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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 December 2011 (08.12.2011)

PCT

(10) International Publication Number
WO 2011/150969 A1

(51) International Patent Classification:
G10L 15/26 (2006.01) *G03B 31/00* (2006.01)

(21) International Application Number:
PCT/EP2010/057747

(22) International Filing Date:
2 June 2010 (02.06.2010)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **NAXOS FINANCE SA** [LU/LU]; 42-44 Avenue de la Gare, L-1610 Luxembourg (LU).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **EZEQUIEL, Ruiz Rodriguez** [PA/PA]; Calle Aquilino de la Guardia n.8, IGRA Building, Panama, 7 (PA).

(74) Agent: **GÖKEN, Klaus G.**; Eisenführ, Speiser & Partner, Am Kaffee-Quartier 3, 28217 Bremen (DE).

(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, RO, RS, RU, 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, 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, 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: APPARATUS FOR IMAGE DATA RECORDING AND REPRODUCING, AND METHOD THEREOF

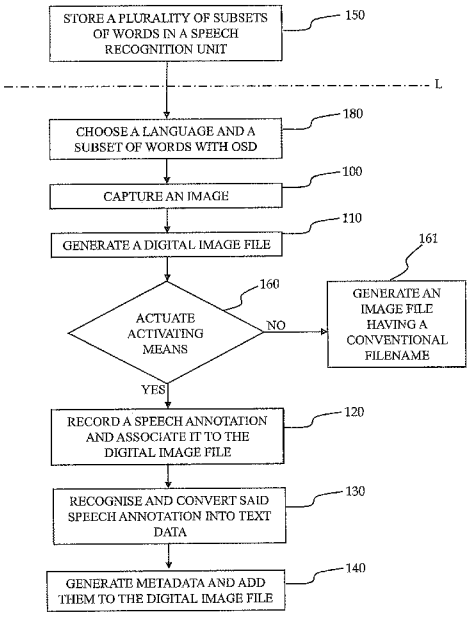

Fig. 2

(57) Abstract: The present invention relates to an apparatus (1) for image data recording and reproducing, said apparatus (1) comprising: - an imaging system (10) for capturing an image; - a signal processor (20) coupled to said imaging system (10) for processing the captured image as a digital image file; - an audio system (30) coupled to said signal processor (20) for acquiring at least one speech annotation apt to be associated with said digital image file; - a speech recognition unit (40) for recognizing said at least one speech annotation and converting the speech annotation into text data, said speech recognition unit (40) being associated to the signal processor (20) for generating metadata using the text data and adding the generated metadata to the digital image file. The invention is characterized in that said speech recognition unit (40) comprises a plurality of subsets (41) of words, each subset (41) having a limited number of words, in order to recognize and convert into text speech annotations acquired from a corresponding plurality of languages.

WO 2011/150969 A1

Naxos Finance SA

19 Rue Eugène Ruppert, L-2453 Luxembourg

APPARATUS FOR IMAGE DATA RECORDING AND REPRODUCING, AND METHOD THEREOF

DESCRIPTION

The present invention relates to an apparatus for image data recording and reproducing according to the preamble of claim 1.

The present invention also relates to a method for image data recording and reproducing, in particular for automatically creating metadata for digital image file.

Apparatuses and methods for image data recording and reproducing are well known at the state of the art; in particular, said apparatuses comprise digital cameras apt to capture images and store them on a digital medium. It should be noted that, in the present text, the words "apparatus" and/or "camera" can be used in order to relate to digital still cameras, digital video cameras, mobile telephones having integrated digital cameras, and the like.

With the apparatuses known at the state of the art, between the time an image is captured and the time it is printed or otherwise displayed, the user (that usually is also the photographer) may forget or lose access to information related to the image, such as the time at which it was captured and/or the location in which it was captured and/or the persons depicted in it.

Some digital cameras allow text, such as text representing the date and the time on which an image was captured, to be associated with a photograph; this text is typically created by the camera and superimposed on the image at a predetermined location and in a predetermined format.

Said text only contains a small amount of information, and it conveys little or no useful information to the user of the digital camera that will help him for distinguishing one image from another.

The same problem arise with the default file naming scheme, that is used in digital cameras in order to identify and track digital image files; in fact, said default file naming scheme only employs:

- a combination of letters (for example: "DSC", "IMG", "PICT", "DSCN", etc.) for indicating the type of digital image file,
- a sequence number (for example: "001", "002", etc.) appended to said indicator to identify a digital image from another, and
- a file type extension (for example, ".TIF", ".JPG", etc.) appended after the sequence number in order to identify the type of the file.

Therefore, also with the default file naming scheme the user has little or no useful information about the contents of a particular image file. In fact, the user must open and view each image file to determine if said image file contains a desired image of a person, of a place, and so on. Eventually the user can edit the naming scheme with the help of a computer, but this possibility is practically of no use when done some time after having recorded the images.

