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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 May 2015 (07.05.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/061972 A1

(51) International Patent Classification:
G06K 9/46 (2006.01)

(21) International Application Number:
PCT/CN2013/086193

(22) International Filing Date:
30 October 2013 (30.10.2013)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **MICROSOFT TECHNOLOGY LICENSING, LLC** [US/US]; One Microsoft Way, Redmond, Washington 98052 (US).

(72) Inventors; and

(71) Applicants *(for US only)*: **CAO, Xudong** [CN/CN]; c/o Microsoft Asia Pacific R & D Headquarters 14F, Building 2, No 5, Dan Ling Street, Haidian District, Beijing 100080 (CN). **WEN, Fang** [CN/CN]; c/o Microsoft Asia Pacific R & D Headquarters 14F, Building 2, No 5, Dan Ling Street, Haidian District, Beijing 100080 (CN). **SUN, Jian** [CN/CN]; c/o Microsoft Asia Pacific R & D Headquarters 14F, Building 2, No 5, Dan Ling Street, Haidian District, Beijing 100080 (CN). **CHEN, Dong** [CN/CN]; c/o Microsoft Asia Pacific R & D Headquarters 14F, Building 2, No 5, Dan Ling Street, Haidian District, Beijing 100080 (CN). **WANG, Yushun** [CN/CN]; c/o Microsoft Asia Pacific R & D Headquarters 14F, Building 2, No 5, Dan Ling Street, Haidian District, Beijing 100080 (CN).

(74) Agent: **SHANGHAI PATENT & TRADEMARK LAW OFFICE, LLC**; 435 Guiping Road, Shanghai 200233 (CN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

[Continued on next page]

(54) Title: HIGH-DIMENSIONAL FEATURE EXTRACTION AND MAPPING

(57) Abstract: A feature extraction and mapping system extracts one or more high-dimensional features associated with an object in an image and maps the one or more high-dimensional features into respective sets of low-dimensional features which may be used for recognizing or classifying the object. The feature extraction and mapping system maintains discriminative and informative characteristics of the object at the front end of feature extraction without increasing the computational and storage burden on the back end of object classification by mapping high-dimensional features into low-dimensional features.

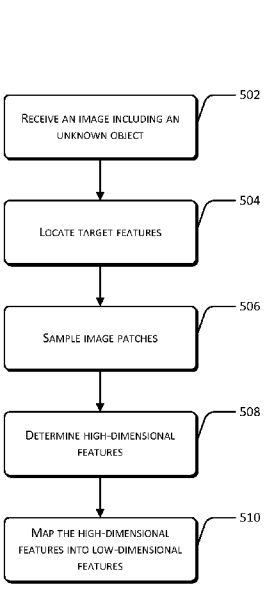

FIG. 5

WO 2015/061972 A1

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

HIGH-DIMENSIONAL FEATURE EXTRACTION AND MAPPING

BACKGROUND

[0001] Object recognition, such as face recognition, has received increasing attention from academic and industrial communities partly because of its potential applications in various technological fields such as computer vision and further because of the advance in computer technologies that make object recognition now implementable in computing devices that are accessible to common users.

[0002] An object recognition algorithm primarily includes two stages: feature extraction and object classification. Although a number of object recognition algorithms have been developed and employed in various applications, these existing object recognition algorithms rely heavily on the sophistication and complexity of object classification with a view to obtain good object recognition. However, in order to offset the high cost of object classification due to its computational complexity and storage capacity, these object recognition algorithms normally select or extract features that are primitive and/or have a low dimension (i.e., low-level or low-dimensional features). While these low-level and/or low-dimensional features may be computationally easy to extract and manipulate, these low-level and/or low-dimensional features often fail to capture discriminative or informative characteristics of different objects, thus leading to poor recognition accuracy regardless of how sophisticated and complicated an object classification algorithm may be.

SUMMARY

[0003] This summary introduces simplified concepts of feature extraction and mapping, which are further described below in the Detailed Description. This summary is not intended to identify essential features of the claimed subject matter, nor is it intended for use in limiting the scope of the claimed subject matter.

[0004] This application describes example embodiments of feature extraction and mapping. In at least one embodiment, a system may receive an image including an unknown object to be recognized. The unknown object may include a face of a person. In response to receiving the image, the system may detect or locate a plurality of target features associated with the unknown object in the image. The system may sample a plurality of image patches within a neighborhood of each target feature of the plurality of target features. In at least one embodiment, the plurality of image patches may include image patches centered at each target feature extracted at multiple levels of different resolutions or scales.

[0005] In at least one embodiment, the system may determine high-dimensional features for the plurality of image patches, and map the high-dimensional features into respective sparse sets of low-dimensional features based on a projection or mapping function. In at least one embodiment, the system may obtain or learn this projection or mapping function based on a plurality of training images and an objective function that includes constraints due to a sparse penalty and a rotational freedom. The system may provide the sparse sets of low-dimensional features to an object recognition mechanism for recognizing or classifying the unknown object.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] The detailed description is set forth with reference to the accompanying figures. In the figures, the left-most digit(s) of a reference number identifies the figure in which the reference number first appears. The use of the same reference numbers in different figures indicates similar or identical items.

[0007] FIG. 1 illustrates an example environment in which an example feature extraction and mapping system may be used.

[0008] FIG. 2 illustrates the example feature extraction and mapping system of FIG. 1 in more detail.

[0009] FIG. 3A-D illustrate different numbers of example target features detected by the feature extraction and mapping of FIG. 2 using a face as an example of an object to be recognized or classified.

[0010] FIG. 4A-D illustrate example multi-scale representations of image patches at or around a target feature that are determined by the feature extraction and mapping of FIG. 2 using a face as an example of an object to be recognized or classified.

[0011] FIG. 5 illustrates a first example method of feature extraction and mapping.

[0012] FIG. 6 illustrates a second example method of feature extraction and mapping.

[0013] FIG. 7 illustrates a first example method of learning a mapping function.

[0014] FIG. 8 illustrates a second example method of learning a mapping function.

DETAILED DESCRIPTION

Overview

[0015] As noted above, existing object recognition algorithms rely heavily on the sophistication and complexity of object classification with a view to obtain a good object recognition while sacrificing or compromising the design of features of an object to be recognized to low-level and/or low-dimensional features of the object. These low-level and/or low-dimensional features often fail to capture discriminative or informative characteristics of the object, thus failing to achieve accurate object recognition regardless of how sophisticated and complicated an object classification algorithm is used. This is especially true when the number of objects to be differentiated or recognized is large, which demands highly discriminative and informative object properties in order to obtain reasonably good object recognition.

[0016] This disclosure describes an example feature extraction and mapping system. The feature extraction and mapping system extracts high-dimensional features associated with an object from an image during a feature extraction stage and maps the high-dimensional features to respective sparse sets of low-dimensional features prior to an object classification stage. In at least one embodiment, high-dimensional features correspond to multi-resolution or multi-scale features that are extracted at one or more target features (or designated landmarks) associated with the object in the image while low-dimensional feature correspond to features that are extracted from the image without utilizing information of the target features and/or exploiting multi-resolution or multi-scale sampling at parts of the object, for example, at the target features. Additionally or

alternatively, the high-dimensional features may have a dimension of greater than or equal to fifty thousand while the low-dimensional features may have a dimension of less than or equal to ten thousand.

[0017] Processes of these two stages preserve informative and discriminative properties of the object during feature extraction on the one hand without increasing the computational burden of object classification on the other hand. Additionally, in some embodiments, the feature extraction and mapping system may further be configured to extract and map features of objects of one or more predetermined types. For example, the feature extraction and mapping system may extract or map features of objects having one or more types including, but not limited to, faces, vehicles, animals, buildings, etc.

[0018] In at least one embodiment, the feature extraction and mapping system may receive an image including an object to be recognized or classified. Upon receiving the image, the feature extraction and mapping system may detect or locate a plurality of target features associated with the object from the image. The feature extraction and mapping system may detect or locate the plurality of target features based on a generic feature detection algorithm such as edge detection, or a specific feature detection algorithm such as pattern matching specialized for a particular type of object. In either case, the plurality of target features may correspond to a set of features that are representative and/or characteristics of the object to be recognized or classified. Additionally or alternatively, the plurality of target features may be salient features of the object that are predetermined by the feature extraction and mapping system and/or a user of the feature extraction and mapping system. In at

least one embodiment, the feature extraction and mapping system may define or identify salient features of an object as features corresponding to a predetermined number of first few highest intensities obtained after applying feature detection filters (such as edge detectors, Gabor filters, etc.) to the object. By way of example and not limitation, in an event that the object to be recognized corresponds to a face, the plurality of target features may include, but are not limited to, an eye, an inner eye corner, an outer eye corner, an eye brow, a left ear, a right ear, an ear lobe, a nose, a nose tip, a chin, a mouth, a left mouth corner, a right mouth corner, etc.

[0019] Upon detecting or locating the plurality of target features, the feature extraction and mapping system may normalize or transform the object to a normalized or mean shape based on the plurality of target features. In some embodiments, the feature extraction and mapping system may compute a similarity transform to transform or normalize a current shape of the object to the normalized or mean shape. The feature extraction and mapping system may estimate the normalized or mean shape by, for example, performing a least square fitting of some or all of the plurality of target features.

[0020] After normalizing or transforming the object, the feature extraction and mapping system may sample a plurality of image patches within a neighborhood of each target feature. In at least one embodiment, the feature extraction and mapping system may sample a plurality of image patches centered at a target feature at multiple resolution or scale levels. Additionally or alternatively, the feature extraction and mapping system may sample a plurality of image patches randomly or heuristically for a target feature. For example, the feature extraction and mapping

system may sample a plurality of image patches within a predetermined distance and/or direction from the target feature.

[0021] In some embodiments, the feature extraction and mapping system may further encode the plurality of image patches using high-dimensional features. For example, the feature extraction and mapping system may divide an image patch into a plurality of cells and represent each cell using a feature descriptor. The feature extraction and mapping system may combine the feature descriptor of each cell to form a high-dimensional feature for that image patch. Examples of the feature descriptor may include LBP (local binary patterns), SIFT (scale-invariant feature transform), HOG (histogram of oriented gradients), Gabor or LE (learning-based) descriptor, or a combination thereof, etc.

