EUROPEAN PATENT SPECIFICATION

SYSTEM AND METHOD FOR MANIPULATION OF A DIGITAL IMAGE
SYSTÈME ET PROCÉDÉ DE MANIPULATION D’UNE IMAGE NUMÉRIQUE

Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO SE SI SK TR

Priority: 08.10.2008 US 247300

Proprietor: Sony Ericsson Mobile Communications AB
221 88 Lund (SE)

Inventor: THÖRN, Karl Ola
21145 Malm (SE)

Representative: Valea AB
Anna Lindhs Plats 4
211 19 Malmö (SE)

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).
TECHNICAL FIELD OF THE INVENTION

The present invention relates to electronic devices that render a digital image, and more particularly to a system and methods for rendering and manipulating a digital image or a portion thereof.

DESCRIPTION OF THE RELATED ART

Contemporary digital cameras typically include embedded digital photo album or digital photo management applications in addition to traditional image capture circuitry. As digital imaging circuitry has become less expensive, other portable devices, including mobile telephones, portable data assistants (PDAs), and other mobile electronic devices often include embedded image capture circuitry (e.g. digital cameras) and digital photo album or digital photo management applications in addition to traditional mobile telephony applications.

Popular digital photo management applications include several photograph manipulation functions for enhancing photo quality, such as correction of red-eye effects, and manipulating a digital image by lightening or darkening an image, digitally zooming in or out, rotating an image, and others. Another popular digital photo management function is a function known as text tagging.

Text tagging is a function in which the user selects a digital photograph or portion thereof and associates a text item therewith. The text item is commonly referred to as a "text tag". When viewing digital photographs, the text tag provides information about the photograph, effectively replacing an age old process of handwriting notes on the back of a printed photograph or in the margins next to a printed photograph in a photo album. Digital text tags also provide an advantage in that they can be easily searched to enable locating and organizing digital photographs within a database.

Similar to text tagging of photographs, there are also applications that provide for the text tagging of documents and other digital images or content. For example, a group of individuals editing a document may provide each other margin comments relating to content. Text tagging may also be employed to highlight content on web pages or comparable image content accessed over the Internet or other network.

When digital photo management or digital image applications are operated on a traditional computer with a traditional user interface (e.g. full QWERTY keyboard, large display, and a convenient pointer device such as a mouse), applying text tags to photographs, documents, or other digital images is relatively easy. The user, for example, simply utilizes the pointer device to select a point within the displayed digital image, mouse-clicks to "open" a new text tag object, types the text tag, and mouse-clicks to apply the text tag to the digital image.

Portable electronic devices, however, such as digital cameras, mobile telephones, portable data assistants (PDAs), and others, typically do not have such a convenient user interface. The display screen is much smaller. The keyboard may have a limited quantity of keys (typically what is known as a "12-key" or "traditional telephone" keyboard), and the pointing device, if present at all, may comprise a touch screen (or stylus activated panel) over the small display. Five-way multi-function keypads also are common. This type of user interface makes the application of text tags to digital photographs and other digital content cumbersome at best.

Speech-to-text applications have been used to generate text without the use of a keyboard, mouse, or associated input devices. Such applications have been employed primarily for document generation. With respect to text tagging a digital image, however, conventional speech-to-text applications have proven less suitable because of the need to place a text tag in a particular location within a digital image. Furthermore, speech-to-text applications may not be suitable in situations in which others are present, who may find audible speech disturbing.

In a separate field of art, eye tracking and gaze direction systems have been contemplated. Eye tracking is the process of measuring the point of gaze and/or motion of the eye relative to the head. The most common contemporary method of eye tracking or gaze direction detection comprises extracting the eye position relative to the head from a video image of the eye. In addition to eye tracking, other forms of face detection are being developed. For example, one form of face detection may detect particular facial features, such as whether an individual is smiling or blinking. It also been contemplated to employ face detection for "lip reading" to convert lip and mouth movements into speech or text. To date, however, such technologies have not been fully utilized in the manipulation of digital photographs and other digital images rendered in a portable electronic device.

SUMMARY

Accordingly, there is a need in the art for an improved system and methods for the manipulation of digital images (and portions thereof) that are rendered on a portable electronic device. In particular, there is a need in the art for an improved system and methods for text tagging and applying visual effects to digital images rendered on a portable electronic device.

Accordingly, a device and method are defined in claims 1 and 8, respectively. Embodiments are further detailed in the dependent claims.

These and further features of the present invention will be apparent with reference to the following description and attached drawings. In the description and drawings, particular embodiments of the invention have been disclosed in detail as being indicative of some of
the ways in which the principles of the invention may be employed, but it is understood that the invention is not limited correspondingly in scope. Rather, the invention includes all changes, modifications and equivalents coming within the spirit and terms of the claims appended hereto.

[0013] Features that are described and/or illustrated with respect to one embodiment may be used in the same way or in a similar way in one or more other embodiments and/or in combination with or instead of the features of the other embodiments.

[0014] It should be emphasized that the terms "comprises" and "comprising," when used in this specification, are taken to specify the presence of stated features, integers, steps or components but do not preclude the presence or addition of one or more other features, integers, steps, components or groups thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015]

FIG. 1 is a schematic diagram of an exemplary electronic device for use in accordance with an embodiment of the present invention.

FIG. 2 is a schematic block diagram of operative portions of a text-tagging/image manipulation system that may be incorporated the electronic device of FIG. 1.

FIG. 3 is a schematic diagram of the exemplary electronic device of FIG. 1 displaying an exemplary modified digital image.

FIG. 4 represents an exemplary sequence of images that may represent changes in the configuration of a user's facial features as may occur during mouthng speech.

FIG. 5 is a flow chart depicting a first exemplary method of manipulating a digital image.

FIG. 6 is a schematic diagram of the exemplary electronic device of FIGs. 1 and 3 in which the digital image is divided into image portions.

FIG. 7 is a flow chart depicting a second exemplary method of manipulating a digital image.

FIGs. 8A-8C are schematic diagrams of an exemplary electronic device in which multiple text tags may be generated for an exemplary image.

FIG. 9 is a flow chart depicting a third exemplary method of manipulating a digital image.