Document No. EP1876596 relates to an apparatus for image data recording and reproducing, said apparatus comprising:

- a signal processor for capturing images, processing the captured images to generate image data, and generating an image file comprising the image data;
- a speech recognition unit for recognizing speech and converting the speech into text data; and
- a controller for generating metadata using the text data and adding the generated metadata to the image file.

According to what is described in document No. EP1876596, the metadata to be included in the image file are generated by using the text data converted by the speech recognition unit, so that it is possible to add reliable metadata (such as, for example, shooting locations or persons being displayed in the image) to the image file just after the capture of the image and/or while reviewing the image file.

In addition, the name of the folder in which the image file is to be stored is generated based on the text data that is converted by using speech recognition, so that it is possible to classify the image files at a time when the image is captured.

However, it has been observed that even the apparatus described in document No. EP1876596 suffers from some drawbacks, since it is adapted to recognize and convert only one predetermined language.

In fact, the programs and software for recognizing speech and converting the speech into text data are expensive, large and very big in size, usually in the order of many megabyte (or a gigabyte) for each language that has to be recognized and converted into text; therefore, said programs and software cannot be utilized in a image data recording and reproducing apparatus without making a choice of only one predetermined language for each apparatus.

This implies that each apparatus realized in accordance with the teachings of the document No. EP1876596 needs to comprise a program apt to recognize and convert into text only one language.

This necessarily means that the apparatus cannot be versatile and eclectic, since it is necessary for the user to have an apparatus comprising a specific program for recognizing his own language, in order to convert said language into text.

This also means that the producer of the apparatus is not able to produce a single product that can be sold in different countries, where the users speak different languages. The consequence of that are an increased number of models for the same product and an increase of cost of production

In this frame, it is the main object of the present invention to overcome the above-mentioned drawbacks by providing an apparatus and a method for image data recording and reproducing which allow to recognize and convert into text a plurality of languages.

It is a further object of the present invention to provide an apparatus and a method for image data recording and reproducing conceived in a manner to be versatile and eclectic.

It is a further object of the present invention to provide a single apparatus and method for image data recording and reproducing able to recognize and convert into text a plurality of different languages.

These objects are achieved by the present invention through an apparatus and a method for image data recording and reproducing, incorporating the features set

out in the appended claims, which are intended as an integral part of the present description.

Further objects, features and advantages of the present invention will become apparent from the following detailed description and from the annexed drawings, which are supplied by way of non-limiting example, wherein:

- Fig. 1 is a block diagram of an apparatus for image data recording and reproducing, in particular a digital camera, according to the present invention;
- Fig. 2 is a block diagram illustrating a first embodiment of a method for image data recording and reproducing according to the present invention;
- Fig. 3 is a block diagram illustrating a second embodiment of a method for image data recording and reproducing according to the present invention.

In Fig. 1, reference numeral 1 designates as a whole an apparatus for image data recording and reproducing, according to the present invention.

The apparatus 1 for image data recording and reproducing according to the exemplary embodiment of the present invention may be a digital still camera, a digital video camera, a mobile telephone having an integrated or associated digital camera, and the like.

Said apparatus 1 comprises:

- an imaging system 10 for capturing an image;
- a signal processor 20 coupled to said imaging system 10 for processing the captured image as a digital image file;
- an audio system 30 coupled to said signal processor 20 for acquiring at least one speech annotation apt to be associated with said digital image file;
- a speech recognition unit 40 for recognizing said at least one speech annotation and converting the speech annotation into text data, said speech recognition unit 40 being associated to the signal processor 20 for generating metadata using the text data and adding the generated metadata to the digital image file.

Said imaging system 10 may comprise a lens/shutter assembly 11, which directs and focuses light onto a sensor 12 for capturing images of a subject; in particular, said sensor 12 can comprise one or more CCD (Charge Coupled Device) or one or more CMOS (Complementary Metal-Oxide Semiconductor).

Therefore, said signal processor 20 controls the operations of the lens/shutter assembly 11 and processes image information received from the sensor 12 for generating an image file containing the captured image in a digital format.

When the image file includes still image data, the digital image file may be in Joint Photographic Experts Group (JPEG) or Tag Image File Format (TIFF) format; when the image file includes moving image data, the digital image file may be in Moving Picture Experts Group (MPEG) format or other video formats known on the state of the art.

Moreover, as known at the state of the art, each of the image files includes an area for storing the image data and an area for storing information regarding the image. This is done in accordance to international standards. In fact there are some entities that have defined how to add metadata to image files, like:

- IPTC Information Interchange Model IIM (International Press Telecommunications Council),
- IPTC Core Schema for XMP, • XMP - Extensible Metadata Platform (an Adobe standard),
- EXIF - Exchangeable image file format, Maintained by CIPA (Camera & Imaging Products Association) and published by JEITA (Japan Electronics and Information Technology Industries Association),
- Dublin Core (Dublin Core Metadata Initiative -DCMI),
- PLUS (Picture Licensing Universal System).

As it can be seen from Fig. 1, the audio system 30 preferably comprises a microphone 31 for allowing a user to record a short audio or voice annotation, record sound for digital video recording, input voice commands, and the like. Said audio system 30 may also comprise a speaker 32.