[0022] Upon encoding the plurality of image patches into a plurality of high-dimensional features, the feature extraction and mapping system may map or project the plurality of high-dimensional features into respective sparse sets of low-dimensional features based on a projection or mapping function. In at least one embodiment, the feature extraction and mapping system may learn or obtain the projection or mapping function based on a plurality of training images and a supervised learning algorithm.

[0023] In some embodiments, after obtaining respective sparse sets of low-dimensional features for the plurality of high-dimensional features associated with the object, the feature extraction and mapping system may further perform object recognition or classification based on these sparse sets of low-dimensional features and an object recognition or classification algorithm. Alternatively, the feature

extraction and mapping system may provide these sparse sets of low-dimensional features to an object recognition or classification mechanism which performs object recognition or classification for the object in the image.

[0024] The described feature extraction and mapping system extracts high-dimensional features associated with an object from an image without ignoring the highly informative and discriminative characteristics of the object upfront during a feature extraction stage. The feature extraction and mapping system later maps the high-dimensional features to respective sparse sets of low-dimensional features prior to performing object recognition, thus avoiding increasing the computational burden at an object classification stage.

[0025] In the examples described herein, the feature extraction and mapping system receives an image including an object, detects or locates a plurality of target features of the object, transforms a current shape of the object, samples a plurality of image patches from the image, encodes each image patch to form a high-dimensional feature, and maps the high-dimensional feature into one or more low-dimensional features based on a pre-trained projection or mapping function. However, in other embodiments, these functions may be performed by one or more services. For example, in at least one embodiment, a detection service may receive an image, detect or locate a plurality of target features of the object, and transform a current shape of the object, while an encoding service may encode each image patch to form a high-dimensional feature. A mapping service may map the high-dimensional feature into one or more low-dimensional features based on a pre-trained projection or mapping function.

[0026] Furthermore, although in the examples described herein, the feature extraction and mapping system may be implemented as software and/or hardware installed in a single device, in other embodiments, the feature extraction and mapping system may be implemented and distributed in multiple devices or as services provided in one or more servers over a network and/or in a cloud computing architecture. Additionally, in some embodiments, the feature extraction and mapping system may be implemented as an add-on, a plug-in and/or a background process for one or more software applications such as a photo editing application, a user login application, etc.

[0027] The application describes multiple and varied implementations and embodiments. The following section describes an example framework that is suitable for practicing various implementations. Next, the application describes example systems, devices, and processes for implementing a feature extraction and mapping system.

Example Environment

[0028] FIG. 1 illustrates an example environment 100 usable to implement a feature extraction and mapping system. The environment 100 may include a feature extraction and mapping system 102. In this example, the feature extraction and mapping system 102 is described to be included in a client device 104. In other instances, the feature extraction and mapping system 102 may be implemented in whole or in part at one or more servers 106 that may communicate data with the client device 104 via a network 108. Additionally or alternatively, some or all of the

functions of the feature extraction and mapping system 102 may be included and distributed among the client device 104 and the one or more servers 106 via the network 108. For example, the one or more servers 106 may include part of the functions of the feature extraction and mapping system 102 while other functions of the feature extraction and mapping system 102 may be included in the client device 104. Furthermore, in some embodiments, some or all the functions of the feature extraction and mapping system 102 may be included in a cloud computing system or architecture.

[0029] The client device 104 may be implemented as any of a variety of conventional computing devices including, for example, a mainframe computer, a server, a notebook or portable computer, a handheld device, a netbook, a mini-computer, an ultra-computer, an Internet appliance, a tablet or slate computer, a mobile device (e.g., a mobile phone, a personal digital assistant, a smart phone, etc.), a gaming console, a set-top box, etc. or a combination thereof.

[0030] The network 108 may be a wireless or a wired network, or a combination thereof. The network 108 may be a collection of individual networks interconnected with each other and functioning as a single large network (e.g., the Internet or an intranet). Examples of such individual networks include, but are not limited to, telephone networks, cable networks, Local Area Networks (LANs), Wide Area Networks (WANs), and Metropolitan Area Networks (MANs). Further, the individual networks may be wireless or wired networks, or a combination thereof.

[0031] In at least one embodiment, the client device 104 may include one or more processors 110 coupled to memory 112. The memory 112 may include one or

more applications or services 114 (e.g., an image editing application, a login application, etc.) and program data 116. The memory 112 may be coupled to, associated with, and/or accessible to other devices, such as network servers, routers, and/or the servers 106.

[0032] In at least one embodiment, a user 118 may want to log into the client device 104. The application 114 (e.g., the login application) of the client device 104, which is supported by the feature extraction and mapping system 102, may capture an image of the user 118 via an image sensor 120 (e.g., a camera) of the client device 104. The application 114 may provide the captured image to the feature extraction and mapping system 102 which extracts high-dimensional features of a face of the user 118 from the image, and maps the high-dimensional features into respective sparse sets of low-dimensional features. The feature extraction and mapping system 102 may then provide the sparse sets of low-dimensional features to an object recognition mechanism or application which performs object recognition based on the sparse sets of low-dimensional features. Upon determining identity information associated with the face in the captured image, the object recognition system may forward the identity information to the feature extraction and mapping system 102 or directly send the identity information to the login application for determining whether to grant access of the client device 104 to the user 118 based on the identity information. Prior to capturing an image of the user 118, the feature extraction and mapping system 102 may request permission of the user 118 to capture the image of the user 118.

Example Feature Extraction and Mapping System

[0033] FIG. 2 illustrates the client device 104 that includes the feature extraction and mapping system 102 in more detail. In at least one embodiment, the client device 104 includes, but is not limited to, one or more processors 110, memory 112, one or more applications or services 114 (e.g., an image editing application, a login application, etc.) and program data 116. In some embodiments, the client device 104 may further include a network interface 200 and an input/output interface 202. The processor(s) 110 is configured to execute instructions received from the network interface 200, received from the input/output interface 202, and/or stored in the memory 112. Additionally or alternatively, some or all of the functionalities of the feature extraction and mapping system 102 may be implemented using an ASIC (i.e., Application-Specific Integrated Circuit), a FPGA (i.e., Field-Programmable Gate Array), a GPU (i.e., Graphics Processing Unit) or other hardware provided in the client device 104.

[0034] The memory 112 may include computer-readable media in the form of volatile memory, such as Random Access Memory (RAM) and/or non-volatile memory, such as read only memory (ROM) or flash RAM. The memory 112 is an example of computer-readable media. Computer-readable media includes at least two types of computer-readable media, namely computer storage media and communications media.

[0035] Computer storage media includes volatile and non-volatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program

modules, or other data. Computer storage media includes, but is not limited to, phase change memory (PRAM), static random-access memory (SRAM), dynamic random-access memory (DRAM), other types of random-access memory (RAM), read-only memory (ROM), electrically erasable programmable read-only memory (EEPROM), flash memory or other memory technology, compact disk read-only memory (CD-ROM), digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other non-transmission medium that can be used to store information for access by a computing device.

[0036] In contrast, communication media may embody computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave, or other transmission mechanism. As defined herein, computer storage media does not include communication media.

[0037] Without loss of generality, a login application is used hereinafter as an example of the application 114 with which the user 118 attempts to log into the client device 104. It is noted, however, that the present disclosure is not limited thereto and can be applied to other applications, such as an image editing application, an image viewing application, a photo catalog application, etc. Furthermore, without loss of generality, a face of the user 118 is used as an object to be processed and/or recognized as an illustrative example. The object extraction and mapping system 102, however, can be used for other types of objects including, for example, vehicles, buildings, animals, or a combination thereof, etc.

[0038] In at least one embodiment, the feature extraction and mapping system 102 may include program modules 204. The program modules 204 may include an input module 206. The input module 206 may receive or obtain an image including an object which is to be recognized or classified. In some embodiments, the input module 206 may receive or obtain the image from an image sensor 120 of the client device 104. Additionally or alternatively, the input module 206 may receive or obtain the image from the application 114 (e.g., the image editing application, etc.). Additionally or alternatively, the input module 206 may receive or obtain the image from a location in the memory 112 that is indicated by the user 118. In some embodiments, the image may include an image that is untouched or has not been edited or modified since the image is created. In some embodiments, the image may include an image that has been edited or modified by the user 118 or another user through some applications such as an image editing application.

[0039] Upon receiving or obtaining the image, a detection module 208 of the feature extraction and mapping system 102 may detect or locate a plurality of target features of the object in the image. In at least one embodiment, the plurality of target features of the object may include features or properties of the object that are determined to be representative or characteristic of the object to be recognized or classified by the feature extraction and mapping system 102 and/or the user 118. Additionally or alternatively, the plurality of target features may include features of the object that correspond to a predetermined number (e.g., ten, fifteen, twenty, fifty, etc.) of first most salient features detected by the detection module 208 based

on a feature detection algorithm specified by the feature extraction and mapping system 102 and/or the user 118.

[0040] By way of example and not limitation, if the object corresponds to a face, the plurality of target features may include, but are not limited to, an eye, an inner eye corner, an outer eye corner, an eye brow, an ear, a nose, a nose tip, a chin, a mouth, a upper middle of the mouth, a lower middle of the mouth, a left mouth corner, and/or a right mouth corner, etc. In some embodiments, the detection module 208 may employ a feature detection algorithm that may or may not be specified for the type of the object to be recognized. For example, the detection module 208 may employ a feature detection algorithm designated or specialized for faces to detect or locate the plurality of target features if the object to be recognized corresponds to a face of a person such as the user 118. Examples of designated or specialized feature detection algorithm for detecting or locating the faces may include, but are not limited to, a pattern matching algorithm with patterns corresponding to certain facial features (such as eyes, mouth, nose, nose tip, etc.), an algorithm of face alignment by explicit shape regression, etc. Details of the algorithm of face alignment by explicit shape regression can be found at a U.S. Patent Application Serial No. 13/728,584, filed on December 27, 2012, the disclosure of which is hereby incorporated in its entirety by reference.