FIG. 10 is an exemplary sequence of figures that depicts an exemplary use of eye detection to select an image portion for manipulation from within a broader digital image.

FIG. 11 is a flow chart depicting a fourth exemplary method of manipulating a digital image.

FIG. 12 is a second exemplary sequence of figures that depicts an exemplary use of eye detection to select an image portion for manipulation from within a broader digital image.

FIG. 13 depicts an exemplary text tagging as applied to a moving video image.

FIG. 14 depicts an exemplary database of text tags generated from various examples described herein.

DETAILED DESCRIPTION OF EMBODIMENTS

[0016] Embodiments of the present invention will now be described with reference to the drawings, wherein like reference numerals are used to refer to like elements throughout. It will be understood that the figures are not necessarily to scale.

[0017] With reference to FIG. 1, an exemplary electronic device 10 is embodied in a digital camera, mobile telephone, mobile PDA, or other electronic device having a digital camera function. Such devices may include any portable radio communication equipment or mobile radio terminal, including pagers, communicators, electronic organizers, smartphones, and any communication apparatus or the like having a camera function. Desktop or laptop computers, gaming devices or games consoles, or similar devices having a web camera also may be employed. The particular electronic device, therefore, may be varied without departing from the scope of the invention.

[0018] The electronic device 10 has a display 12 for the rendering of information, and particularly for rendering a digital image 15. In the example of the figure, the digital image represents a still photograph. The digital image may be a still or moving video image, and may be derived from any source such as the camera viewfinder, an internal memory, an external or network storage device, or the like. The digital image also may represent an image from a network, such as a portion of an Internet webpage, may be a document image, or may comprise both text and/or non-text portions. The following examples are described primarily in reference to the manipulation of a digital photograph, although it will be appreciated that comparable principles may be applied to any digital image.

[0019] The electronic device may have a camera assembly 20 for taking photographs or moving video images. The camera assembly may include a subject image capturing lens 17 that is represented by hidden lines to indicate its presence on the side of the device opposite
the user. In this manner, the lens 17 has a field of view directed away from the back side of the display 12 for capturing still and/or motion video images 15, such that the display 12 may operate as the camera viewfinder. The camera assembly 20 may also include a user image capturing device 42 for capturing an image of a user. For example, the user image capturing device may be a user image capturing lens 42 that faces toward the user during use. The lens 42, therefore, may monitor the motions of the user. For example, if the electronic device has a telephone calling function, the lens 42 and the camera assembly 20 may provide a video telephony function that captures an image of the user when the user is participating in a telephone call. As is further described below, lens 42 and camera assembly 20 also may capture an image of the user for face detection, including lip reading and eye tracking functions, in accordance with embodiments of the present invention.

[0020] FIG. 2 represents a functional block diagram of operative portions of a text-tagging/image manipulation system 90 that may be incorporated into the electronic device 10. Further to the above, the system 90 may include the camera assembly 20 having the subject image capturing lens 17 and the user image capturing lens 42. Camera assembly 20 may also include one or more image sensors 16 for receiving the light from the lenses to generate the images. Each lens 17 and 42 may be associated with its own sensor, or sensing may be performed for both lenses by a common sensor. Camera assembly 20 may also include other features common in conventional digital cameras, such as light meter, a flash, and the like.

[0021] It will be appreciated that the digital images to be manipulated as described herein need not come from the camera assembly 20. For example, digital images may be stored in and retrieved from a memory 50. In addition, digital images may be accessed from an external or network source via any conventional wired or wireless network interface (not shown).

[0022] Referring again to FIG. 2, the system 90 may include a primary control circuit 30 that is configured to carry out overall control of the functions and operations of the system 90, as well as other components of the broader electronic device 10. The control circuit 30 may include a processing device 34, such as a CPU, microcontroller or microprocessor.

[0023] Among their functions, to implement the features of the present invention, the control circuit 30 and/or processing device 34 may comprise a controller that may execute program code stored on a machine-readable medium embodied as the text tagging/image manipulation application 37. Application 37 may be a stand-alone software application or form a part of a software application that carries out additional tasks related to the electronic device 10. It will be apparent to a person having ordinary skill in the art of computer programming, and specifically in application programming for mobile telephones, servers or other electronic devices, how to program an electronic device to operate and carry out logical functions associated with the application 37. Accordingly, details as to specific programming code have been left out for the sake of brevity. Also, while the code may be executed by control circuit 30 in accordance with exemplary embodiments, such controller functionality could also be carried out via dedicated hardware, firmware, software, or combinations thereof, without departing from the scope of the invention.

[0024] The system 90 may include a rendering device, such as a display 12 for rendering digital images and associated visual content to a user. The display 12 may be coupled to the control circuit 30 by a video processing circuit 64 that converts video data to a video signal used to drive the display. The processing circuit 64 may include any appropriate buffers, decoders, video data processors and so forth. The video data may be generated by the control circuit 30, retrieved from a video file that is stored in the memory 50, derived from an incoming video data stream received from an external network, or obtained by any other suitable method.

[0025] FIG. 3 depicts the electronic device 10 displaying the digital image of FIG. 1. In FIG. 3, a text tag 59, stating "Vacation", has been added to associate a subject matter with the image. In accordance with embodiments of the present invention, the text tag may be generated as follows.

[0026] Referring to FIGs. 1-3, a text tag may be applied to any digital image rendered on a rendering device, such as the display 12. The text tagging application 37 may include a conversion module 38 for converting a sequence of images of a user, captured by the lens 42, into a manipulation command. In one exemplary embodiment, the conversion module is a lip reading module 38. In general, the lip reading module 38 may interpret the movement and configuration of a user's lips and mouth into text for text tagging.

[0027] The lip reading module 38 may be actuated by a variety of ways of inputting commands in an electronic device. In one embodiment, lip reading may be activated using a dedicated activation key input or button, such as an activation key 13. In this example, if a user presses and holds the key 13, the user image capturing lens 42 of the camera assembly 20 may be employed to detect the motion and configuration of a user's lips and mouth as a sequence of images 52. The sequence of images 52 may then be passed to the lip reading module 38, which interprets the motion and configuration of the sequence of images as speech text.