In accordance with the present invention, said speech recognition unit 40 comprises a plurality of subsets 41 of words, each subset 41 having a limited number of words, in order to recognize and convert into text speech annotations acquired from a corresponding plurality of languages.

In particular, each subset 41 of words does not comprise a complete dictionary of words of a specific language, but each subset 41 of words comprises a relative

translation in a determined language only of a limited number of words, choosing and memorizing them at the manufacturer site only between the words more frequently used for being associated to a determined image.

In particular, said plurality of words may comprise:

- terms indicating a celebration and/or a recurrence and/or a festivity (such as, for example: "Party", "Holiday", "Baptism", "Marriage", "Birthday", "Christmas", "Easter", etc.);
- terms indicating a geographic place (such as, for example: "Sea", "Desert", "Hill", "Mountain", "Lake", etc.);
- terms indicating countries all around the world (such as "Germany", "France", "Italy", "The United States of America", "Japan", "China", "Korea" etc.) and the major cities in these countries (such as "Frankfurt", "Munich", "Paris", "Rome", "Los Angeles", "Las Vegas", "Tokyo" "Shanghai", "Hong Kong", "Macau", "Seoul"), as well as famous buildings and pieces of fine art in these cities (such as "Chinese Wall", "Casino", "Coliseum", "Tour Eiffel", etc.;
- terms indicating a season (such as: "Spring", "Summer", "Autumn", "Winter") and/or a month and/or a day of the week;
- terms indicating a number, in particular numbers from zero to nine in order to be able to compose each number;
- terms indicating a relationship with a person (such as, for example: "Brother", "Sister", "Father", "Mother ", "Grandfather", "Grandmother", "Uncle", "Aunt", "Cousin", "Friend", "Husband", "Wife");
- terms indicating the name of a person (such as, for example: "Carl", "Paul", "Peter", "John" , "Frank", "Robert", "Abbie", "Jane", "Mary", "Beth");
- terms indicating an animal (such as, for example: "Dog", "Cat", "Horse", "Bird") and/or a thing (such as, for example: "House", "Office", "Garden", "Church", "Cathedral", "Car" , "Bike").

This provision allows to obtain an apparatus and a method for image data recording and reproducing which allow to recognize and convert into text a plurality of languages, even if limited to a subset of words.

It is clear that if the word that the user wants to associate to a certain image is not provided by the limited subset of words memorized and recognizable by the apparatus, this particular word can be edited manually by making use of one of the several tools known in the state of the art for writing words: keyboards, touch screen systems, etc.

In particular, the apparatus 1 and the method according to the present invention allows to recognize speech and to convert the speech into text data without the need of using a speech recognition unit 40 expensive, large and very big in size, usually in the order of many megabyte (or a gigabyte), for each language that has to be recognized and converted into text. Therefore, this solution can be implemented in consumer products like digital still cameras, digital video cameras, mobile telephones having integrated digital cameras, and the like, without charging these products with a cost that cannot accepted by the market.

It is therefore clear that said speech recognition unit 40 can be utilized in the apparatus 1 without making a choice at the manufacturer site of a predetermined language to be used, and that said speech recognition unit 40 allows to indicate one single apparatus 1 and method conceived in such a manner to be extremely versatile and eclectic.

Preferably, said speech recognition unit 40 is associated to activating means 42 that allow the user to activate the speech recognition unit 40 in order to convert the speech annotation into text data.

In particular, said activating means 42 can be actuated by the user before the image is captured and/or displayed; otherwise, said activating means 42 can be actuated by the user after the image is captured, in particular when said image is displayed. For example, said activating means 42 may comprise a button (not shown in the drawings) preferably positioned on an external surface of the apparatus 1.

The apparatus 1 comprises also a memory 50 coupled to the signal processor 20 for storing the digital image file and/or the speech annotation and/or the speech annotation converted into text data. Said memory 50 can comprise a Random Access Memory (RAM), a Read Only Memory (ROM), an Electrically Erasable

Programmable Read Only Memory (EEPROM), or the like.

Moreover, the apparatus 1 further comprises a display 60 associated to the signal processor 20. As known, said display 60 can be used for a plurality of purposes, in particular:

- for displaying the image to be captured to the user; in this case the display 60 allows the user to center and focus the image, pose persons appearing in the image, and the like;
- for displaying a captured image, stored in the memory 50 as a digital image files;
- for displaying menus apt to convey information to the user,
- for selecting features of the apparatus 1;
- for controlling operation of the apparatus 1, and the like.

In a preferred embodiment of the present invention, said display 60 comprises an On Screen Display (OSD) system apt to choose both a language between a plurality of languages for displaying the operation of the apparatus 1, both one of said subsets 41 of words.

As said before, it is clear that the apparatus 1 can comprise input means (not shown in Fig. 1) for generating metadata in a traditional manner and in accordance to international standards, i.e. producing text data for generating metadata to be added to the digital image file; for example, said input means may comprise a keyboard or a touch screen.