[0041] In some instances, the detection module 208 may not be able to detect or locate all target features that may be designated by the feature extraction and mapping system 102 and/or the user 118. For example, the image may include a left side view of a face of the user 118. If the plurality of target features include features

associated with a left eye and a right eye, the detection module 208 may fail to detect or locate features associated with the right eye. In this instance, the detection module 208 may include a detection threshold which may be predefined by the feature extraction and mapping system 102 and/or the user 118, and determine that a detection of a target feature fails in response to determining that no result returned by the feature detection algorithm (such as a pattern matching algorithm with a pattern corresponding to the right eye, for example) is greater than or equal to the detection threshold. In this case, the detection module 208 may assign a default value (such as zero, one, etc.) to this target feature. In some embodiments, the detection module 208 may further flag this target feature as missing or unreliable so that the feature extraction and mapping system 102 may ignore or lessen the effect or influence of this missing target feature during subsequent operations.

[0042] Upon detecting or locating the plurality of target features, a transformation module 210 may transform or normalize the object based on the plurality of target features. By way of example and not limitation, the transformation module 210 may compute a similarity transform to transform or normalize the current shape of the object to a normalized or mean shape. In at least one embodiment, the transformation module 210 may estimate the normalized or mean shape by performing a least square fitting of some or all of the plurality of target features. An example of the similarity transform can be found at the U.S. Patent Application Serial No. 13/728,584, filed on December 27, 2012 as described above.

[0043] Additionally or alternatively, the transformation module 210 may transform or normalize the object by rescaling the object in the image to allow at least two target features of the plurality of target features to have a distance within a predetermined distance range. Additionally or alternatively, the transformation module 210 may transform or normalize the object by reorienting the object in the image to allow at least two target features of the plurality of target features to have an orientation or direction within a predetermined orientation or direction range. For example, if the object to be recognized corresponds to a face and the at least two target features correspond a left eye corner and a right eye corner, the transformation module 210 may transform or normalize the object by setting a distance between the left eye corner and the right eye corner to be a certain number of pixels (such as ten pixels, twenty pixels, thirty pixels, etc.) and/or orienting a line joining the left eye corner and the right eye corner to be horizontal.

[0044] Regardless of which algorithm is used for transforming or normalizing the object, the transformation module 210 may ignore any missing target feature as determined by the detection module 208 from determination of the similarity transform as described above. After transforming or normalizing the object, a sampling module 212 may sample or extract a plurality of image patches centered at and/or around each target feature of the plurality of target features. In at least one embodiment, the sampling module 212 may sample or extract more image patches at one target feature than another target feature based on, for example, respective sizes, degrees of relevancy (with respect to recognition of the object, for example), degrees of distinctness (among different objects of a same type or a different type,

for example), etc., of the target features. By way of example and not limitation, if the object to be recognized is a face of a person, the sampling module 212 may sample or extract more image patches at or around an eye than a mouth as the eye may include more distinct information about the person than the mouth, and/or sample or extract more image patches at or around the mouth than a nose as the mouth may have a size larger than the nose, etc.

[0045] In some embodiments, the sampling module 212 may compute a multi-resolution or multi-scale pyramid of the image and sample or extract image patches of multiple resolutions or scales centered at some or all of the plurality of target features. The number of different resolutions or scales may be set up the feature extraction and mapping system 102 and/or the user 118 in advance. In at least one embodiment, the feature extraction and mapping system 102 and/or the user 118 may set this number of different resolutions or scales differently for different target features of the plurality of target features and/or differently for different types of objects (e.g., faces, vehicles, buildings, animals, etc.) to be recognized. In some embodiments, the sampling module 212 may sample or extract image patches of a single resolution or scale centered at some or all of the plurality of target features.

[0046] Additionally or alternatively, in at least one embodiment, the sampling module 212 may sample a plurality of image patches of multiple resolutions or scales for one or more target features to have a same size. For example, the sampling module 212 may sample a plurality of multi-scale or multi-resolution image patches centered at a target feature with each image patch thereof having a same or fixed pixel size (or a number of pixels) regardless of which resolution or scale a respective

image patch corresponds to. This enables a lower-resolution or scale image patch to have a more global scope or information about the target feature as compared to a higher-resolution or scale image patch. In some embodiments, the sampling module 212 may sample a plurality of image patches of multiple resolutions or scales for one or more target features to have different sizes, e.g., a smaller size for a lower-resolution or scale image patch, etc.

[0047] Additionally or alternatively, the sampling module 212 may sample or extract a pre-set number of image patches of one or more different resolutions or scales (i.e., one or more single-scale image patches and/or one or more multi-scale image patches) within a neighborhood of one or more target features of the plurality of target features randomly or heuristically. For example, in an event that the object to be recognized is a face and the target feature is a mouth, the sampling module 212 may heuristically sample image patches more at both left and right corners or ends of the mouth than the upper middle or lower middle of the mouth.

[0048] Upon sampling the plurality of image patches, an encoding module 214 may encode each image patch into a mathematical form. In at least one embodiment, prior to encoding an image patch into a mathematical form, the encoding module 214 may divide the image patch into a plurality of cells. The encoding module 214 may divide the image patch into a plurality of non-overlapping cells or a plurality of partially overlapping cells with a degree of overlap (e.g., 5%, 10%, 20%, 50%, etc.) predetermined by the feature extraction and mapping system 102 and/or the user 118. After dividing the image patch into a plurality of cells (e.g., a plurality of 4×4 cells, a plurality of 6×6 cells, a plurality of 8×8 cells, etc.), the

encoding module 214 may encode each cell with one or more feature descriptors. Examples of the one or more feature descriptors may include, but are not limited to, LBP (local binary patterns), SIFT (scale-invariant feature transform), HOG (histogram of oriented gradients), Gabor or LE (learning-based) descriptor, etc. For example, the encoding module 214 may apply respective filters for the one or more feature descriptors, such as a Gabor filter, to a cell to obtain a Gabor feature descriptor for that cell.

[0049] In some embodiments, the encoding module 214 may further combine feature descriptors associated with one or more image patches into a high-dimensional feature. For example, the encoding module 214 may concatenate feature descriptors of an image patch into a high-dimensional feature. In at least one embodiment, a feature descriptor may be expressed in a form of a vector of S dimensions or a matrix of dimensions $N \times M$, where S, N and M are integers equal to or greater than one. The encoding module 214 may combine or concatenate k feature descriptors of an image patch into a vector of kN or a matrix of $kN \times M$ or $N \times kM$, etc.

[0050] Upon obtaining a high-dimensional feature for each image patch of the plurality of image patches, a mapping module 216 may map or project the high-dimensional feature into one or more low-dimensional features. In at least one embodiment, the mapping module 216 may map or project a high-dimensional feature into a sparse set of low-dimensional features based on a pre-trained projection or mapping function. The pre-trained projection or mapping function may include, for example, a linear projection, a sparse projection that outputs a sparse set

of low-dimensional features given an input of a high-dimensional feature, a rotated sparse projection (i.e., a sparse projection with rotational freedom), etc. The feature extraction and mapping system 102 includes a learning module 218 that is configured to obtain the pre-trained projection or mapping function based on a plurality of training images and a learning algorithm, which will be described hereinafter shortly.

[0051] After mapping the high-dimensional features into respective sparse sets of low-dimensional features (or respective one or more low-dimensional features), in at least one embodiment, the mapping module 216 may provide these low-dimensional features to the application 114 (such as an image editing application, a login application, etc.) or an object recognition mechanism that may perform recognition or classification of the object based on the low-dimensional features. In other embodiments, the feature extraction and mapping system 102 may further include a recognition module 220 that is configured to perform object classification or recognition for the unknown object in the image. The recognition module 220 may recognize or classify the object in the image based on the low-dimensional features and an object recognition or classification algorithm. Examples of the object recognition or classification algorithm may include, but are not limited to, a Bayesian inference, an artificial neural network, etc. Upon recognizing or classifying an identity and/or a type of the object in the image, the recognition module 220 may provide information of the identity and/or the type of the object to the application 114 or the client device 104 for subsequent operations such as determining whether to grant an access to the user 118 based on the recognized identity of the user 118.

Example Mapping Function and Example Learning Algorithm

[0052] As described in the foregoing description, the feature extraction and mapping system 102 includes the learning module 218 that is configured to obtain the pre-trained mapping function based on a plurality of training images and a learning algorithm. The learning module 218 may employ a supervised learning algorithm or an unsupervised learning algorithm. Examples of the supervised learning algorithm may include, but are not limited to, a Joint Bayesian algorithm, a LDA (linear discriminant analysis) algorithm, a PLDA (probabilistic linear discriminant analysis) algorithm, a SVM (support vector machine) algorithm, etc. The unsupervised learning algorithm may include, for example, an ANN (artificial neural network) algorithm, a PCA (principal component analysis) algorithm, an ICA (independent component analysis) algorithm, a SOM (self-organizing map) algorithm, etc.

[0053] By way of example and not limitation, the following description provides a learning process that employs the Joint Bayesian algorithm as a supervised learning algorithm for recognizing or classifying a face. It is noted, however, that the Joint Bayesian algorithm for recognizing or classifying a face is used as an illustrative example only and should not be construed as a limitation to the present disclosure.

[0054] In one embodiment, the learning module 218 may receive a plurality of training images with identity information of faces included in the training images via the input module 206. In response to receiving the plurality of images, the learning module 218 may employ the detection module 208 to detect or locate a plurality of

target features associated with a face included in each training image. FIG. 3A-D illustrate various number of example target features 300 of a face included in a training image that are detected by the detection module 208. FIG. 3A shows a detection of five target features of a human face that correspond to centers of left and right eyes, a nose tip, left and right mouth corners. FIG. 3B shows a detection of denser target features of the face, which include left and right corners of the left eye, left and right corners of the right eye, a nose tip, left and right nose lobes, left and right mouth corners. FIG. 3C and FIG. 3D show increasingly denser target features that are detected or located by the detection module 208. In one embodiment, the learning module 218 may choose target features that are closer to each other and/or correspond to a same or similar characteristic of the face, such as a mouth to improve performance of the target features as these target features may be complementary to each other.