[0028] The motion and configuration detection may be interpreted by means of object recognition, edge detection, silhouette recognition, or other means for detecting motion as are known in the art. For example, FIG. 4 represents a sequence of images 52a-c that may represent changes in the configuration of a user's mouth as may occur during mouthing speech. As indicated by the arrows in the figure, a change in the position, configuration, and movement of the mouth and lips represent both a
change in orientation and a change in velocity of these facial features. There may also be changes in the optical flow in that subtle differences in shadowing and light balance, caused by the changes in feature orientation, may result in gray value changes over consecutive images in the sequence. The changes in feature orientation, feature velocity, and/or optical flow may be analyzed by the lip reading module 38, which may then interpret the images 52a-c to generate a text representation of the words mouthed by the user. The text may be converted into a text tag for inclusion in the digital image. The user may mouth the text tag as a sequence of any recognizable portions of speech, such as individual letters, whole words, and/or syllables.

[0029] Once the user’s mouthing of the text tag is complete, the user may release the key or button 13 to terminate the generation of the text tag. The text representation may then be outputted from the text tagging application 37, via the control circuit 30, to drive the video processing circuit 64 and display 12. The result is the altering of the rendering of the digital image in the display by placement of the text tag 59 in the digital image. Thus, in the example of FIG. 3, a user presses and holds the key 13 while mouthing the word “Vacation”. The text tag “Vacation” may then be placed within the digital image. In this manner, a text tag may be added to the digital image 15 without having to employ a conventional keypad. In addition, by employing lip reading, the text tag may be added quietly, and thus reduce disturbances to others who may be nearby.

[0030] It will be appreciated that the lip reading module may be employed to manipulate a digital image in ways other than text tagging. For example, rather than pressing and holding, a user may press the key 13 without holding to toggle between a visual effects mode and a text tagging mode. Once a mode is selected, a user may press and hold the key 13 to mouth either a visual effects/manipulation command or a text tag, as is commensurate with the selected mode. For example, when in the visual effects mode, the lip reading module may recognize the mouthing of such image manipulation commands as lighten, darken, rotate (left or right), zoom in, zoom out, and others. The manipulation command may then be implemented via the control circuit 30, to drive the processing circuit 64 and display 12 to alter the rendering of the image as desired.

[0031] In accordance with the above, FIG. 5 is a flow chart depicting an exemplary method of manipulating a digital image. Although the exemplary method is described as a specific order of executing functional logic steps, the order of executing the steps may be changed relative to the order described. Also, two or more steps described in succession may be executed concurrently or with partial concurrence. It is understood that all such variations are within the scope of the present invention.

[0032] The method may begin at step 100, at which a digital image is rendered. At step 110, the user may select either a text tagging mode or a visual effects mode. Regardless of the selected mode, at step 120a or 120b, the user may activate the conversion module, which in this case is a lip reading module, such as by pressing and holding the key 13. If in the text tagging mode, the lip reading module at step 130a may detect lip and mouth configuration changes corresponding to a mouthed text tag. When the user has completed mouthing the text tag, at step 140a the user may deactivate the lip reading module, such as by releasing the key 13. At step 150a, the system may generate the text tag, such as by employing the processing circuit to alter the rendering of the digital image to include the text tag. Similarly, if in the visual effects mode, the lip reading module may detect at step 130b lip and mouth configuration changes corresponding to a mouthed visual effects command. When the user has completed mouthing the visual effects command, at step 140b the user may deactivate the lip reading module, again such as by releasing the key 13. At step 150b, the system may implement the visual effects command, such as by employing the processing circuit to alter the rendering of the digital image in accordance with the command.

[0033] Although the above description is in connection with lip reading to generate the text tag, other forms of facial detection may be employed in addition or in alternative to lip reading. For example, default text tags may be associated with other facial gestures, such as blinking, non-verbal mouth movements, other head movements, and the like. Speak recognition techniques also may be employed together with lip reading to enhance the text tag creation function.

[0034] Note that once a text tag or visual effect has been generated, the digital image may be stored in its altered format. In one embodiment, the altered portions are incorporated into or referenced in the metadata of the image file. The altered digital image may be stored in any conventional storage device, such as the memory 50 of the electronic device 10, or an external or network storage device.

[0035] It is desirable, of course, that a text tag be placed at a suitable location within the digital image. For example, in FIG. 3 the text tag is placed adjacent the upper-right corner of the image, apart from the principle subject matter. In this manner, the text tag results in minimal interference with the appreciation of the image. In one embodiment, a text tag may be placed in a default location that, as a matter of probability, would represent a desirable location. For example, a photograph typically would have the principle subject matter centered within the image. An appropriate default location may be adjacent one of the corners, such as the upper-right corner in FIG. 3. If the user deems the default location to be unsuitable for a particular image, the user may move the text tag using conventional input devices, such as a stylus, keypad or touch screen inputs, or the like.

[0036] In a preferred embodiment, however, the text tag application 37 may automatically determine an appropriate location for the text tag 59. Referencing again to
from portions that do not contain principle subject matter. The image analysis module may differentiate landmarks, and various other prominent objects within the image. The image analysis module may differentiate portions of the image containing principle subject matter from portions that do not contain principle subject matter (such as relatively featureless land, water, sky, and the like).

Referring to FIG. 6, the image analysis module 39 may divide a digital image 15 into portions with a grid 44. In this example, the grid 44 divides the image 15 into nine rectangular portions of substantially equal area. It should be appreciated that although the grid 44 is represented as rectangles, other shapes and sizes may also be implemented to divide the image. For example, other polygons or shapes may be employed, or the image may be divided by pixels or groups of pixels. In addition, the grid may be formed as part of a computing process, and thus may be invisible to a user and would not actually be displayed or rendered as part of the image 15.