Figures 2 and 3 respectively relate to a first and to a second representation of a method for image data recording and reproducing according to the present invention.

In particular, said method comprises the following steps:

- storing (step 150) at the manufacturer site a plurality of subsets 41 of a limited number of words in said speech recognition unit 40 for recognising and converting into text speech annotations acquired from a corresponding plurality of languages;
- capturing an image by means of an apparatus 1 comprising an imaging system 1 (step 100);

- processing the captured image as a digital image file through a signal processor 20 coupled to said imaging system 10 (step 110);
- recording at least one speech annotation, in particular in a memory 50, by means of an audio system 30 coupled to said signal processor 20, said at least one speech annotation being apt to be associated with said digital image file (step 120);
- recognising said at least one speech annotation and converting the speech annotation into text data by means of a speech recognition unit 40 associated to the signal processor 20 (step 130);
- generating metadata using the text data and adding the generated metadata to the digital image file (step 140).

According to the present invention, said step 130 of recognising and converting the speech annotation into text data is performed by making use of one of the plurality of subsets 41 of words stored in said speech recognition unit 40 for recognising and converting into text speech annotations acquired from a corresponding plurality of languages.

In Figs. 2 and 3, the line L indicates the fact that said step 150 of storing a plurality of subsets 41 of a limited number of words in said speech recognition unit 40 is accomplished at the manufacturer site.

In particular, the method according to the present invention is performed through the step 160 of actuating activating means 42 of the speech recognition unit 40, said activating means 42 allowing the user to activate the speech recognition unit 40 in order to convert the speech annotation into text data.

As can be seen in particular in Fig. 2, said step 160 of actuating said activating means 42 can be performed after the step 110 of processing the captured image, i.e. when said image is already recorded in a memory 50 of the apparatus 1. In this case, said step 160 can be preceded by a step 161 of generating an image file having a conventional filename. Moreover, in the case the user decides not to actuate said activating means 42, the apparatus 1 can perform the step 161 of generating an image file having a conventional filename.

Alternatively, as can be appreciated in particular from Fig. 3, said step 160 of

actuating said activating means 42 can be performed before said step 100 of capturing an image.

Moreover, the method according to the present invention comprises the further step 180 of choosing both a language between a plurality of languages for displaying the operation of the apparatus 1, both one of said subsets 41 of words by means of an On Screen Display (OSD) system comprised in said display 60.

Preferably, with reference to the method of Fig. 2, said step 180 of choosing a language and a subset of words is performed before the step 100 of capturing an image; with reference to the method of Fig. 3, said step 180 of choosing a language and a subset of words is performed after the step 160 of actuating said activating means 42.

Moreover, it must be noticed that the present invention can also be embodied as computer readable metadata on a computer readable storage medium/data. The computer readable storage medium/data is any data storage device that can store data, which can be thereafter read by a computer system. Examples of the computer readable recording medium include Electrically Erasable Programmable Read Only Memory (EEPROM), random-access memory (RAM), CD-ROMs, magnetic tapes, floppy disks, optical data storage devices, and the like.

The advantages offered by an apparatus and a method for image data recording and reproducing according to the present invention are apparent from the above description.

In particular, such advantages are due to the fact that the provision of a speech recognition unit 40 comprising a plurality of subsets 41 of words allows to recognize and convert into text a plurality of languages; in particular, this can be done without the need of using a speech recognition unit 40 expensive, large and very big in size, usually in the order of many megabyte (or a gigabyte), for each language that has to be recognized and converted into text.

It is therefore clear that clear that said speech recognition unit 40 can be utilized in the apparatus 1 without making a choice of a predetermined language that has to be recognized and converted into text, therefore, the particular realization of the speech recognition unit 40 according to the present invention

allows to indicate an apparatus 1 and a method conceived in such a manner to be versatile and eclectic.

The apparatus and method described herein by way of example may be subject to many possible variations without departing from the novelty spirit of the inventive idea; it is also clear that in the practical implementation of the invention the illustrated details may have different devices or be replaced with other technically equivalent elements, as well as providing different sequences of steps.

For instance with respect to the embodiments shown in Fig. 2 and 3, the step 180 of choosing the language can be followed immediately from the step 160 of actuating the activating means, making it manually be the user or automatically by the apparatus 1, as the consequence of having chosen both the language for displaying the operation of the apparatus 1 and one of said subsets 41 of words.

It can therefore be easily understood that the present invention is not limited to the above-described apparatus and method, but may be subject to many modifications, improvements or replacements of equivalent parts and elements without departing from the inventive idea, as clearly specified in the following claims.

* * * * * * * * *

CLAIMS

1. Apparatus (1) for image data recording and reproducing, said apparatus (1) comprising:

- an imaging system (10) for capturing an image;
- a signal processor (20) coupled to said imaging system (10) for processing the captured image as a digital image file;
- an audio system (30) coupled to said signal processor (20) for acquiring at least one speech annotation apt to be associated with said digital image file;
- a speech recognition unit (40) for recognizing said at least one speech annotation and converting the speech annotation into text data, said speech recognition unit (40) being associated to the signal processor (20) for generating metadata using the text data and adding the generated metadata to the digital image file,

characterized in that

said speech recognition unit (40) comprises a plurality of subsets (41) of words, each subset (41) having a limited number of words, in order to recognize and convert into text speech annotations acquired from a corresponding plurality of languages.