[0055] After detecting or locating the plurality of target feature of each training image, the learning module 218 may instruct the transformation module 210 to transform or normalize the face in each training image based on some or all of respective target features. In at least one embodiment, the learning module 218 may further configure the sampling module 212 to sample a plurality of image patches at or around the plurality of target features from each training image. FIG. 4A-D illustrate example multi-scale representations of image patches 400 at or around a target feature 402 associated with a face in a training image. FIG. 4A shows an image patch obtained at the center of a right eye in its original image resolution or scale, with FIG. 4B-D show image patches of progressively lower resolutions or scales

obtained at the center of the right eye. As shown in FIG. 4A-D, the sampling module 212 captures a larger (or more global) scope or range around the target feature in an image patch of lower resolution or scale than an image patch of higher resolution or scale, while the image patch of higher resolution or scale includes more detailed information of the target feature than that of the image patch of lower resolution or scale.

[0056] In at least one embodiment, the learning module 218 may forward the plurality of image patches to the encoding module 214 to obtain respective high-dimensional features of each training image. Upon obtaining the high-dimensional features, the learning module 218 may apply a dimension reduction or compression algorithm, such as PCA, ICA, etc., to compress the high-dimensional features into features of lower or reduced dimensions. Additionally, in some embodiments, the learning module 218 may further apply supervised subspace learning algorithms such as LDA, Joint Bayesian algorithm, etc., to extract discriminative information from the image patches for face recognition and further reduce the dimension of the features.

[0057] By way of example and not limitation, a plurality of high-dimensional features (or a high-dimensional feature set) for a training image may be represented as $X = [x_1, x_2, \dots, x_N]$, while $Y = [y_1, y_2, \dots, y_N]$ may represent a corresponding low-dimensional feature set that is obtained based on a learning algorithm such as a subspace learning algorithm. *N* corresponds to a number of training images. In at least one embodiment, the learning module 218 may be configured to find a sparse

linear projection or mapping function, *B*, which maps *X* to *Y* within a predetermined error threshold based on an objective function as follows:

$$\min_B \|Y - B^T X\|_2^2 + \lambda \|B\|_1 \quad (1)$$

where the first term, $\|Y - B^T X\|_2^2$, corresponds to a reconstruction error and the second term, $\|B\|_1$, corresponds to a sparse penalty, i.e., a constraint that enforces sparseness. The scalar λ represents a factor that balances the first and the second terms.

[0058] In some embodiments, the learning module 218 may further introduce a degree of freedom in addition to the constraint that enforces sparseness. By way of example and not limitation, the learning module 218 may introduce a degree of rotational freedom by using, for example, a distance metric (such as Euclidean and/or Cosine) which is invariant to rotational transformation. With an additional rotation matrix, *R*, the learning module 218 may employ a rotated sparse regression, which is a linear regression with constraints related to a sparse penalty and a degree of freedom in rotation, as described in the following formulation:

$$\min_{B,R} \|R^T Y - B^T X\|_2^2 + \lambda \|B\|_1 \quad (2)$$

s.t. $$R^T R = I$$

where *I* is an identity matrix.

[0059] In at least one embodiment, the learning module 218 may employ an optimization algorithm to obtain *B* and *R*. In some embodiments, the learning module 218 may employ an alternative optimization algorithm to obtain *B* and *R*. The learning module 218 may initialize iterations of the alternative optimization algorithm by setting the matrix *B* or *R* as a predefined matrix such as an identity

matrix. For example, by setting the matrix *R* as an identity matrix, the learning module 218 may solve for the matrix B given the matrix *R* as the identity matrix. In at least one embodiment, if $\tilde{Y} = R^T Y$, the learning module 218 may employ the following objective function:

$$\min_B \left\| \tilde{Y} - B^T X \right\|_2^2 + \lambda \|B\|_1 \quad (3)$$

[0060] Since columns of the matrix *B* are independent of one another in Equation (3), the learning module 218 may optimize each column of the matrix *B* in parallel or concurrently. In at least one embodiment, the learning module 218 may employ a coordinate descent algorithm which may be initialized by values obtained in a previous iteration for obtaining values of the columns of the matrix *B*. Details of the coordinate descent algorithm can be found at J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of statistical software*, 33(1):1–22, 2010.

[0061] Upon obtaining the matrix *B* in an iteration, the learning module 218 may fix the matrix B, i.e., setting the sparse penalty term to a constant. By removing the sparse penalty term, the objective function in Equation (2) becomes:

$$\min_{B,R} \|R^T Y - B^T X\|_2^2 \quad (4)$$

s.t. $$R^T R = I$$

[0062] This problem has a closed form solution. For example, if a SVD decomposition of $YX^T B$ is UDV^T, the closed form solution of the matrix R may become:

$$R = UV^T \quad (5)$$

where *U* and *V* are matrices.

[0063] By iteratively optimizing these two sub-problems until a predetermined number of iterations and/or the predetermined error threshold is/are reached, the learning module 218 learns the rotated sparse regression to obtain the matrix *B*. With this learned matrix *B* (i.e., the projection or mapping function), the mapping module 216 may map a high-dimensional feature *x* into a low-dimensional feature *y* by $x = B^T y$. Because of the sparse penalty, the number of nonzero elements in the matrix B is reduced by orders of magnitude. As the complexity of a linear projection in computation and storage is linear to the number of nonzero elements, the computational and storage cost of the linear projection is dramatically reduced.

Example Applications

[0064] In at least one embodiment, the feature extraction and mapping system 102 may be employed in object recognition such as face verification/recognition, etc, and object attribute recognition, e.g., gender classification and age estimation, etc. The feature extraction and mapping system 102 may be deployed or included as a part of an object recognition system or an object attribute recognition system, and extract and provide discriminative and informative characteristics of an object to be recognized or classified to be used by other parts or component of the object recognition system or the object attribute recognition system. The feature extraction and mapping system 102 may pre-select or pre-defined a plurality of target features that are representative of an attribute of an object to be recognized, and extract or detect these target features of an unknown object from an image as described in the foregoing embodiments for subsequent operations by the other parts or components of the object recognition system or the object attribute recognition system.

[0065] Additionally or alternatively, the feature extraction and mapping system 102 may be deployed or included in one or more software applications such as a login application, an image editing application, etc. For example, the feature extraction and mapping system 102 may be deployed as a plug-in or an add-on program to be used by the one or more software applications. Additionally or alternatively, the feature extraction and mapping system 102 may be deployed as a background or common process or service that provides one or more functions to the one or more software applications in the client device 104 and/or the one or more servers 106 via the network 108.

Example Methods

[0066] FIG. 5 is a flow chart depicting a first example method 500 of feature extraction and mapping. FIG. 6 is a flow chart depicting a second example method 600 of feature extraction and mapping. FIG. 7 is a flow chart depicting a first example method 700 of learning a mapping function. FIG. 8 is a flow chart depicting a second example method 800 of learning a mapping function. The methods of FIG. 5-8 may, but need not, be implemented in the environment of FIG. 1 and using the device of FIG. 2. For ease of explanation, methods 500, 600, 700 and 800 are described with reference to FIGS. 1 – 2. However, the methods 500, 600, 700 and 800 may alternatively be implemented in other environments and/or using other systems.

[0067] Methods 500, 600, 700 and 800 are described in the general context of computer-executable instructions. Generally, computer-executable instructions can include routines, programs, objects, components, data structures, procedures, modules, functions, and the like that perform particular functions or implement particular abstract data types. The method can also be practiced in a distributed computing environment where functions are performed by remote processing devices that are linked through a communication network. In a distributed computing environment, computer-executable instructions may be located in local and/or remote computer storage media, including memory storage devices.

[0068] The example methods are illustrated as a collection of blocks in a logical flow graph representing a sequence of operations that can be implemented in hardware, software, firmware, or a combination thereof. The order in which the

method is described is not intended to be construed as a limitation, and any number of the described method blocks can be combined in any order to implement the method, or alternate methods. Additionally, individual blocks may be omitted from the method without departing from the spirit and scope of the subject matter described herein. In the context of software, the blocks represent computer instructions that, when executed by one or more processors, perform the recited operations. In the context of hardware, some or all of the blocks may represent application specific integrated circuits (ASICs) or other physical components that perform the recited operations.

[0069] Referring back to FIG. 5, at block 502, the feature extraction and mapping system 102 receives an image including an object to be recognized or classified.

[0070] At block 504, the feature extraction and mapping system 102 detects or locates a plurality of target features associated with the object in the image.

[0071] At block 506, the feature extraction and mapping system 102 samples a plurality of image patches within a neighborhood of each target feature of the plurality of target features. In at least one embodiment, the feature extraction and mapping system 102 may sample a plurality of multi-scale or multi-resolution image patches centered at a target feature.

[0072] At block 508, the feature extraction and mapping system 102 determines a high-dimensional feature for each image patch of the plurality of image patches.

[0073] At block 510, the feature extraction and mapping system 102 maps the high-dimensional feature into a sparse number of low-dimensional features based on a linear projection function. The linear projection function has been trained by

the feature extraction and mapping system 102 based on a plurality of training images and an objective function that include constraints of a sparse penalty and a rotational freedom.

[0074] Referring back to FIG. 6, at block 602, the feature extraction and mapping system 102 receives an image including an object to be recognized or classified.

[0075] At block 604, the feature extraction and mapping system 102 detects or locates a plurality of target features associated with the object in the image.

[0076] At block 606, the feature extraction and mapping system 102 transforms or normalizes a current shape of the object in the image to a normalized or mean shape based on some or all of the plurality of target features. In at least one embodiment, the feature extraction and mapping system 102 may transform or normalize the object by explicit shape regression. Additionally or alternatively, the feature extraction and mapping system 102 may transform or normalize the object by rescaling the object in the image to allow at least two target features of the plurality of target features to have a distance within a predetermined distance range and/or orienting the object in the image to allow at least two target features of the plurality of target features to have an orientation within a predetermined orientation range.

[0077] At block 608, the feature extraction and mapping system 102 samples a plurality of image patches within a neighborhood of each target feature of the plurality of target features. In at least one embodiment, the feature extraction and mapping system 102 may sample a plurality of multi-scale or multi-resolution image patches centered at a target feature.

[0078] At block 610, the feature extraction and mapping system 102 divides each image patch into a plurality of cells.

[0079] At block 612, the feature extraction and mapping system 102 codes each cell using a feature descriptor. Examples of the feature descriptor may include, but are not limited to, LBP (local binary patterns), SIFT (scale-invariant feature transform), HOG (histogram of oriented gradients), Gabor or LE (learning-based) descriptor.

