensor information from a sensing device; and

using the first set of sensor information to determine that the user's face is at the first distance from the display; and
determining that the user's face has moved to a second distance from the display comprises:
receiving a second set of sensor information from the sensing device; and
using the second set of sensor information to determine that the user's face is at the second distance from the display.

16. The method of claim 15, wherein the sensing device is one of: an infrared distance sensing device; a laser distance sensing device; a SONAR distance sensing device; and an image capture device for capturing an image of the user's face.

17. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 12.

18. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 13.

19. An apparatus, comprising:
one or more processors; and
one or more storages having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform the operations of:
determining that a user's face is at a first distance from a display;
determining, based at least in part upon the first distance, a first scaling factor;
causing a set of visual content on the display to be sized according to the first scaling factor;
determining that the user's face has moved to a second distance from the display,
wherein the second distance is different from the first distance;
determining, based at least in part upon the second distance, a second scaling factor,
wherein the second scaling factor is different from the first scaling factor; and
causing the set of visual content to be sized according to the second scaling factor to cause a display size of the set of visual content to change.

20. The apparatus of claim 19, further comprising:

a sensing device, which is one of: an infrared distance sensing device; a laser distance sensing device; a SONAR distance sensing device; and an image capture device for capturing an image of the user's face.

21. A method, comprising:

from a first captured image of a user's face, determining that a particular facial feature has a first size;

determining, based at least in part upon the first size, a first scaling factor;

causing a set of visual content on a display to be sized according to the first scaling factor;

from a second captured image of the user's face, determining that the same particular facial feature is of a second size, wherein the second size is different from the first size;

determining, based at least in part upon the second size, a second scaling factor, wherein the second scaling factor is different from the first scaling factor; and

causing the set of visual content to be sized according to the second scaling factor to cause a display size of the set of visual content to change.

22. The method of claim 21, wherein the particular facial feature is a separation between two distinct portions of the user's face.

23. The method of claim 22, wherein the first and second captured images of the user's face comprise a plurality of pixels, wherein the first size indicates a first number of pixels spanned by the separation between the two distinct portions of the user's face in the first captured image, and wherein the second size indicates a second number of pixels spanned by the separation between the two distinct portions of the user's face in the second captured image.

24. The method of claim 21, further comprising:

performing a calibration procedure, wherein the calibration procedure comprises:

from a first calibration image of the user's face captured while the user's face is at a first distance from the display, determining that the particular facial feature has a first calibration size;

while the user's face is at the first distance from the display, receiving input from the user indicating a first desired scaling factor;
from a second calibration image of the user's face captured while the user's face is at a second distance from the display, determining that the particular facial feature has a second calibration size, wherein the second distance is different from the first distance and the second calibration size is different from the first calibration size; and
while the user's face is at the second distance from the display, receiving input from the user indicating a second desired scaling factor, wherein the second desired scaling factor is different from the first scaling factor.

25. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 21.

26. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 23.

27. A computer readable storage medium having instructions stored thereon which, when executed by one or more processors, cause the one or more processors to perform the method of claim 24.

28. An apparatus, comprising:
one or more processors; and
one or more storages having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform the operations of:
from a first captured image of a user's face, determining that a particular facial feature has a first size;
determining, based at least in part upon the first size, a first scaling factor;
causing a set of visual content on a display to be sized according to the first scaling factor;

from a second captured image of the user's face, determining that the same particular facial feature is of a second size, wherein the second size is different from the first size;

determining, based at least in part upon the second size, a second scaling factor,

wherein the second scaling factor is different from the first scaling factor; and

causing the set of visual content to be sized according to the second scaling factor to cause a display size of the set of visual content to change.

29. The apparatus of claim 28, further comprising:

an image capturing device for capturing the first and second captured images of the user's face.

