



- (51) International Patent Classification:
G01C 21/00 (2006.01) *G06K 9/00* (2006.01)
- (21) International Application Number:
PCT/CN2018/105132
- (22) International Filing Date:
12 September 2018 (12.09.2018)
- (25) Filing Language: English
- (26) Publication Language: English
- (71) Applicant: **BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD.** [CN/CN]; Building 34, No. 8 Dongbeiwang West Road, Haidian District, Beijing 100193 (CN).
- (72) Inventor: **YU, Guangda**; Building 34, No. 8 Dongbeiwang West Road, Haidian District, Beijing 100193 (CN).
- (74) Agent: **METIS IP (CHENGDU) LLC**; (No. 846 South Tianfu Road) Tianfu Innovation Center, Chengdu, Sichuan 610213 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: SYSTEMS AND METHODS FOR DROWSINESS DETECTION

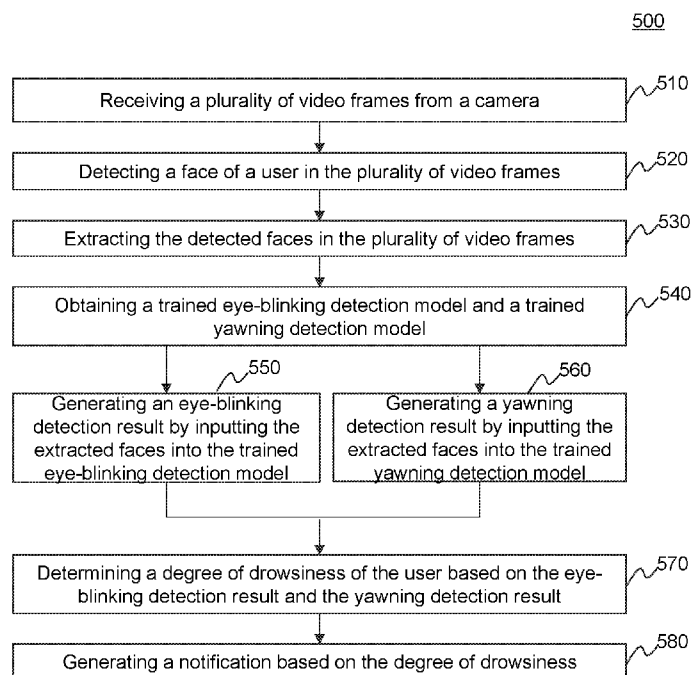


FIG. 5

(57) Abstract: A method for determining a degree of drowsiness of a driver is provided. The method may include receiving a plurality of video frames from a camera(510). The method may also include detecting a face of a driver in the plurality of video frames(520) and extracting the detected faces in the plurality of video frames(530). The method may further include obtaining a trained eye-blinking detection model and a trained yawning detection model(540). The method may further include generating an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model(550), and generating a yawning detection result by inputting the extracted faces into the trained yawning detection model(560). The method may further include determining a degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result(570). The method may further include generating a notification based on the degree of drowsiness(580).

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))*

SYSTEMS AND METHODS FOR DROWSINESS DETECTION

TECHNICAL FIELD

[0001] The present disclosure generally relates to systems and methods for user behavior, and in particular, to systems and methods for detecting drowsiness of a user.

BACKGROUND

[0002] With the dramatic increase in the number of vehicles on the road, traffic accidents occur more frequently. A large proportion of the traffic accidents are caused by drowsy driving of drivers. The existing methods of drowsy driving detection include analyzing images of the driver and determining that the driver is sleepy if the driver's eyes frequently kept closed. However, the reliability of assessment of the driver's sleepy condition may deteriorate if there is noise in the driver's images, given that the eyes of the driver may appear small in these images. It is desirable to provide systems and methods for detecting drowsiness of the driver more reliably and efficiently.

SUMMARY

[0003] According to one aspect of the present disclosure, a system is provided. The system may include at least one computer-readable storage device including a set of instructions for determining a degree of drowsiness of a driver, and at least one processor in communication with the at least one computer-readable storage device. When executing the set of instructions, the at least one processor may be configured to cause the system to receive a plurality of video frames from a camera. The at least one processor may be also configured to cause the system to detect a face of a driver in the plurality of video frames, and extract the detected faces in the plurality of video frames. The at least one processor may be further configured to cause

the system to obtain a trained eye-blinking detection model and a trained yawning detection model. The at least one processor may be further configured to cause the system to generate an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model, and generate a yawning detection result by inputting the extracted faces into the trained yawning detection model. The at least one processor may be further configured to cause the system to determine a degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result. The at least one processor may be further configured to cause the system to generate a notification based on the degree of drowsiness.