2. Apparatus (1) according to claim 1, characterized in that each subset (41) of words comprises a relative translation in a determined language only of a limited number of words, choosing and memorizing them at the manufacturer site only between the words more frequently used for being associated to a determined image.

3. Apparatus (1) according to one or more of the preceding claims, characterized in that said speech recognition unit (40) is associated to activating means (42) that allow the user to activate the speech recognition unit (40) in order to convert the speech annotation into text data.

4. Apparatus (1) according to claim 1, characterized in that said apparatus (1) comprises a memory (50) coupled to the signal processor (20) for storing the digital image file and/or the speech annotation and/or the speech annotation converted into text data.

5. Apparatus (1) according to claim 1, characterized in that said apparatus (1) comprises a display (60) associated to the signal processor (20).

6. Apparatus (1) according to claim 5, characterized in that said display (60) comprises an On Screen Display (OSD) system apt to choose both a language between a plurality of languages for displaying the operation of the apparatus (1), both one of said subsets (41) of a limited number of words.

7. Apparatus (1) according to claim 1, characterized in that said apparatus (1) comprises input means for generating metadata using said text data and coding them according to a determined international standard.

8. Method for image data recording and reproducing comprising the following steps:

- capturing an image by means of an apparatus (1) comprising an imaging system (10) (step 100);
- processing the captured image as a digital image file through a signal processor (20) coupled to said imaging system (10) (step 110);
- recording at least one speech annotation, in particular in a memory (50), by means of an audio system (30) coupled to said signal processor (20), said speech annotation being apt to be associated with said digital image file (step 120);
- recognising said speech annotation and converting at least one speech annotation into text data by means of a speech recognition unit (40) associated to the signal processor (20) (step 130);
- generating metadata using the text data and adding the generated metadata to the digital image file (step 140),

said method being characterized by the fact that

said step (130) of recognising and converting the at least one speech annotation into text data is performed by means of a step (150) of storing at the manufacturer site a plurality of subsets (41) of a limited number of words in said speech recognition unit (40) and using them for recognising and converting into text the speech annotations acquired from a corresponding plurality of languages.

9. Method according to claim 8, characterized by comprising a step (160) of

actuating activating means (42) of the speech recognition unit (40), said activating means (42) allowing the user to activate the speech recognition unit (40) in order to convert the speech annotation into text data.

10. Method according to claim 9, characterized in that said step (160) of actuating said activating means (42) is performed after the step (110) of processing the captured image.

11. Method according to claim 9, characterized in that said step (160) of actuating said activating means (42) is performed before said step (100) of capturing an image.

12. Method according to claim 11, characterized by comprising a step of characterized in that said step (160) of actuating said activating means (42) is preceded by a step (161) of generating an image file having a conventional filename.

13. Method according to claim 8, characterized by comprising a step (180) of choosing both a language between a plurality of languages for displaying the operation of the apparatus (1), both one of said subsets (41) of a limited number of words by means of an On Screen Display (OSD) system comprised in said display (60).

14. Method according to claim 13, characterized in that said step (180) of choosing a language and a subset of a limited number of words is performed before said step (100) of capturing an image.

15. Method according to claim 13, characterized in that said step (180) of choosing a language and a subset of words is performed after said step (160) of actuating said activating means (42).