The image analysis module 39 may identify one or more image portions 43 whose boundaries are defined by the grid, and which depict principle subjects or portions of principle subjects within the digital image. The image portions 43 may be identified by any of object detection, face detection, edge detection, or other means for identifying principle subject matter within a digital image 15. For example, in FIG. 6 five portions 43 have been identified that include portions of principle subject matter. The center column portions 43 include portions of the monument and its reflection in the water. The image analysis module has also identified the left-center and right-center portions 43 as containing principle subject matter, recognizing the combination of forest and water. The remaining portions of the image as defined by the grid 44 are, for the most part, relatively featureless regions of sky or water. Accordingly, the text tagging application 37 has caused the text tag to be rendered in one of the predominantly featureless regions (adjacent the upper-right corner), apart from the principle subject matter. In the event the user does not agree with the automatic placement, the user still may have the option of manually altering the placement of the text tag to a more suitable location.

In accordance with the above, FIG. 7 is a flow chart depicting a second exemplary method of manipulating a digital image. Although the exemplary method is described as a specific order of executing functional logic steps, the order of executing the steps may be changed relative to the order described. Also, two or more steps described in succession may be executed concurrently or with partial concurrence. It is understood that all such variations are within the scope of the present invention.

The method of FIG. 7 is analogous to the method of FIG. 5, except in the manner by which the text tag 59 is generated. In FIG. 7, once the user has mouthed the complete text tag and deactivated the lip reading module at step 140a, at step 142 the image may be analyzed for portions containing principle subject matter. At step 150a, the text tag is rendered apart from the image portions containing principle subject matter. In this manner, the potential for interference with the appreciation of the digital image may be reduced.

It will be appreciated that in the examples of FIGs. 3 and 6, a single text tag 59 has been applied to be associated with the entire image 15. It may be desirable, however, particularly when an image has more than one principle subject, to apply a text tag to a specific portion of the image, rather than the whole. Similarly, it may be desirable to apply multiple text tags, with each text tag pertaining to a specific portion of a digital image.

In FIGs. 8A-8C depict an exemplary image 15a-c in which multiple text tags may be generated. In the manner described above, the image analysis module 39 may divide the image with a grid 44 and identify image portions 43 containing principle subject matter. In this example, as shown in FIG. 8A, four image portions 43 have been identified that contain principle subject matter - a house, a boat, a dog, and a person. A user may navigate the grid to select a specific image portion 43 with a navigation device. For example, the electronic device 10 may have a navigation device in the form of a keypad 11 (see also FIG. 2). In the example of FIGs. 8A-8C, the keypad 11 is a conventional five-way navigation ring, including four directional input surfaces and a center "select" button. A selected image portion may be indicated with an indicator 41, shown in the figures as being slanted lines. It will be appreciated that other indicators may be employed, such as highlighting, shading, and the like. In FIG. 8A, the house image portion 43 has been selected as indicated by the slanted lines. In FIG. 8B, the user has navigated to the boat image portion 43, and in FIG. 8C, the user has navigated to the dog image portion 43, also as indicated by the slanted lines.

Note that as depicted in FIGs. 8A-8C, once a user has navigated to a particular image portion 43, the user may generate a text tag specific to that image portion. Although each figure depicts only one text tag, it will be appreciated that more than one text tag may be generated for a digital image. For example, the resultant digital image 15 may contain all three text tags of "house", "boat", and "dog", or any combinations thereof. A user also may generate text tags additionally to and/or alternatively to those depicted in FIGs. 8A-8C.

Each of the text tags may be generated in the manner described above. For a selected image portion 43, the user may execute the text tagging operation with a press- and-hold key input, such as with key 13. The user image capturing lens 42 may then capture the mouthing of a text tag as a sequence of images, after which the user may release the key input. The lip reading...
module may interpret the mouth and lip movements, and convert such movements to text, which may then be incorporated as a text tag within the digital image. In addition, the image analysis module 39 may determine a placement of the text tag within the digital image so as to place the text tag within (or partially within) or adjacent to the associated selected image portion, while minimizing interference with the subject of the image portion. A user may repeat such sequence for additional selected image portions to add additional text tags.

[0045]  In accordance with the above, FIG. 9 is a flow charting depicting a third exemplary method of manipulating a digital image. Although the exemplary method is described as a specific order of executing functional logic steps, the order of executing the steps may be changed relative to the order described. Also, two or more steps described in succession may be executed concurrently or with partial concurrence. It is understood that all such variations are within the scope of the present invention. The method of FIG. 9 is comparable to the methods of FIGs. 5 and 7, except that the method of FIG. 9 includes additional steps pertaining to the manipulation of specific image portions within the broader digital image.

[0046]  The method may begin at step 200 at which the image may be rendered. At step 203, the image analysis module 39 may identify image portions having principle subject matter. At step 206, a selection of one of such image portions may be received, either as a default input or a user input from among the image portions having principle subject matter. At step 210, as before, the user may select either a text tagging mode or a visual effects mode. Regardless of the selected mode, at step 220a or 220b the user may activate the conversion module, which in this example is a lip reading module, such as by pressing and holding the key 13. If in the text tagging mode, the lip reading module at step 230a may detect lip and mouth configuration changes corresponding to a mouthed text tag. When the user has completed mouthinging the text tag, at step 240a the user may deactivate the lip reading module, such as by releasing the key 13. At step 250a, the system may generate the text tag, such as by employing the processing circuit to alter the rendering of digital image to include a text tag associated with the selected image portion. Similarly, if in the visual effects mode, the lip reading module may detect at step 230b lip and mouth configuration changes corresponding to a mouthed visual effects command. When the user has completed mouthing the visual effects command, at step 240b the user may deactivate the lip reading module, again such as by releasing the key 13. At step 250b the system may implement the visual effects command, such as by employing the processing circuit to alter the rendering of the selected image portion of the digital image in accordance with the command.

[0047]  In this embodiment, at step 255 a determination may be made as to whether there is an additional image portion to be manipulated. For example, a user may employ the keypad/navigation ring 11 to select another image portion. If an input of an additional image portion is received at step 255, the method may return to step 206 to manipulate the next image portion. When no additional image portions are selected at step 255, i.e., the user is finished manipulating the image, the method may end.

[0048]  Referring again to FIG. 2, the text tagging application 37 may also include a facial detection module 40 for detecting one or more facial features of a user. In one embodiment, the facial detection module 40 is an eye detection module 40 for detecting the orientation of the eyes and adjacent facial features. As further described below, the facial/eye detection module 40 may constitute an alternative navigation device for navigating within the digital image and selecting an image portion. The eye detection module may be employed instead of the five-way navigation ring 11 so as provide additional automation and reduce the use of manual inputs. Also as further described below, the user image capturing lens 42 may also capture the movement and orientation of the eyes and adjacent facial features (e.g., the nose). The eye detection module 40 may translate such movements and orientation into a directional input or command for selecting an image portion 43 within the digital image.