[0080] At block 614, the feature extraction and mapping system 102 determines or obtain a high-dimensional feature for each image patch of the plurality of image patches by combining the feature descriptor of each cell to form a high-dimensional feature for each image patch.

[0081] At block 616, the feature extraction and mapping system 102 maps the high-dimensional feature into a low-dimensional feature based on a linear projection function. The linear projection function has been trained by the feature extraction and mapping system 102 based on a plurality of training images and an objective function that include constraints of a sparse penalty and a rotational freedom.

[0082] At block 618, the feature extraction and mapping system 102 recognizes the object based on the low-dimensional feature of each multi-scale image patch and an object recognition algorithm. In at least one embodiment, the object recognition algorithm may include any object recognition algorithm such as a recognition algorithm based on pattern matching, regression, etc. In some embodiments, rather than recognizing the object, the feature extraction and mapping system 102 may send the low-dimensional features to an object recognition

mechanism which is not a part of the feature extraction and mapping system 102 for object recognition.

[0083] Referring back to FIG. 7, at block 702, the feature extraction and mapping system 102 receives a plurality of high-dimensional features and a plurality of low-dimensional features. In at least one embodiment, each high-dimensional feature of the plurality of high-dimensional features may be associated with one or more low-dimensional features of the plurality of low-dimensional features.

[0084] At block 704, the feature extraction and mapping system 102 determines a projection function that maps the plurality of high-dimensional features into the plurality of low-dimensional features under constraints of sparse penalty and rotational freedom.

[0085] Referring back to FIG. 8, at block 802, the feature extraction and mapping system 102 receives a plurality of training images.

[0086] At block 804, the feature extraction and mapping system 102 extracts a plurality of target features of an object from one or more training images. In at least one embodiment, the plurality of target features of the object may be predetermined by the feature extraction and mapping system 102 for that particular type of the object.

[0087] At block 806, the feature extraction and mapping system 102 extracts or obtains a plurality of image patches within a neighborhood of each target feature of the plurality of target features. In at least one embodiment, the feature extraction and mapping system 102 may obtain a plurality of multi-scale or multi-resolution image patches within a neighborhood of each target feature.

[0088] At block 808, the feature extraction and mapping system 102 derives the plurality of high-dimensional features from the plurality of image patches. For example, the feature extraction and mapping system 102 may derive the plurality of low-dimensional features from the plurality of high-dimensional features based on a feature compression algorithm and a learning algorithm. By way of example and not limitation, the feature extraction and mapping system 102 may compress the plurality of high-dimensional features into a plurality of features of reduced dimension based on the feature compression algorithm, and extract discriminative features from the plurality of features of reduced dimension to form the plurality of low-dimensional features based on the learning algorithm.

[0089] At block 810, the feature extraction and mapping system 102 determines a projection function that maps the plurality of high-dimensional features into the plurality of low-dimensional features under constraints of sparse penalty and rotational freedom. In at least one embodiment, the projection function includes a linear regression.

[0090] At block 812, the feature extraction and mapping system 102 receives an image including an unknown object to be recognized or classified.

[0091] At block 814, the feature extraction and mapping system 102 extracts or determines one or more high-dimensional features associated with the unknown object.

[0092] At block 816, the feature extraction and mapping system 102 maps the one or more high-dimensional features into respective one or more low-dimensional

features of the plurality of low-dimensional features based on the determined projection function.

[0093] At block 818, the feature extraction and mapping system 102 recognizes the unknown object based on the respective one or more low-dimensional features and an object recognition algorithm.

[0094] Any of the acts of any of the methods described herein may be implemented at least partially by a processor or other electronic device based on instructions stored on one or more computer-readable media. By way of example and not limitation, any of the acts of any of the methods described herein may be implemented under control of one or more processors configured with executable instructions that may be stored on one or more computer-readable media such as one or more computer storage media.

Conclusion

[0095] Although embodiments have been described in language specific to structural features and/or methodological acts, it is to be understood that the claims are not necessarily limited to the specific features or acts described. Rather, the specific features and acts are disclosed as exemplary forms of implementing the claimed subject matter. Additionally or alternatively, some or all of the operations may be implemented by one or more ASICS, GPUs, or other hardware.

WHAT IS CLAIMED IS:

1. A system comprising:

one or more processors;

memory storing executable instructions that, when executed by the one or more processors, cause the one or more processors to perform acts comprising:

detecting a plurality of target features of an unknown object in an image;

sampling a plurality of image patches within a neighborhood of each target feature of the plurality of target features;

determining a high-dimensional feature for each image patch of the plurality of image patches;

mapping the high-dimensional feature into a low-dimensional feature based on a linear projection function, the linear projection function having been trained based on a plurality of training images and an objective function that include constraints of a sparse penalty and a rotational freedom.

2. A method comprising:

detecting a plurality of target features of an object in an image;

transforming the object in the image to a normalized form based on the plurality of target features;

sampling a plurality of image patches within a neighborhood of each target feature of the plurality of target features; and

determining a high-dimensional feature for each image patch of the plurality of image patches.

3. The method as recited in claim 2, wherein determining the high-dimensional feature for each image patch comprises:

dividing each image patch into multiple cells;

coding each cell of the multiple cells using a respective feature descriptor; and

combining the respective feature descriptors of the multiple cells to form the high-dimensional feature for each image patch.

4. The method as recited in claim 3, wherein the respective feature descriptor comprises LBP (local binary patterns), SIFT (scale-invariant feature transform), HOG (histogram of oriented gradients), Gabor or LE (learning-based) descriptor.

5. The method as recited in claim 2, wherein the object comprises a face and the plurality of target features comprise at least two of an eye, an eye brow, a nose, a mouth and/or an ear.

6. The method as recited in claim 2, wherein the transforming comprises rescaling the object in the image such that a distance between at least two target features of the plurality of target features is within a predetermined distance range.

7. The method as recited in claim 2, wherein the transforming comprises orienting the object in the image such that an orientation between at least two

target features of the plurality of target features is within a predetermined orientation range.

8. The method as recited in claim 2, further comprising mapping the high-dimensional feature to a low-dimensional feature based on a linear projection.

9. The method as recited in claim 8, further comprising recognizing the object based on the low-dimensional feature of each multi-scale image patch and an object recognition algorithm.

10. The method as recited in claim 9, wherein the object comprises a face of a user, and the method further comprises recognizing an identity of the user based on recognizing the object.

11. The method as recited in claim 8, wherein the linear projection is associated with a sparse penalty and a factor for a degree of freedom in rotation.

12. The method as recited in claim 8, wherein the low-dimensional feature is one of a plurality of low-dimensional features that are obtained based on a learning algorithm and a plurality of training images.

13. The method as recited in claim 12, wherein the plurality of low-dimensional features are derived by:

compressing a plurality of high-dimensional features of the plurality of training images into a plurality of features of reduced dimensions; and

extracting discriminative information from the plurality of features of reduced dimensions based on the learning algorithm.

14. One or more computer-readable media storing executable instructions that, when executed by one or more processors, cause the one or more processors to perform acts comprising:

receiving a plurality of high-dimensional features and a plurality of low-dimensional features, a high-dimensional feature of the plurality of high-dimensional features being associated with one or more low-dimensional features of the plurality of low-dimensional features; and

determining a projection function that maps the plurality of high-dimensional features into the plurality of low-dimensional features under constraints of sparse penalty and rotational freedom.

15. The one or more computer-readable media as recited in claim 14, the acts further comprising:

extracting a plurality of target features of an object from one or more training images;

obtaining a plurality of image patches within a neighborhood of each target feature of the plurality of target features;

deriving the plurality of high-dimensional features from the plurality of image patches.

16. The one or more computer-readable media as recited in claim 14, the acts further comprising deriving the plurality of low-dimensional features from the plurality of high-dimensional features based on a feature compression algorithm and a learning algorithm.

17. The one or more computer-readable media as recited in claim 16, wherein deriving the plurality of low-dimensional features comprises:

compressing the plurality of high-dimensional features into a plurality of features of reduced dimension based on the feature compression algorithm; and

extracting discriminative features from the plurality of features of reduced dimension to form the plurality of low-dimensional features based on the learning algorithm.

18. The one or more computer-readable media as recited in claim 14, the acts further comprising:

receiving one or more high-dimensional features associated with an unknown object; and

mapping the one or more high-dimensional features into respective one or more low-dimensional features of the plurality of low-dimensional features based on the projection.

19. The one or more computer-readable media as recited in claim 18, the acts further comprising recognizing the unknown object based on the respective one or more low-dimensional features and an object recognition algorithm.

20. The one or more computer-readable media as recited in claim 14, wherein the projection function comprises a linear regression.