16. A computer program product adapted to perform the method of any one of claims 8 to 15.

17. A computer readable storage medium/data carrier used in association with the computer program product of claim 16.

* * * * * * * * *

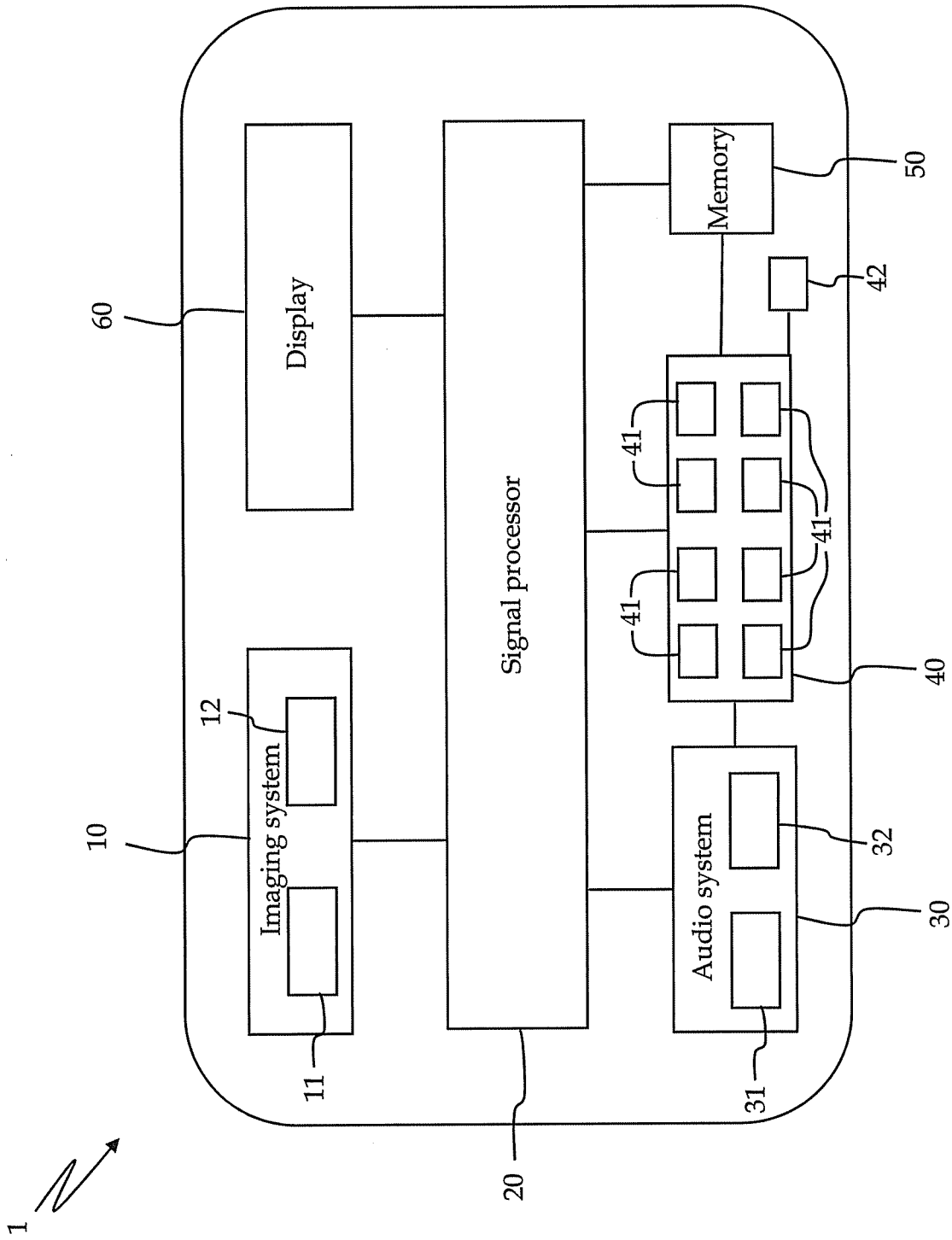
1
10
11
12
20
30
31
32
40
41
41
41
42
50
60
Display
Signal processor
Memory
Imaging system
Audio system