[0049]  FIG. 10 is a sequence of figures that depicts the use of eye detection to select an image portion 43 within the digital image 15a-c from the previous figures. For simplicity, the components of the electronic device 10 are omitted from FIG. 10 to focus on the facial detection, and particularly the eye detection-based, selection features.

[0050]  Similar to FIGs. 8A-8C, in the example of FIG. 10, image portions 43 are determined which contain principle subject matter within the digital image. The image portions 43 may contain such objects commonly tagged in photographs such as people, houses, animals, or other prominent objects, as opposed to relatively featureless portions. In the top drawing of FIG. 10, the left-center image portion 43 has been selected from the digital image 15a, as indicated by the slanted lines indicator 41. As stated above, other indications of selection may be employed, such as shading or highlighting a selected image portion. The user has generated a text tag 59a "House" for this image portion in the manner described above.

[0051]  Referring to FIGs. 2 and 10, the eye detection module 40 may be employed to select a different image portion 43 as follows. The user image capturing lens 42 may detect the motion and orientation of the user's facial features, particularly the movement and orientation of the user's eyes and adjacent facial features. Such movement and orientation may be captured as a sequence of images and determined by object recognition, edge detection, silhouette recognition or other means for detecting motion of any item or object detected within such sequence of images. The movement of the facial features may then be converted into a directional input as follows. For example, reflections of illumination off the user's cornea may be utilized to determine where on the display 12 the user has focused, and/or to determine a change
in position of the user's focus on the display 12. In general, the eye detection module 40 may monitor the sequence of images captured by the user image capturing lens 42, and upon detecting a qualified motion, generate a direction vector representative of the direction of such motion. The direction vector, therefore, corresponds to a directional input. The eye detection module may then generate a repositioning output to change the selection of the particular image portion 43 in accordance with the direction vector. The repositioning output may also cause the video processing circuit 64 to alter the display of the indicator 41 commensurately to reflect the change in the selection of the particular image portion.

[0052] For example, in FIG. 10 elements 45a-d represent a sequence of images as may be detected by the eye detection module 40. Thus, the eye detection module 40 monitors the sequence of motion represented by frames 45a-45d. As in previous embodiments, the digital image 15a-c may be segmented into nine segments by the grid 44 by dividing the digital image vertically into thirds and horizontally into thirds. As seen in the top drawing of FIG. 10, after processing by the image analysis module 39, those segments that meet the criteria for principle subject matter are deemed image portions 43. In this example, the left center segment including an image of a house, the center segment including an image of a boat, the left lower segment including an image of a dog, and the right lower segment including an image of a person, may meet the subject criteria and be deemed as image portions 43. The remaining segments include relatively featureless sea or sky and may not meet the selection criteria for principle subject matter. As represented by image 15a, the selection indicator 41 (the slanted lines) is initially positioned at the left-center image portion 43.

[0053] To change the image portion selection and reposition the indicator 41, the eye detection module 40 may first detect that the user has turned his head to the right, as depicted in the thumbnail frames from 45a to 45b. The eye detection module 40 may define a direction vector 49 corresponding to the direction of motion of at least a portion of the user's face, as represented, for example, by the change in configuration and orientation of the user's eyes and adjacent facial features. The direction vector 49 may be derived from determining the relative displacement and distortion of a triangle formed by the relative position of the user's eyes and nose tip within the sequence of images captured by the user image capturing lens. For example, triangle 47a represents the relative positions of the user's eyes and nose within frame 45a, and triangle 47b represents the relative position of the user's eyes and nose within frame 45b. The relative displacement between triangle 47a and 47b along with the relative distortion indicate that the user has looked to the right as represented by direction vector 49.

[0054] The direction vector corresponds to a directional input based on the change in orientation of the user's facial features. In response to a movement as represented by the direction vector 49, the eye detection module 40 may generate an output signal to reposition the indicator 41 to the image portion 43 to the right of the previously selected image portion. This is depicted in the second rendering 15b, in which the slanted lines indicator 41 now indicates that the center image portion 43 has been selected. The user may now generate a text tag 59b "Boat" for this image portion using the lip reading module in the manner described above.

[0055] Similarly, when the user, as depicted in frame 45c, turns his head downward and to the left as depicted in frame 45d, the eye detection module 40 may determine a direction vector 51 corresponding to the direction of the movement of the user's face. Based on the new directional input based on direction vector 51, the eye detection module 40 may generate an output signal to reposition the indicator 41 to the image portion 43 to the lower-left image portion 43 from the previously selected center image portion. This is depicted in the third rendering 15c, in which the indicator 41 now indicates that the lower-left image portion 43 has been selected. The user may now generate a text tag 59c "Dog" for this image portion using the lip reading module in the manner described above.

[0056] It should be appreciated that if each of the nine segments represented an image portion 43, there may exist an ambiguity as follows. For example, overlaying the direction vector 49 on the digital image indicates that the movement of the indicator 41 (from the center right position as depicted in rendering 15a) could be to the upper-center portion of the digital image, the center portion of the digital image, or the upper right portion of the digital image. However, by first utilizing the image analysis module 39 to identify only those image portions meeting selection criteria for principle subject matter (as opposed to relatively featureless portions), only those image portions that depict principle subject matter will represent image portions 43. As such, there is little ambiguity that only the center portion is positioned relative to the center right portion in the direction of the direction vector 49. Accordingly, the motion represented by displacement of the user's face between frame 45a to 45b (resulting in vector 49) and from 45c to 45d (resulting in vector 51) results in the selection of an appropriate image portion 43 containing principle subject matter. Should a digital image indeed contain principle subject matter in every or nearly every portion of the grid 44, then the ambiguity may be resolved by always moving the selection to an adjacent image portion. Using multiple, single step movements, any image portion may be selected regardless of the starting position.