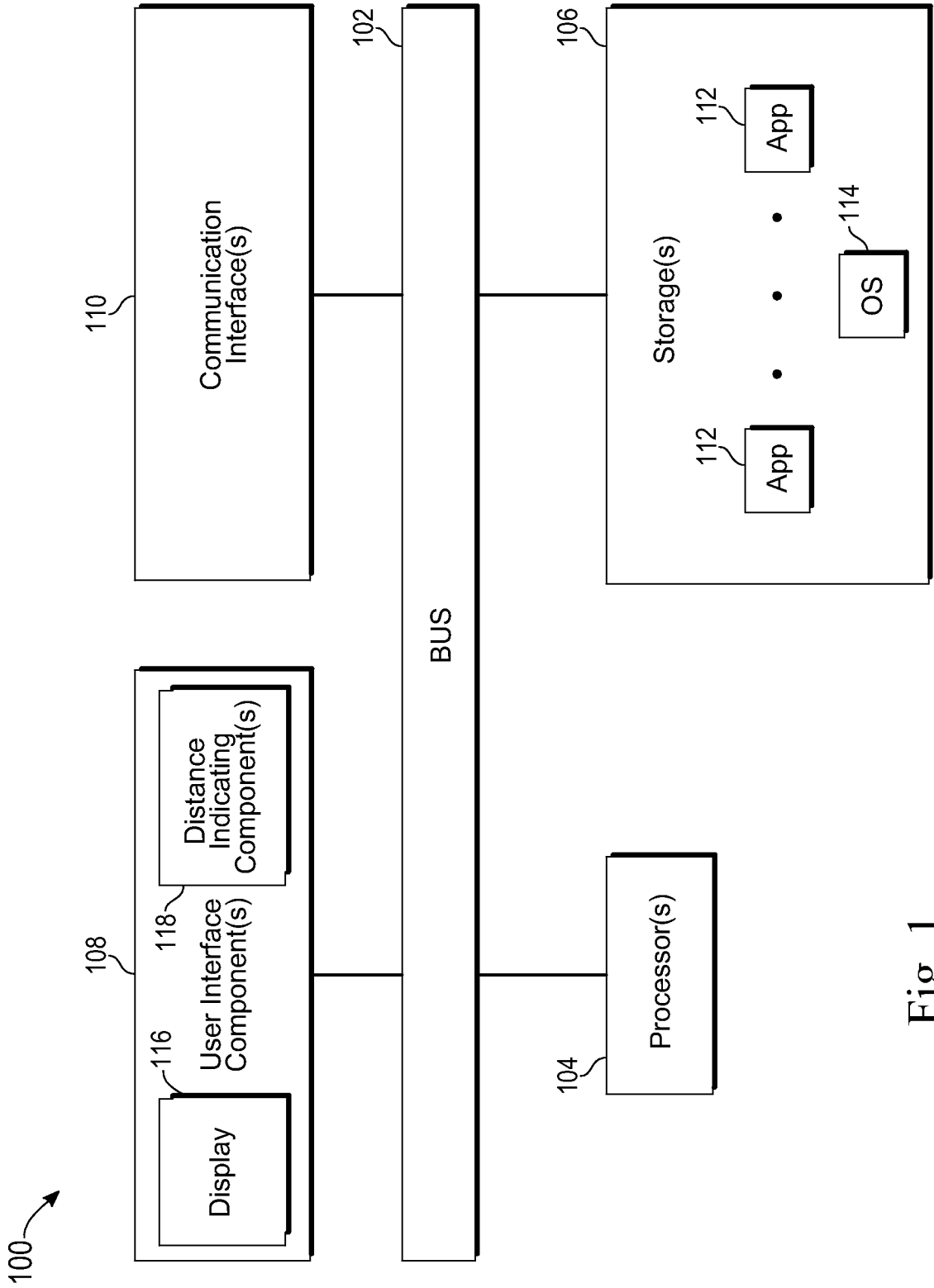


Fig. 1

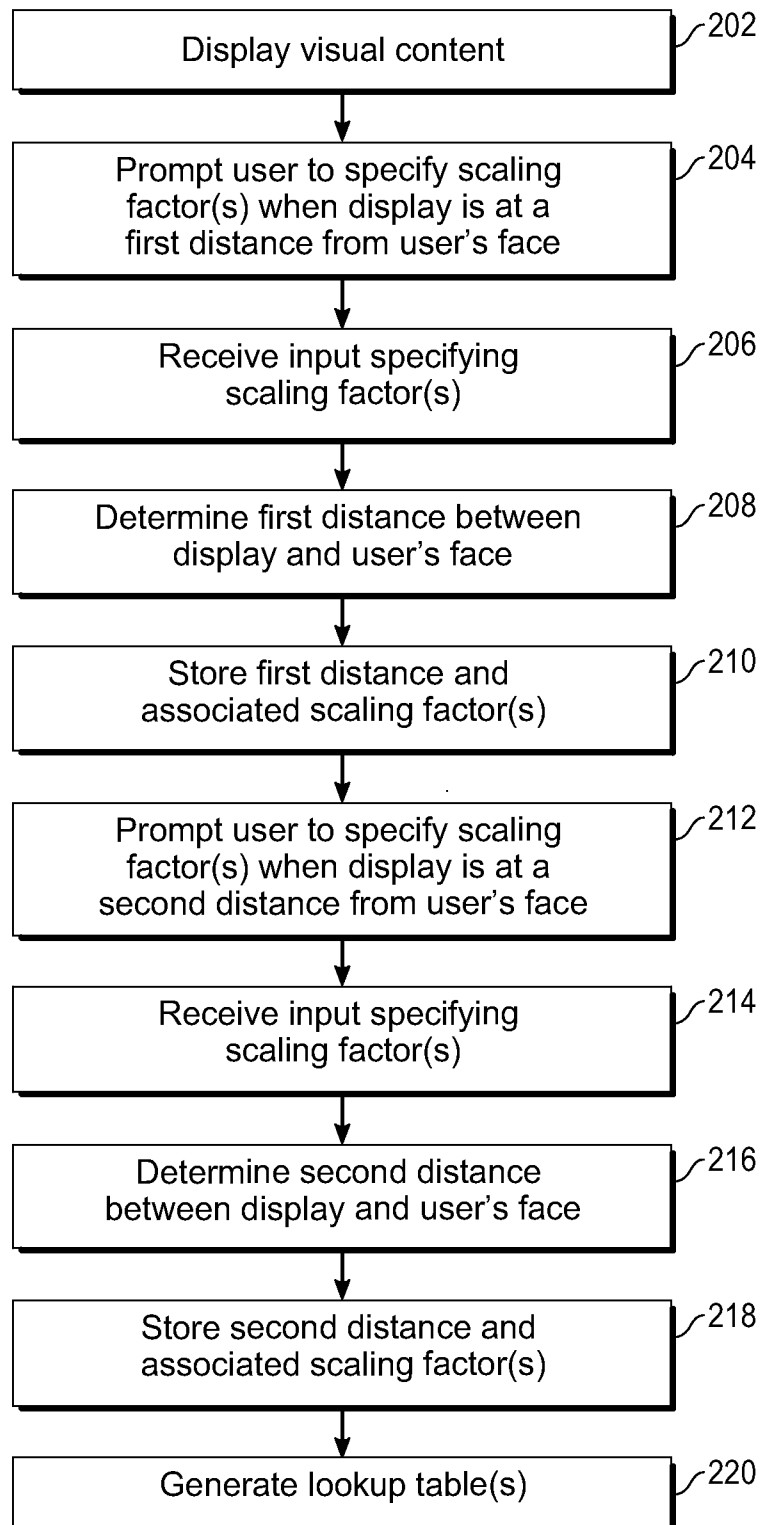


Fig. 2

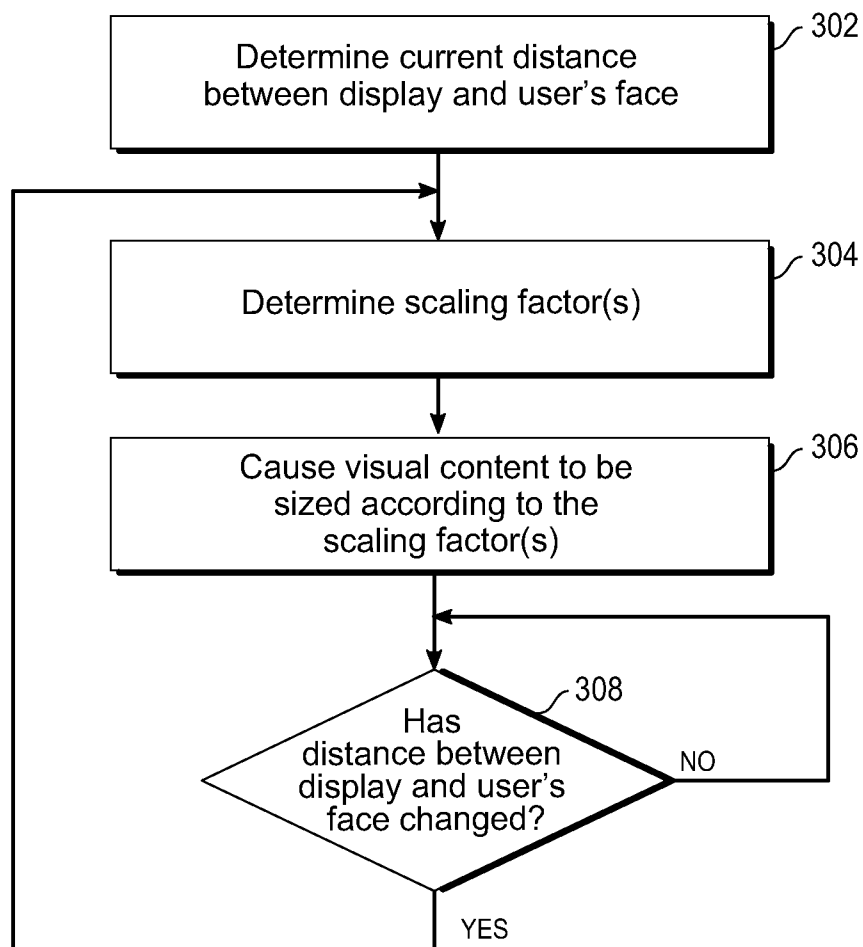


Fig. 3

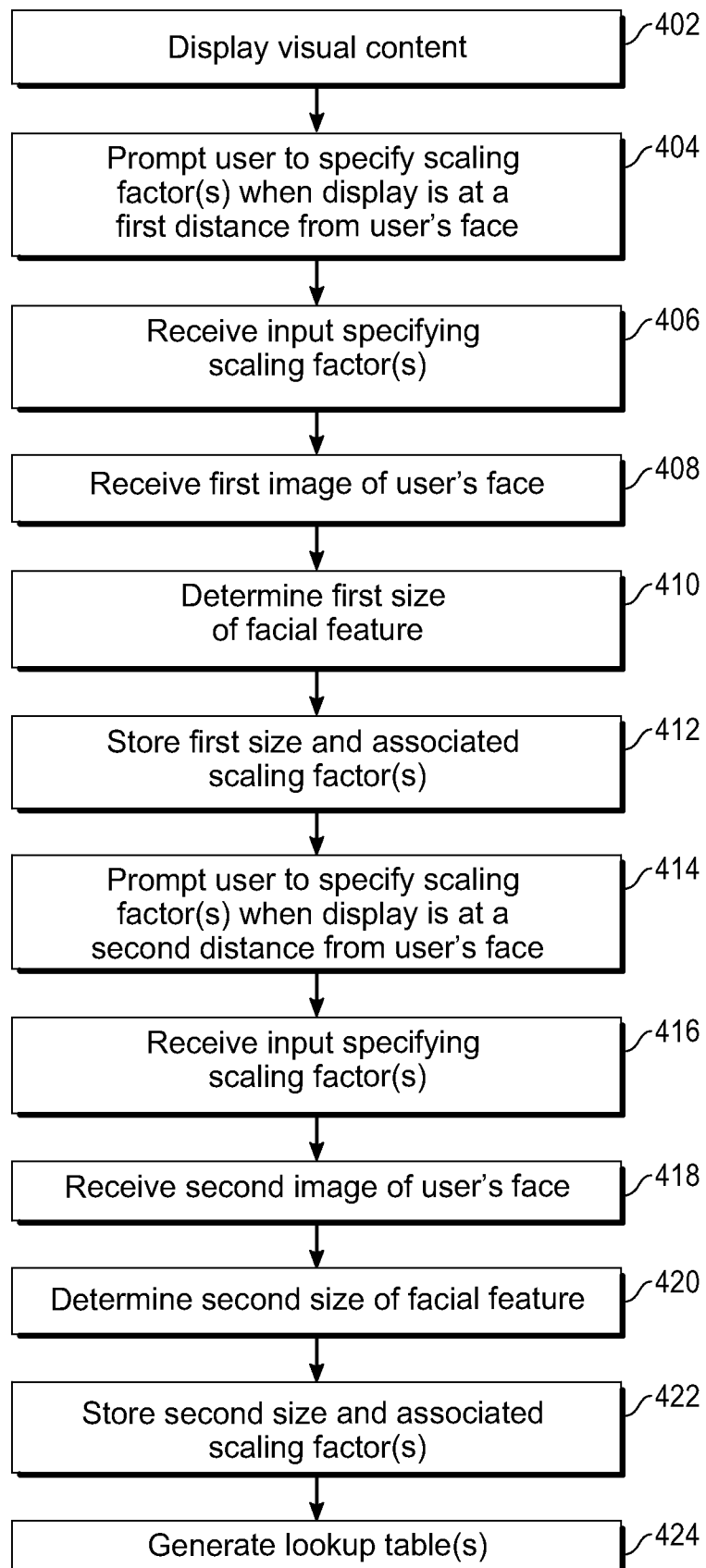


Fig. 4

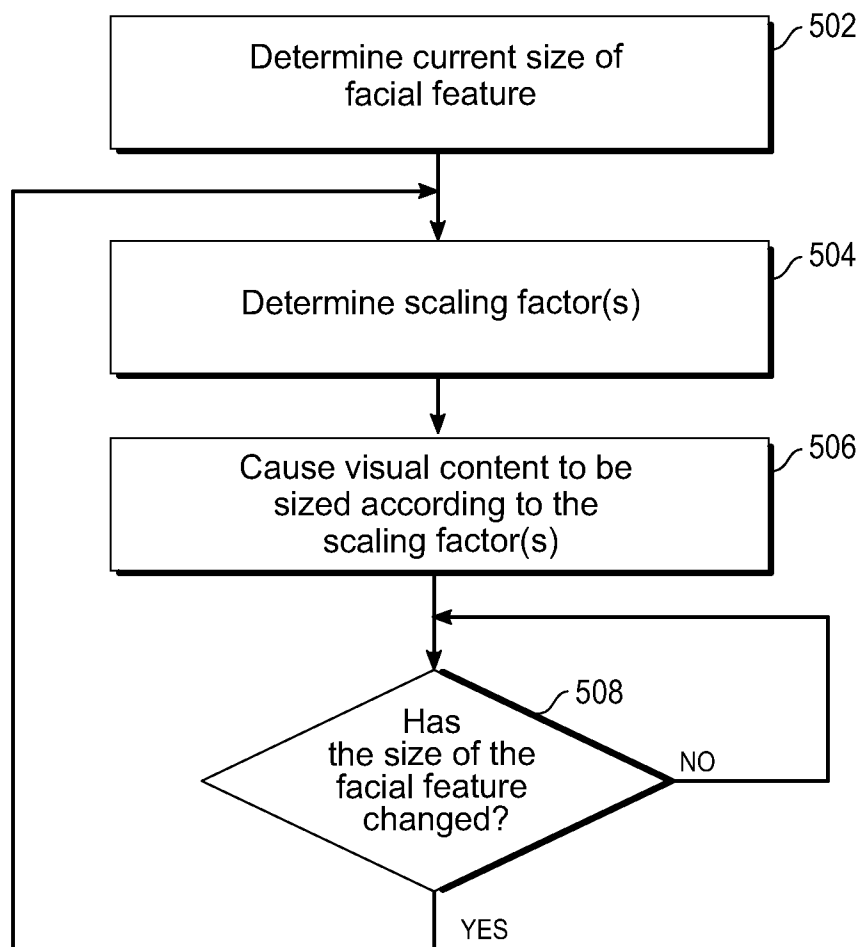


Fig. 5

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2012/033505

A. CLASSIFICATION OF SUBJECT MATTER

INV. G09G5/00 G06F3/01
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G09G G06F H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2009/164896 A1 (THORN KARL OLA [SE] THOEM KARL OLA [SE]) 25 June 2009 (2009-06-25) figures 1-7 paragraphs [0004] - [0010], [0040] - [0049], [0054], [0056], [0058] - [0060], [0062], [0063], [0067] claims 1,4	1-29
X	US 2003/234799 A1 (LEE BEOM-SEOK [KR]) 25 December 2003 (2003-12-25) figures 1-6 paragraphs [0008] - [0013], [0033], [0038] - [0042] ----- -/--	1-29



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

5 July 2012

Date of mailing of the international search report

13/07/2012

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Maciu, Emanoil

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2012/033505

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 7 181939 A (ROHM CO LTD) 21 July 1995 (1995-07-21) figures 1-3 the whole document -----	1,2,4-8, 10-12, 14-17, 19-29
X	US 2004/160386 A1 (MICHELITSCH GEORG [DE] ET AL) 19 August 2004 (2004-08-19) figures 1-3 paragraphs [0009] - [0028], [0035] - [0039], [0042] - [0045], [0051], [0061] - [0064] -----	1-3,5-8, 10-12, 14-17, 19-29
X	US 2009/201314 A1 (TANAKA KAZUMASA [JP] ET AL) 13 August 2009 (2009-08-13) figures 21-24 paragraphs [0012], [0107] - [0121] -----	1,2,4, 6-8, 10-12, 14-17, 19-29 5
A		

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2012/033505

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009164896 A1	25-06-2009	EP 2220548 A1 JP 2011510364 A US 2009164896 A1 WO 2009081239 A1	25-08-2010 31-03-2011 25-06-2009 02-07-2009
US 2003234799 A1	25-12-2003	KR 20030097310 A US 2003234799 A1	31-12-2003 25-12-2003
JP 7181939 A	21-07-1995	NONE	
US 2004160386 A1	19-08-2004	EP 1426919 A1 JP 2004185007 A US 2004160386 A1	09-06-2004 02-07-2004 19-08-2004
US 2009201314 A1	13-08-2009	CN 101510978 A EP 2091240 A2 JP 2009192737 A US 2009201314 A1	19-08-2009 19-08-2009 27-08-2009 13-08-2009