Fig. 1

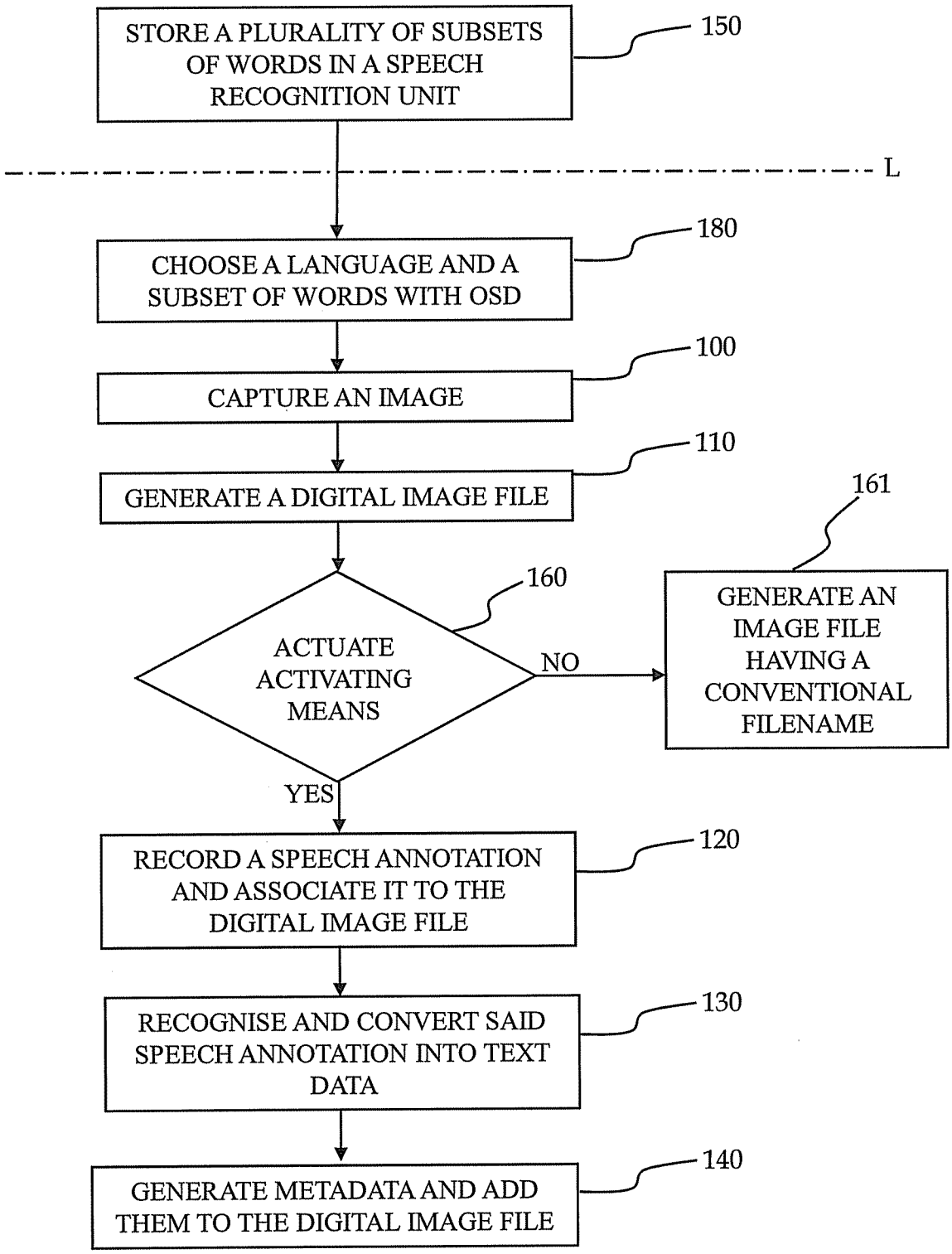
STORE A PLURALITY OF SUBSETS OF WORDS IN A SPEECH RECOGNITION UNIT
150
L
CHOOSE A LANGUAGE AND A SUBSET OF WORDS WITH OSD
180
CAPTURE AN IMAGE
100
GENERATE A DIGITAL IMAGE FILE
110
ACTUATE ACTIVATING MEANS
160
NO
GENERATE AN IMAGE FILE HAVING A CONVENTIONAL FILENAME
161
YES
RECORD A SPEECH ANNOTATION AND ASSOCIATE IT TO THE DIGITAL IMAGE FILE
120
RECOGNISE AND CONVERT SAID SPEECH ANNOTATION INTO TEXT DATA
130
GENERATE METADATA AND ADD THEM TO THE DIGITAL IMAGE FILE
140

Fig. 2

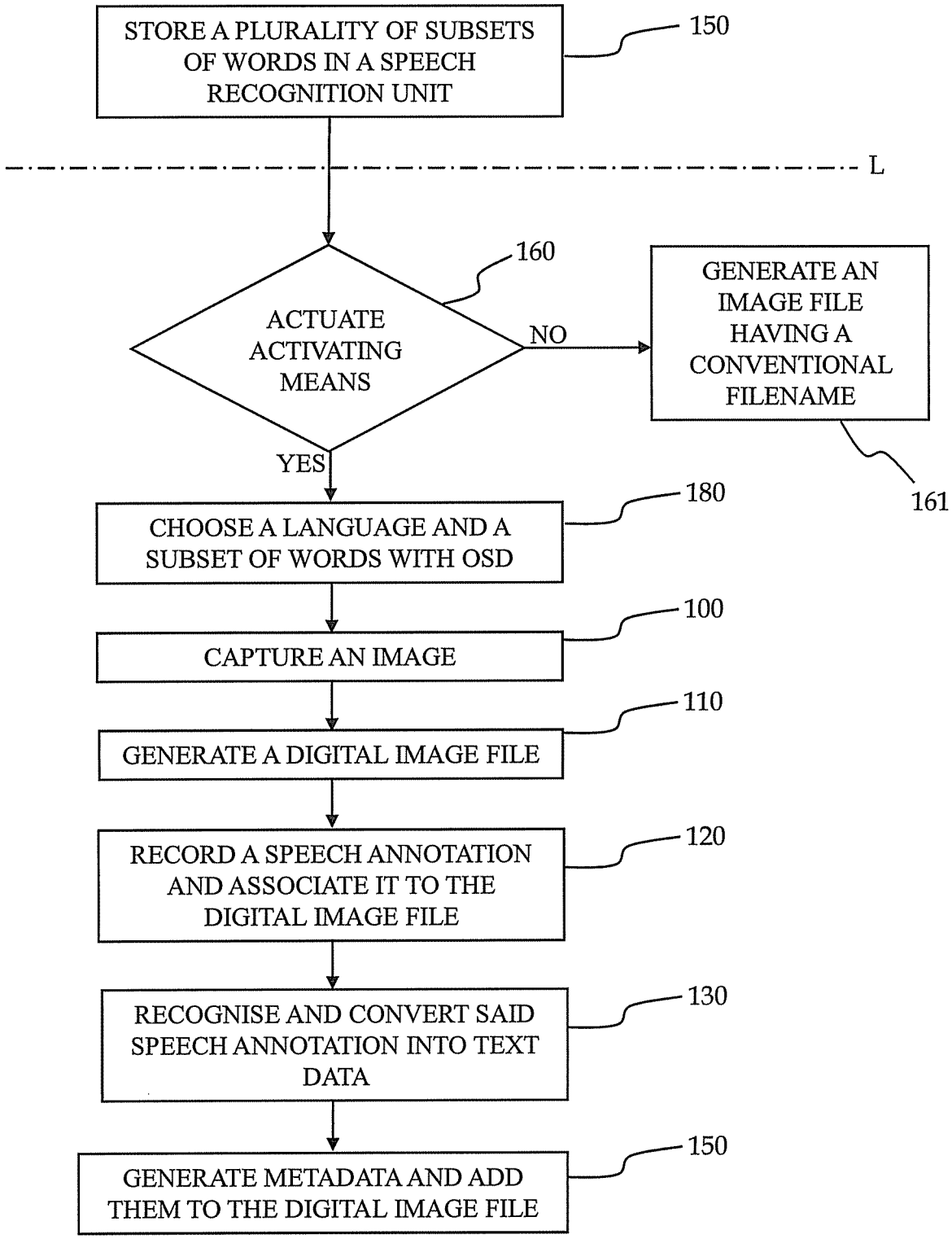
STORE A PLURALITY OF SUBSETS OF WORDS IN A SPEECH RECOGNITION UNIT
150
L
160
ACTUATE ACTIVATING MEANS
NO
GENERATE AN IMAGE FILE HAVING A CONVENTIONAL FILENAME
161
YES
180
CHOOSE A LANGUAGE AND A SUBSET OF WORDS WITH OSD
100
CAPTURE AN IMAGE
110
GENERATE A DIGITAL IMAGE FILE
120
RECORD A SPEECH ANNOTATION AND ASSOCIATE IT TO THE DIGITAL IMAGE FILE
130
RECOGNISE AND CONVERT SAID SPEECH ANNOTATION INTO TEXT DATA
150
GENERATE METADATA AND ADD THEM TO THE DIGITAL IMAGE FILE