[0057] In the example of FIG. 10, only one text tag at a time is depicted within the digital image. As stated above respect to other embodiments, multiple text tags may be placed on a single digital image. The text tags may also be highlighted so that eye detection may interpret eye movements comparable to the manner described above to position the text tag within the image. Also as stated above, the lip reading module may be
employed to manipulate a digital image in ways other than text tagging. For example, in a visual effects mode, the lip reading module may recognize the mouthing of such image manipulation commands as lighten, darken, rotate (left or right), zoom in, zoom out, and others. The manipulation command may then be implemented via the control circuit 30, to drive the video processing circuit 64 and display 12 to manipulate the selected image portion as desired. In another embodiment, the lip reading module may be used in the visual effects mode to select the image portion 43. For example, the lip reading module may recognize directional commands such as "left", "right", "up" and "down" to generate a repositioning of the selection indicator 41.

In accordance with the above, FIG. 11 is a flow chart depicting a fourth exemplary method of manipulating a digital image. Although the exemplary method is described as a specific order of executing functional logic steps, the order of executing the steps may be changed relative to the order described. Also, two or more steps described in succession may be executed concurrently or with partial concurrence. It is understood that all such variations are within the scope of the present invention.

The method of FIG. 11 is analogous to the method of FIG. 9, except that FIG. 11 includes the use of eye detection to receive the selection of the additional image portion. In FIG. 11, once a text tag or visual effects command has been applied for the selected image portion at either of steps 250a or 250b, at step 256 a determination may be made as to whether a facial motion or reorientation is detected, such as a reorientation of the user's eyes and adjacent facial features. If so, at step 257 a direction vector may be generated commensurate with the motion or reorientation of the user's facial features. Based on the direction vector, at step 258 a new image portion may be selected for manipulation. The method may then return to step 210 to apply a text tag or visual effects command to the newly selected image portion as desired by the user.

FIG. 12 depicts another example in which the text tagging and image manipulation features are applied to an exemplary digital image. In this example, a digital image 14a-c depicts several people. The image analysis module 40 has identified each face as representing a principle subject, and rather than a grid, a circular indicator 60 is employed to identify each face as an image portion 43 for text tagging or other manipulation. For example, in the top rendering 14a, the left-most person has been selected as indicated by the indicator 60, and the user has generated a text tag 58a "Rebecka". To select another individual within the digital image, the facial/eye detection module 40 may receive a sequence of images of the user via the user image capturing lens 42. The sequence of images is again represented in FIG. 12 as elements 45a-d. In particular, from the user motion from 45a to 45b, the eye detection module 40 generates a direction vector 49 that results in the selection of the commensurate image portion. As shown in rendering 14b, the image portion to the upper-right of Rebecka has been selected, and the user generates a text tag 58b "Karl" for the newly selected image portion. From the user motion from 45c to 45d, the eye detection module 40 generates a direction vector 51 that results in the selection of the commensurate image portion. As shown in rendering 14c, the image portion to the lower-left of Karl has been selected, and the user generates a text tag 58c "Johan" for the newly selected image portion.

As before, although only one text tag is shown at a time in FIG. 12, multiple text tags may be generated simultaneously for the same digital image. In addition, when the subject matter of an image portion 43 is a person, an additional visual effects command may be red-eye correction.

As referenced above, once any of the text tags or visual effects has been generated in accordance with any of the described embodiments, the altered digital image may be stored in its altered format. In one embodiment, the altered portions are incorporated into or referenced in the metadata of the image file. The altered digital image may be stored in any conventional storage device, such as the memory 50 of the electronic device 10, or an external or network storage device.

Although the invention is been described with reference to a digital photograph, the embodiments may be implemented with respect to other digital images. For example, a user may add comments to a document rendered on a display, a webpage downloaded from the Internet or other network, or any other digital image.

For example, FIG. 13 depicts an exemplary application of the present invention to a moving video image 96. The exemplary video image 96 includes a plurality of frames 96a, 96b, and 96c. The frames of the video image 96 may be stored in a database, such as the memory 50 of the electronic device 10, may be real-time frames generated in the viewfinder (display 12) by the camera assembly 20, or real time frames streamed from an external video source or network.

In general, utilizing the features of previous embodiments, a text tag 98 may be added to one of the frames (for example frame 96a). Such text tag 98 may then be recorded as part of the video file as described above. In this manner, the text tag becomes associated with the particular tagged frame. Thus, frame 96a may be identified in the video file as a "tagged frame" 62. Subsequent motion of the portion of the image that was tagged (e.g. the depiction of Karl) is recorded as object motion data 64, as indicated by the arrows in the figure. As such, when subsequent frames 96b or 96c of the video clip 96 are rendered, the image analysis module recognizes the same depiction in such subsequent frames, and the text tag 98 remains with the portion of the image originally tagged, even as that portion is relocated within the frame. The text tag 98 thus “follows” Karl throughout the video.

In the above embodiment, the video may be tagged while a user replays a video. In another embod-
iment, the user may tag the video comparable to the above while the video is being recorded. Because the tagging is based upon lip reading by the lip reading module, the video may be tagged in the form of comments without the need to incorporate an audio soundtrack. The comments may appear in the video as text tags while the video is replayed. Additionally or alternatively, the tags may be stored to provide a means of searching for video about a particular subject matter. For example, if a user tags multiple videos of a particular person with the person’s name, the tags may be used to search a database to locate such videos that include the person of interest.

[0067] As referenced above, once any of the text tags or visual effects has been generated in accordance with any of the described embodiments, the altered digital image may be stored in its altered format. In one embodiment, the altered portions are incorporated into or referenced in the metadata of the image file. The altered digital image may be stored in any conventional storage device, such as the memory 50 of the electronic device 10, or an external or network storage device. In additional, the alteration may be stored in a separate database that would permit searching features. In one embodiment, the alterations, such as text tags and image manipulations, may be stored as a metadata fields associated with the respective digital images.

[0068] For example, FIG. 14 depicts an exemplary database 32 of text tags from the various examples described above. Each digital file may be associated with an identification (Photo ID) 52, a position 54 of the text tag within the digital image, and the text tag content 59. In one embodiment, the system may also store the sequence of images interpreted by the lip reading module as a visual tag 56. If the image is a video, such as the image of FIG. 13, the database may also include data pertaining to the tagged frame 62 and object motion data 64. It will be appreciated that the database described herein is an example, and the precise content and format of the database may be varied. In particular, the metadata may constitute stored metadata fields that are associated with the respective digital images.