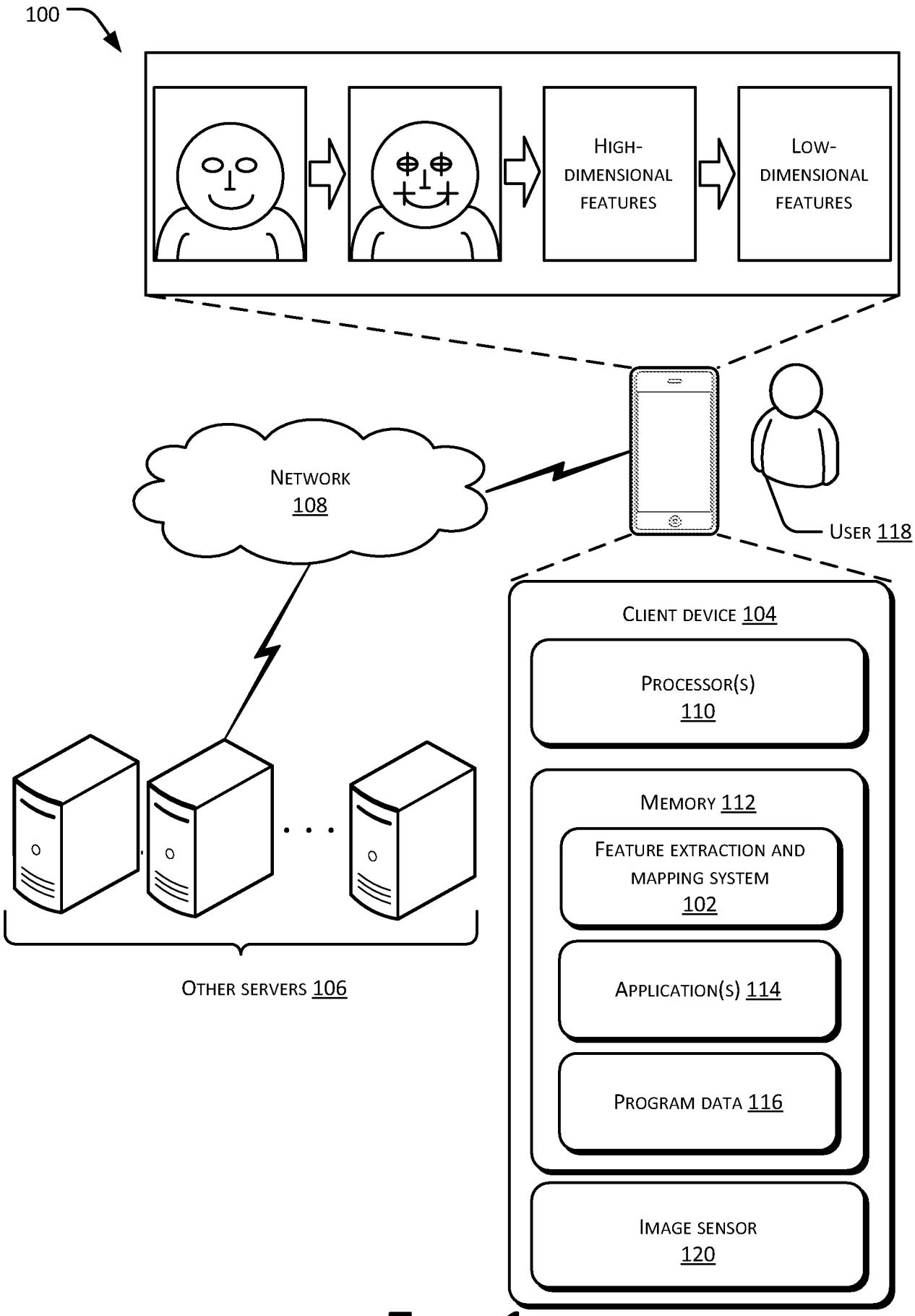
100
HIGH-DIMENSIONAL FEATURES
LOW-DIMENSIONAL FEATURES
NETWORK 108
USER 118
OTHER SERVERS 106
CLIENT DEVICE 104
PROCESSOR(S) 110
MEMORY 112
FEATURE EXTRACTION AND MAPPING SYSTEM 102
APPLICATION(S) 114
PROGRAM DATA 116
IMAGE SENSOR 120

FIG. 1

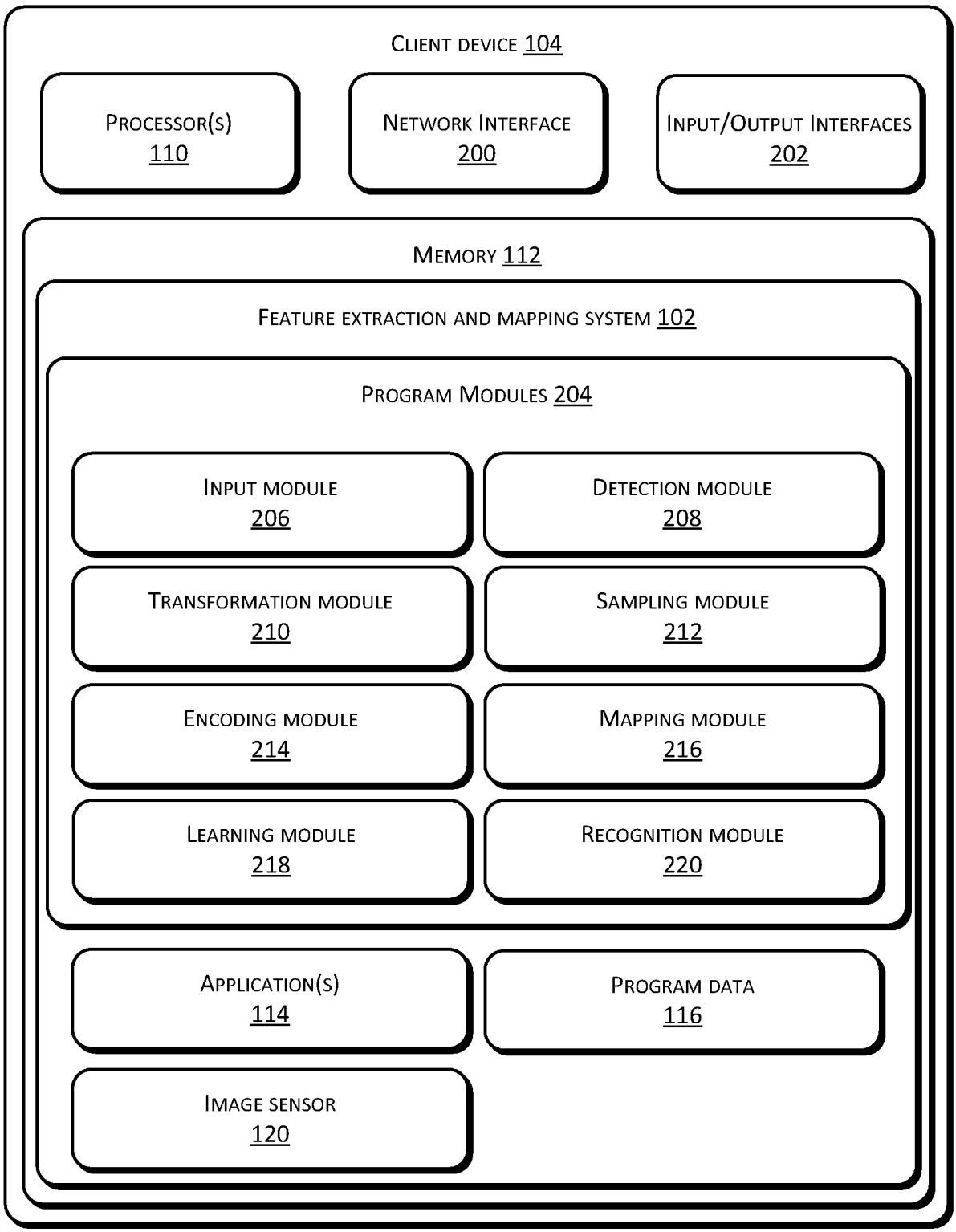
CLIENT DEVICE 104
PROCESSOR(S) 110
NETWORK INTERFACE 200
INPUT/OUTPUT INTERFACES 202
MEMORY 112
FEATURE EXTRACTION AND MAPPING SYSTEM 102
PROGRAM MODULES 204
INPUT MODULE 206
DETECTION MODULE 208
TRANSFORMATION MODULE 210
SAMPLING MODULE 212
ENCODING MODULE 214
MAPPING MODULE 216
LEARNING MODULE 218
RECOGNITION MODULE 220
APPLICATION(S) 114
PROGRAM DATA 116
IMAGE SENSOR 120