Fig. 3

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2010/057747

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/26 G03B31/00
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)
G10L G03B

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 practical, search terms used)

EPO-Internal, WPI Data, COMPENDEX, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6 879 958 B1 (KOJIMA KIYONOBU [JP] ET AL) 12 April 2005 (2005-04-12) column 9, line 7 - line 11 column 15, line 50 - column 16, line 45 column 20, line 60 - line 64 column 23, line 4 - line 6 figures 1,17	1-17
Y	US 5 546 145 A (BERNARDI BRYAN D [US] ET AL) 13 August 1996 (1996-08-13) column 4, line 47 - column 5, line 24 column 10, line 7 - line 28 column 11, line 8 - line 14 column 11, line 32 - line 52 column 12, line 38 - line 55	1-17

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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 document 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	Date of mailing of the international search report
2 February 2011	10/02/2011

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Geißler, Christian

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2009/298529 A1 (MAHAJAN YOGESH DAGADU [US]) 3 December 2009 (2009-12-03) paragraphs [0042], [0051], [0064]	1-17
Y	US 5 991 719 A (YAZAKI MASATOMO [JP] ET AL) 23 November 1999 (1999-11-23) column 3, line 64 - column 4, line 11	1-17
Y	US 5 758 023 A (BORDEAUX THEODORE AUSTIN [US]) 26 May 1998 (1998-05-26) column 5, line 17 - line 31	1-17
A	US 2008/062280 A1 (WANG GANG [US] ET AL) 13 March 2008 (2008-03-13) paragraphs [0002], [0008], [0012], [0022] - [0024], [0031], [0032], [0037], [0039]	1-17

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2010/057747

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6879958 B1	12-04-2005	EP 1081586 A2	07-03-2001
		JP 4200607 B2	24-12-2008
		JP 2001075777 A	23-03-2001
US 5546145 A	13-08-1996	DE 69526548 D1	06-06-2002
		DE 69526548 T2	05-12-2002
		EP 0699941 A1	06-03-1996
		JP 8076288 A	22-03-1996
US 2009298529 A1	03-12-2009	WO 2009148892 A1	10-12-2009
US 5991719 A	23-11-1999	EP 0953896 A1	03-11-1999
		JP 4036528 B2	23-01-2008
		JP 11312073 A	09-11-1999
US 5758023 A	26-05-1998	AT 200590 T	15-04-2001
		AU 682380 B2	02-10-1997
		AU 7328294 A	13-02-1995
		CA 2167200 A1	26-01-1995
		DE 69427083 D1	17-05-2001
		DE 69427083 T2	06-12-2001
		EP 0708958 A1	01-05-1996
		JP 9500223 T	07-01-1997
		WO 9502879 A1	26-01-1995
US 2008062280 A1	13-03-2008	CA 2576597 A1	12-03-2008
		EP 1901284 A2	19-03-2008
		JP 2008072685 A	27-03-2008