[0069] The database 32 (or comparable metadata) may provide a searching function. A user may enter a search term corresponding to a text tag in any conventional manner. The search results may encompass digital images having the searched text tag. For example, if a user desired access to stored images that have been tagged as including “Rebecka”, the user may enter “Rebecka” as a search term to find such digital images.

[0070] Although the invention has been shown and described with respect to certain preferred embodiments, it is understood that equivalents and modifications will occur to others skilled in the art upon the reading and understanding of the specification. The present invention includes all such equivalents and modifications, and is limited only by the scope of the following claims.

Claims

1. An electronic device (10) comprising:

   a rendering device (12) for rendering a digital image;
   a user image capturing device (42) for capturing a sequence of images of a user;
   a conversion module (38) for converting the sequence of images into a text tag for the digital image by measuring at least one of a change in feature orientation, velocity or optical flow of depicted subject matter over consecutive images in the sequence; and
   a processing circuit (64) for altering the rendering of the digital image to include the text tag;
   wherein the conversion module (38) comprises a lip reading module and the sequence of images is a sequence of motion of the user’s facial features, and the lip reading module converts the motion of the user’s facial features into the text tag;
   and further comprising an image analysis module (39) for determining the placement of the text tag within the digital image, wherein the image analysis module divides the digital image into a plurality of image portions and performs object detection on each image portion to determine at least one image portion in which principal subject matter is contained; and the text tag is placed within the digital image apart from the image portion depicting the principal subject matter.

2. The electronic device (10) of claim 1, wherein the lip reading module (38) converts the motion of the user’s facial features into a visual effects command, and the processing circuit (64) alters the rendering of the digital image to implement the visual effects command.

3. The electronic device (10) of claim 2, wherein the visual effects command is at least one of lightening the digital image, darkening the digital image, zooming in, zooming out, rotating the digital image, or performing red-eye correction on a subject contained in the digital image.

4. The electronic device (10) of any of claims 1-3, further comprising a navigation device (11) for selecting a second image portion within the digital image; wherein the user image capturing device (42) captures the second sequence of images of the user, the conversion module converts (38) the second sequence of images into a second text tag for the digital image in a manner associated with the second image portion, and the processing circuit (64) alters the rendering of the digital image in a manner associated
5. The electronic device (10) of claim 4, wherein the navigation device (11) comprises a key (13) for receiving a navigation input for selecting the second image portion.

6. The electronic device (10) of any of claims 4-5, wherein the navigation device (11) comprises a facial detection module (40) for detecting a facial motion of the user corresponding to a directional input.

7. The electronic device (10) of claim 6, wherein the facial detection module (40) comprises an eye detection module for detecting the user’s eyes, and the directional input is a detected change in the orientation of the user’s eyes.

8. A method of manipulating a digital image comprising the steps of:

   rendering the digital image;
   capturing a sequence of images of a user;
   converting the sequence of images of the user into a text tag for the digital image by measuring at least one of a change in feature orientation, velocity or optical flow of depicted subject matter over consecutive images in the sequence, wherein the sequence of images is a sequence of motion of the user’s facial features and this motion is converted to a text tag via lip reading, and
   altering the rendering of the digital image to include the text tag;
   and further comprising determining the placement of the text tag within the digital image by dividing the digital image into a plurality of image portions and performing object detection on each image portion to determine at least one image portion in which principal subject matter is contained; and
   placing the text tag within the digital image apart from the image portion depicting the principal subject matter.

9. The method of claim 8, further comprising:

   selecting a second image portion within the digital image in which principal subject matter is depicted;
   capturing a second sequence of images of a user;
   converting the second sequence of images of the user into a second text tag for the digital image in a manner associated with the second image portion; and
   altering the rendering of the digital image in a manner associated with the second text tag.

Patentansprüche

1. Elektronisches Gerät (10), das Folgendes umfasst:

   eine Rendervorrichtung (12) zum Rendern eines Digitalbildes;
   eine Nutzerbildaufnahmeverrichtung (42) zum Aufnehmen einer Abfolge von Bildern eines Nutzers;
   ein Umwandlungsmodul (38) zum Umwandeln der Abfolge von Bildern in einen Text-Tag für das Digitalbild durch Messen einer Änderung einer Merkmalausrichtung und/oder einer Geschwindigkeit und/oder eines optischen Flusses eines dargestellten Themengegenstandes über aufeinanderfolgenden Bildern in der Abfolge; und
   einen Verarbeitungsschaltkreis (64) zum Ändern des Renderns des Digitalbildes dergestalt, dass es den Text-Tag enthält; wobei das Umwandlungsmodul (38) ein Lippenlesemodul umfasst und die Abfolge von Bildern eine Abfolge von Bewegungen der Gesichtszüge des Nutzers ist, und das Lippenlesemodul die Bewegungen der Gesichtszüge des Nutzers in den Text-Tag umwandelt; und
   das Bildanalysemodul (39) zum Bestimmen der Anordnung des Text-Tags innerhalb des Digitalbildes umfasst, wobei das Bildanalysemodul das Digitalbild in mehrere Bildabschnitte aufteilt und eine Objektdetektion an jedem Bildabschnitt ausführt, um mindestens einen Bildabschnitt zu ermitteln, in dem ein Hauptthemengegenstand enthalten ist; und der Text-Tag innerhalb des Digitalbildes getrennt von dem Bildabschnitt, der den Hauptthemengegenstand zeigt, angeordnet wird.

2. Elektronisches Gerät (10) nach Anspruch 1, wobei das Lippenlesemodul (38) die Bewegungen der Gesichtszüge des Nutzers in einen Visual-Effects-Befehl umwandelt, und der Verarbeitungsschaltkreis (64) das Rendern des Digitalbildes ändert, um den Visual-Effects-Befehl zu implementieren.