FIG. 2

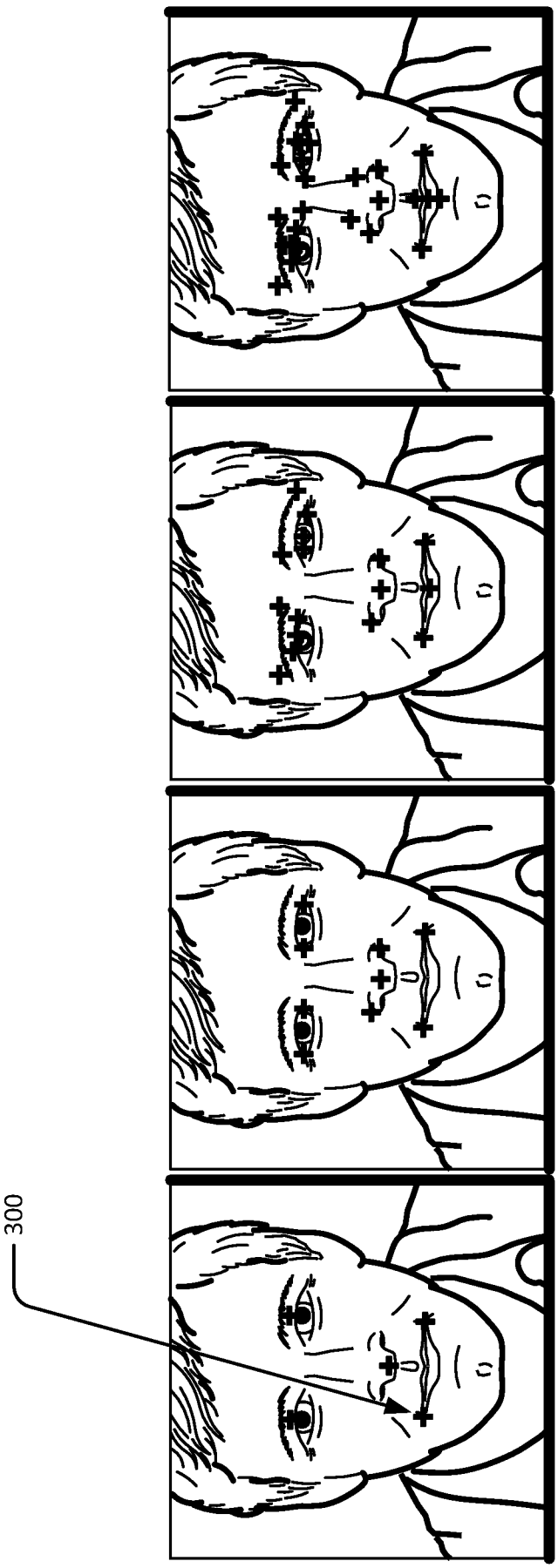

FIG. 3A FIG. 3B FIG. 3C FIG. 3D

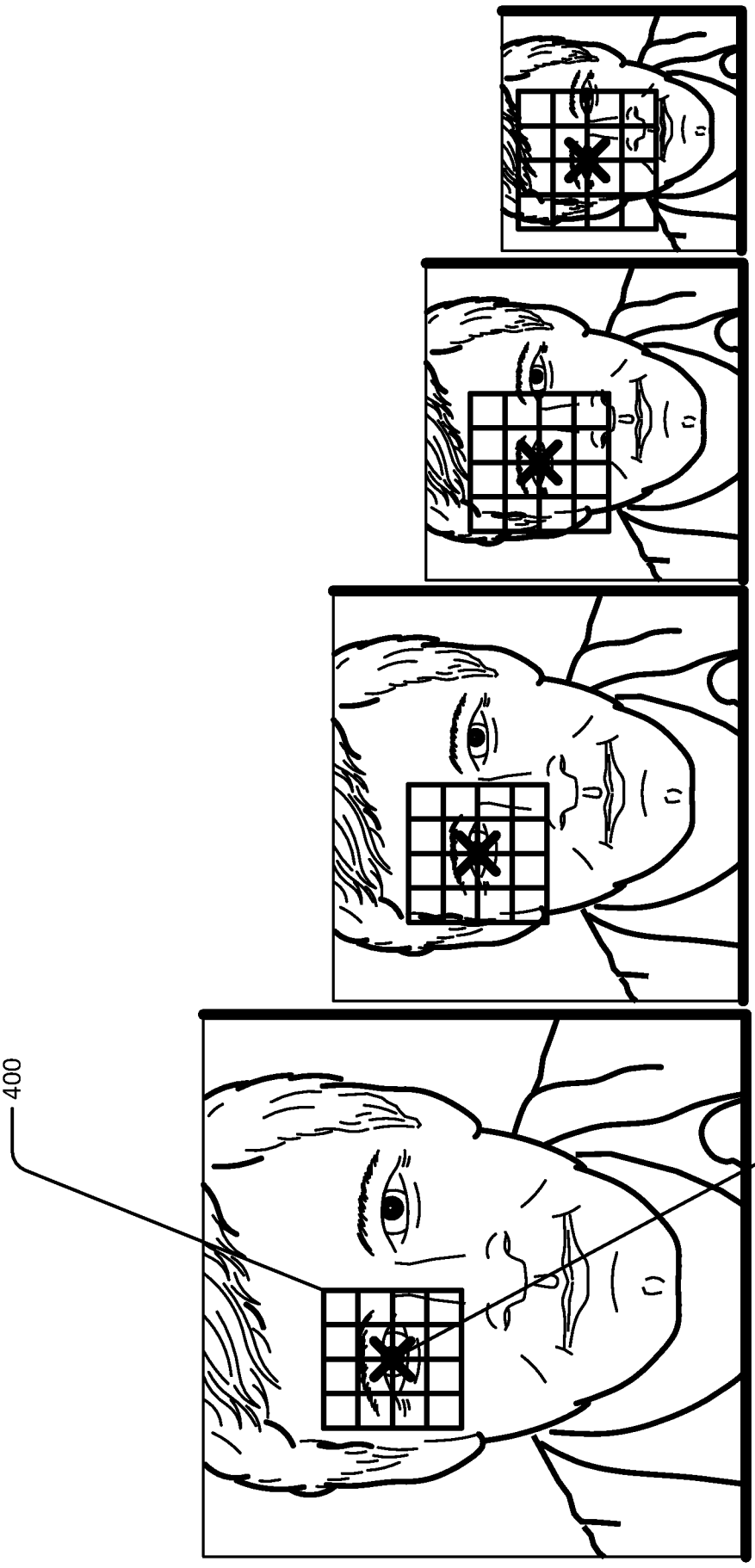

FIG. 4A FIG. 4B FIG. 4C FIG. 4D

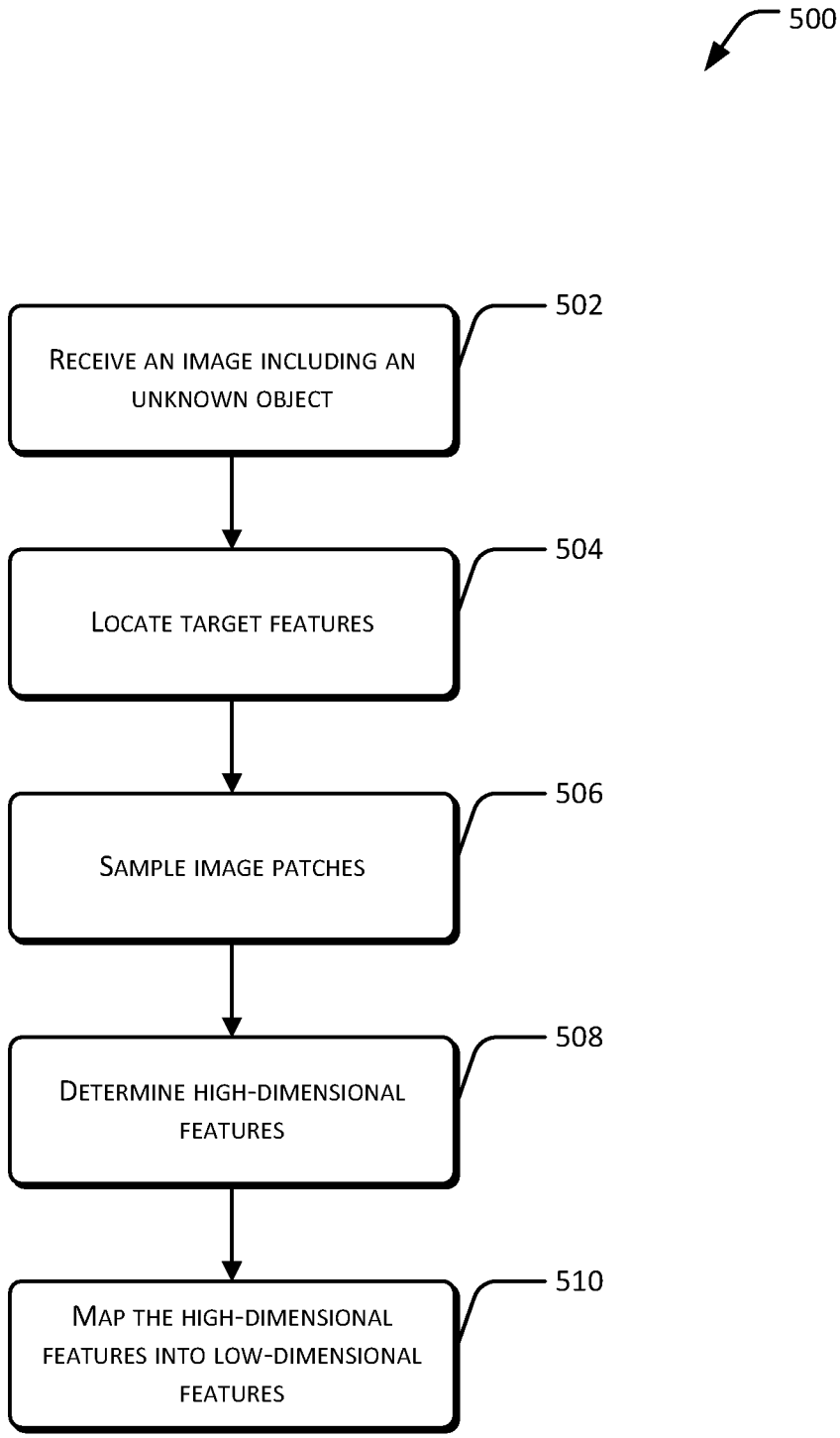
500
RECEIVE AN IMAGE INCLUDING AN UNKNOWN OBJECT
502
LOCATE TARGET FEATURES
504
SAMPLE IMAGE PATCHES
506
DETERMINE HIGH-DIMENSIONAL FEATURES
508
MAP THE HIGH-DIMENSIONAL FEATURES INTO LOW-DIMENSIONAL FEATURES
510

FIG. 5

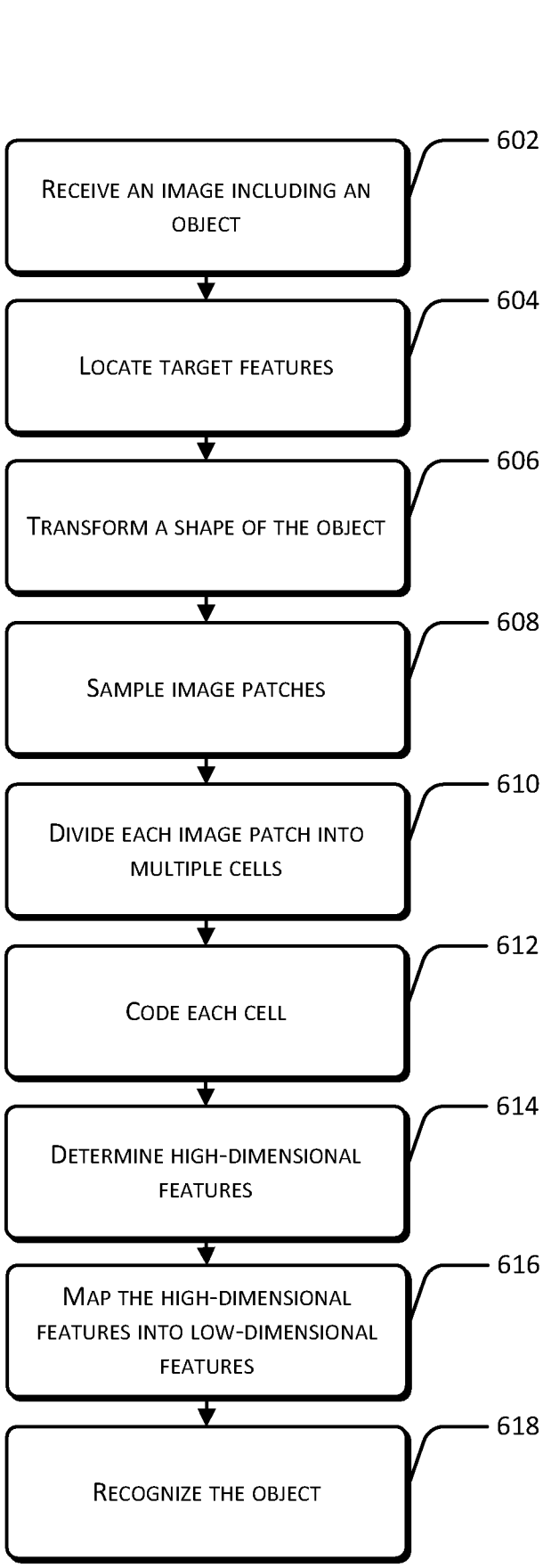
600
602
RECEIVE AN IMAGE INCLUDING AN OBJECT
604
LOCATE TARGET FEATURES
606
TRANSFORM A SHAPE OF THE OBJECT
608
SAMPLE IMAGE PATCHES
610
DIVIDE EACH IMAGE PATCH INTO MULTIPLE CELLS
612
CODE EACH CELL
614
DETERMINE HIGH-DIMENSIONAL FEATURES
616
MAP THE HIGH-DIMENSIONAL FEATURES INTO LOW-DIMENSIONAL FEATURES
618
RECOGNIZE THE OBJECT