4. Elektronisches Gerät (10) nach einem der Ansprüche...
5. Elektronisches Gerät (10) nach Anspruch 4, wobei die Navigationsvorrichtung (11) eine Taste (13) zum Empfangen einer Navigationseingabe zum Auswählen des zweiten Bildabschnitts umfasst.


7. Elektronisches Gerät (10) nach Anspruch 6, wobei das Gesichtsdetektionsmodul (40) ein Augendetektionsmodul zum Detektieren der Augen des Nutzers umfasst und die Richtungseingabe eine detektierte Änderung in der Ausrichtung der Augen des Nutzers ist.

8. Verfahren zum Manipulieren eines Digitalbildes, das folgende Schritte umfasst:

   Rendern des Digitalbildes;
   Aufnehmen einer Abfolge von Bildern eines Nutzers;
   Ändern des Renderns des Digitalbildes dergestalt, dass der Text-Tag aufgenommen wird; und
   des Weiteren Folgendes umfasst: Bestimmung der Anordnung des Text-Tags innerhalb des Digitalbildes durch Teilen des Digitalbildes in mehrere Bildabschnitte, und Ausführen einer Objektdetektion an jedem Bildabschnitt, um mindestens einen Bildabschnitt zu bestimmen, in dem ein Hauptthemengegenstand enthalten ist; und


9. Verfahren nach Anspruch 8, das des Weiteren Folgendes umfasst:

   Auswählen eines zweiten Bildabschnitts innerhalb des Digitalbildes, in dem ein Hauptthemengegenstand gezeigt ist;
   Aufnehmen einer zweiten Abfolge von Bildern eines Nutzers;
   Umwandeln der zweiten Abfolge von Bildern des Nutzers in einen zweiten Text-Tag für das Digitalbild in einer Weise, die dem zweiten Bildabschnitt zugeordnet ist; und
   Ändern des Renderns des Digitalbildes in einer Weise, die dem zweiten Bildabschnitt zugeordnet ist, dergestalt, dass der zweite Text-Tag enthalten ist.

Revendications

1. Dispositif électronique (10) comprenant :

   un dispositif (12) de rendu destiné à rendre une image numérique ;
   un dispositif (42) de prise de vue destiné à prendre une séquence d’images d’un utilisateur ;
   un module (38) de conversion destiné à convertir la séquence d’images en une étiquette de texte pour l’image numérique en mesurant au moins l’un d’un changement dans l’orientation de formes, la vitesse ou le débit optique du contenu représenté sur des images successives de la séquence ; et
   un circuit (64) de traitement destiné à retoucher le rendu de l’image numérique pour inclure l’étiquette de texte,

   dans lequel le module (38) de conversion comprend un module de lecture sur les lèvres et la séquence d’images est une séquence de mouvement des traits de visage de l’utilisateur, et dans lequel le module de lecture sur les lèvres convertit le mouvement des traits de visage de l’utilisateur en étiquette de texte ; et

   comprenant en outre un module (39) d’analyse d’image destiné à déterminer l’emplacement de l’étiquette de texte à l’intérieur de l’image numérique, dans lequel le module d’analyse d’image divise l’image numérique en une pluralité de parties d’image et effectue une détection d’objet sur chaque partie d’image pour déterminer au moins une partie d’image dans laquelle le contenu principal est contenu ; et dans lequel l’étiquette de texte est placée à l’intérieur de
l'image numérique en dehors de la partie d'image représentant le contenu principal.

2. Dispositif électronique (10) selon la revendication 1, dans lequel le module (38) de lecture sur les lèvres convertit le mouvement des traits de visage de l'utilisateur en une instruction d'effets visuels, et dans lequel le circuit (64) de traitement retouche le rendu de l'image numérique pour mettre en œuvre l'instruction d'effets visuels.

3. Dispositif électronique (10) selon la revendication 2, dans lequel l'instruction d'effets visuels est au moins l'un d'un éclaircissement de l'image numérique, d'un assombrissement de l'image numérique, d'une opération de zoom avant, d'une opération de zoom arrière, d'une rotation de l'image numérique, ou de l'exécution d'une correction d'œil rouge sur un sujet contenu dans l'image numérique.

4. Dispositif électronique (10) selon l'une quelconque des revendications 1 à 3, comprenant en outre un dispositif (11) de navigation destiné à choisir une deuxième partie d'image à l'intérieur de l'image numérique, dans lequel le dispositif (42) de prise de vue d'utilisateur prend une deuxième séquence d'images de l'utilisateur, le module (38) de conversion convertit la deuxième séquence d'images en une deuxième étiquette de texte pour l'image numérique de manière associée à la deuxième partie d'image associée à la deuxième partie d'image pour inclure la deuxième étiquette de texte.

5. Dispositif électronique (10) selon la revendication 4, dans lequel le dispositif (40) de détection de visage comprend une touche (13) pour recevoir une entrée de navigation destinée à choisir la deuxième partie d'image.

6. Dispositif électronique (10) selon l'une quelconque des revendications 4 et 5, dans lequel le dispositif (11) de navigation comprend un module (40) de détection de visage destiné à détecter un mouvement de visage de l'utilisateur correspondant à une entrée directionnelle.

7. Dispositif électronique (10) selon la revendication 6, dans lequel le module (40) de détection de visage comprend un module de détection des yeux destiné à détecter les yeux de l'utilisateur, et dans lequel l'entrée directionnelle est un changement détecté de l'orientation des yeux de l'utilisateur.

8. Procédé de manipulation d'une image numérique comprenant les étapes consistant :
   à rendre l'image numérique ;
FIG. 5

Render Image 100

Text Tag or Visual Effects? 110

Activate Lip Reading Module 120a

Detect Lip Reading For Text 130a

Deactivate Lip Reading Module? 140a

Generate Text Tag 150a

Y

N

Activate Lip Reading Module 120b

Detect Lip Reading For Visual Effects Command 130b

Deactivate Lip Reading Module? 140b

Implement Visual Effects Command 150b

FIG. 6
FIG. 7
FIG. 9
FIG. 12
FIG. 13

<table>
<thead>
<tr>
<th>Photo Database 32</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Photo ID</strong></td>
</tr>
<tr>
<td>Photo #1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Photo #2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Video #3</td>
</tr>
</tbody>
</table>

FIG. 14