FIG. 6

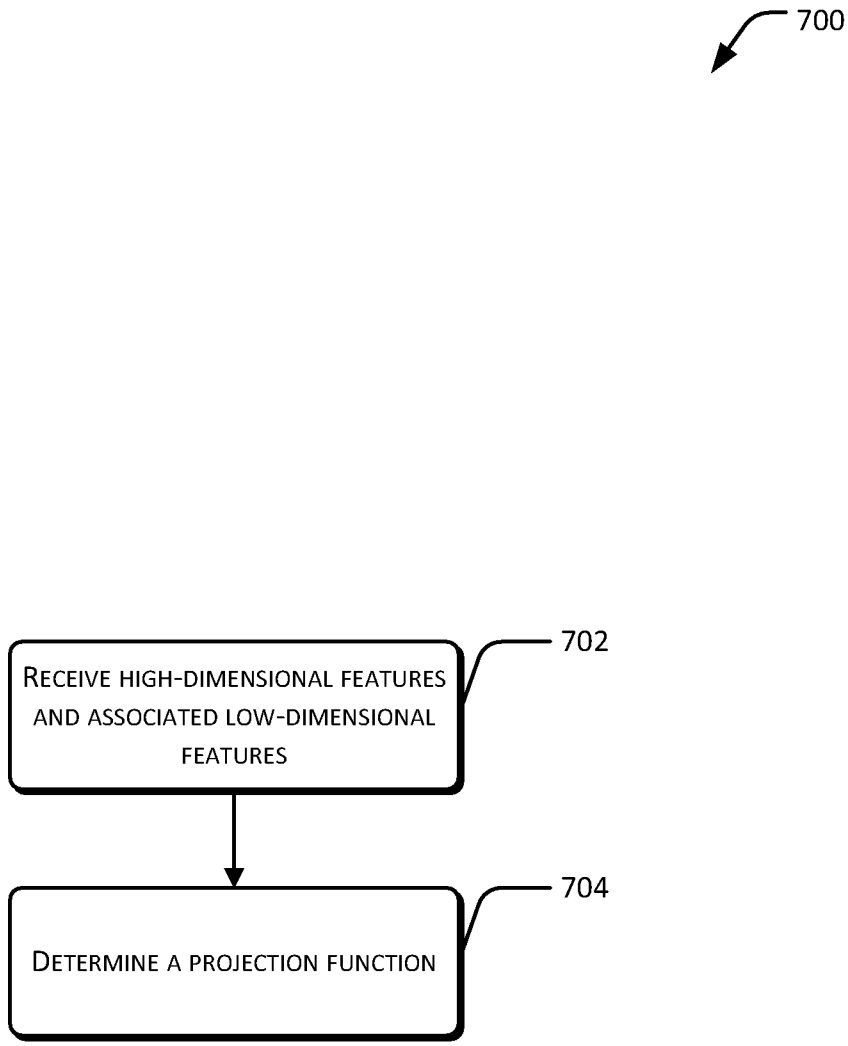
700
RECEIVE HIGH-DIMENSIONAL FEATURES AND ASSOCIATED LOW-DIMENSIONAL FEATURES
702
DETERMINE A PROJECTION FUNCTION
704

FIG. 7

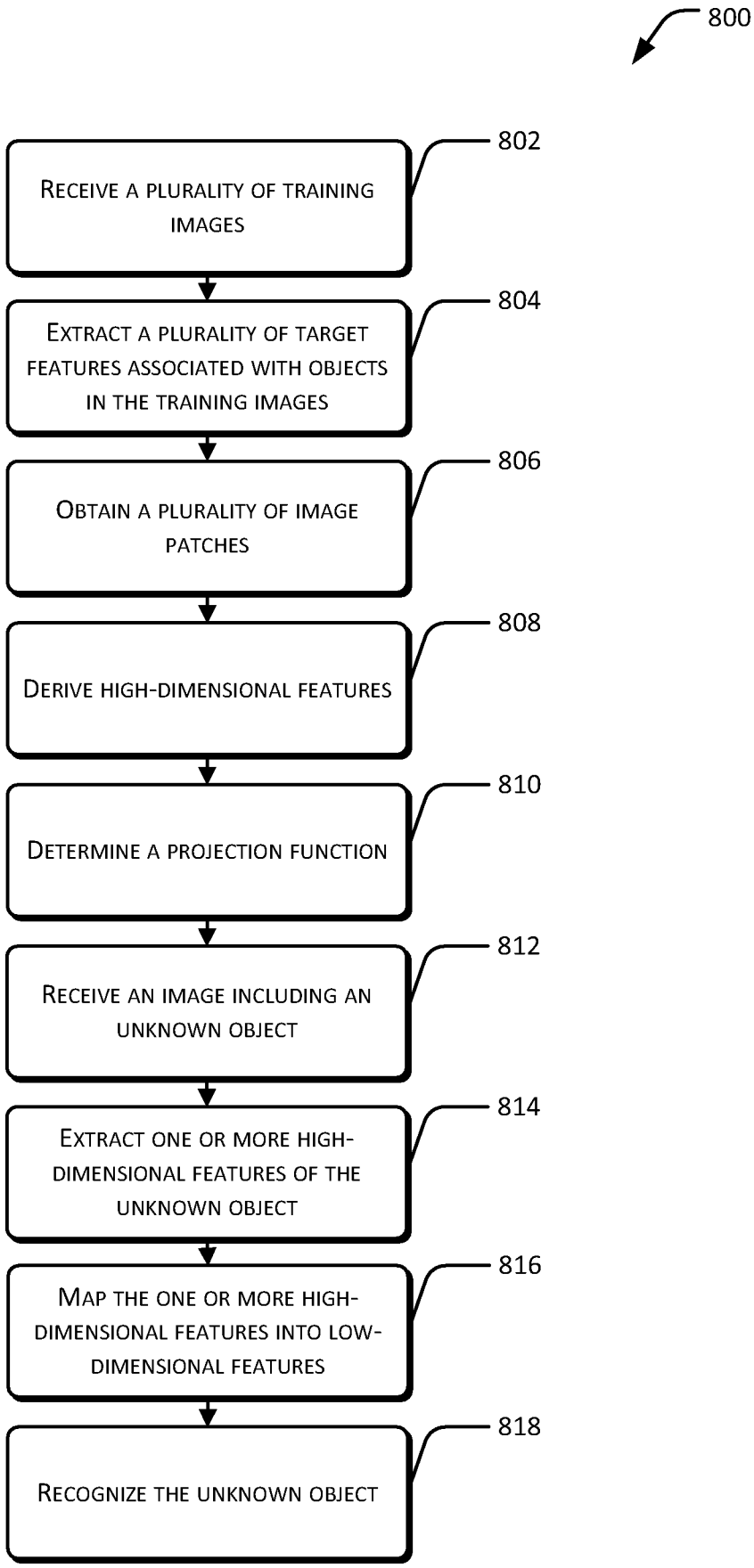
800
802
RECEIVE A PLURALITY OF TRAINING IMAGES
804
EXTRACT A PLURALITY OF TARGET FEATURES ASSOCIATED WITH OBJECTS IN THE TRAINING IMAGES
806
OBTAIN A PLURALITY OF IMAGE PATCHES
808
DERIVE HIGH-DIMENSIONAL FEATURES
810
DETERMINE A PROJECTION FUNCTION
812
RECEIVE AN IMAGE INCLUDING AN UNKNOWN OBJECT
814
EXTRACT ONE OR MORE HIGH-DIMENSIONAL FEATURES OF THE UNKNOWN OBJECT
816
MAP THE ONE OR MORE HIGH-DIMENSIONAL FEATURES INTO LOW-DIMENSIONAL FEATURES
818
RECOGNIZE THE UNKNOWN OBJECT

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/086193

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/46(2006.01)i

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)

G06K

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

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

CPRSABS,CNKI,DWPI,IEEE:face,facial,recogni+,identif+,mapp+,projection,linear,high dimension,low dimension,dimension+ 2d reduc+,multi+ 2d resolution

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011293189A1 (MICROSOFT CORPORATION) 01 December 2011 (2011-12-01) description, paragraphs [0013]-[0047] and figures 1, 2, 4 and 6	2-7
Y	US 2011293189A1 (MICROSOFT CORPORATION) 01 December 2011 (2011-12-01) description, paragraphs [0013]-[0047] and figures 1, 2, 4 and 6	1, 8-13, 15
X	CN 101038622A (SHANGHAI JIAO TONG UNIVERSITY) 19 September 2007 (2007-09-19) see description, page 2 line 14-page 5 line 3	14, 16-20
Y	CN 101038622A (SHANGHAI JIAO TONG UNIVERSITY) 19 September 2007 (2007-09-19) see description, page 2 line 14-page 5 line 3	1, 8-13, 15
A	US 2010232657A1 (WANG, J.) 19 September 2007 (2007-09-19) see the whole document	1-20

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 May 2014	**09 June 2014**

Name and mailing address of the ISA/	Authorized officer
STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA(ISA/CN) **6,Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088 China**	**ZHANG,Hui**
Facsimile No. **(86-10)62019451**	Telephone No. **(86-10)62089289**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/086193

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 2011293189A1	01 December 2011	WO 2011149976A3	26 January 2012
		WO 2011149976A2	01 December 2011
		CN 102906787A	30 January 2013
		EP 2577606A2	10 April 2013
CN 101038622A	19 September 2007	CN 100483445C	29 April 2009
US 2010232657A1	19 September 2007	US 8224042B2	17 July 2012
		JP 2010218551A	30 September 2010