[0004] In some embodiments, the trained eye-blinking detection model may be generated by a process for training an eye-blinking detection model. The process may include obtaining a preliminary eye-blinking detection model; obtaining a plurality of training face samples; classifying the plurality of training face samples into a set of positive eye-blink training samples and a set of negative eye-blink training samples, wherein eyes in each of the positive eye-blink training samples blink and eyes in each of the negative eye-blink training samples do not blink; and training the preliminary eye-blinking detection model based on the set of positive eye-blink training samples and the set of negative eye-blink training samples to generate the trained eye-blinking detection model.

[0005] In some embodiments, the trained yawning detection model may be generated by a process for training a yawning detection model. The process may include obtaining a preliminary yawning detection model; obtaining a plurality of training face samples; classifying the plurality of training face samples into a set of positive yawn training samples and a set of negative yawn training samples, wherein a person of the face in each of the positive yawn training samples yawns and a person of the face in each of the negative yawn training samples does not yawn; and training the preliminary yawning detection model based on the set of positive yawn training samples and the

set of negative yawn training samples to generate the trained yawning detection model.

[0006] In some embodiments, the at least one processor may be further configured to cause the system to determine whether an angle between a direction perpendicular to a face of the driver in the at least one video frame of the plurality of video frames and a capturing direction of the camera is greater than a threshold for at least one video frame of the plurality of video frames. In response to the determination that the angle between the direction perpendicular to the face of the driver and the capturing direction of the camera is greater than the threshold, the at least one processor may be further configured to cause the system to discard the at least one video frame of the plurality of video frames from the plurality of video frames to be inputted into the trained eye-blinking detection model or the trained yawning detection model.

[0007] In some embodiments, to determine the angle between the direction perpendicular to the face of the driver and the capturing direction of the camera, the at least one processor may be configured to cause the system to input the extracted face in each of the plurality of video frames into a trained angle determination model to generate a result that whether the angle between the direction perpendicular to the face of the driver and the capturing direction of the camera is greater than the threshold.

[0008] In some embodiments, the trained angle determination model may be generated by a process for training an angle determination model. The process may include obtaining a preliminary angle determination model; obtaining a plurality of training face samples; classifying the plurality of training face samples into a set of positive angle training samples and a set of negative angle training samples, wherein the angle between the direction perpendicular to a face of a driver and the capturing direction of a camera in each positive angle training sample is greater than 60, and the angle between the direction perpendicular to a face of the driver and the capturing direction

of the camera in each negative angle training sample is less than or equal to 60 degrees; and training the preliminary angle determination model based on the set of positive training samples and the set of negative training samples to generate the trained angle determination model.

[0009] In some embodiments, to determine the degree of drowsiness based on the eye-blinking detection result and the yawning detection result, the at least one processor may be further configured to cause the system to determine a frequency of eye-blinking by the driver based on the eye-blinking detection result; determine a frequency of yawning by the driver based on the yawning detection result; and determine the degree of drowsiness based on the frequency of eye-blinking and the frequency of yawning.

[0010] In some embodiments, to determine the frequency of eye-blinking of the driver based on the eye-blinking detection result, the at least one processor may be further configured to cause the system to determine a number count of one or more blinks by the driver detected in the plurality of video frames based on the eye-blinking detection result; determine a total time length of the plurality of video frames; and determine the frequency of eye-blinking based on a number count of the one or more blinks and the total time length of the plurality of video frames.

[0011] In some embodiments, to determine the frequency of yawning of the driver based on the plurality of yawning detection results, the at least one processor may be further configured to cause the system to determine the number of one or more yawns by the driver detected in the plurality of successive frames based on the plurality of yawning detection results; determine a total time length of the plurality of video frames; and determine the frequency of yawning based on a number count of the one or more yawns of the driver and the total time length of the plurality of video frames.

[0012] In some embodiments, the notification may include at least one of a sound, a vibration, or a light.

[0013] In some embodiments, a loudness or frequency of the sound, a

strength of the vibration, or an intensity or frequency of the light may depend on the degree of drowsiness.

[0014] According to another aspect of the present disclosure, a method for determining a degree of drowsiness of a driver is provided. The method may include receiving a plurality of video frames from a camera. The method may also include detecting a face of a driver in the plurality of video frames. The method may further include extracting the detected faces in the plurality of video frames. The method may further include obtaining a trained eye-blinking detection model and a trained yawning detection model. The method may further include generating an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model, and generating a yawning detection result by inputting the extracted faces into the trained yawning detection model. The method may further include determining a degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result. The method may further include generating a notification based on the degree of drowsiness.

[0015] According to another aspect of the present disclosure, a system for determining a degree of drowsiness of a driver is provided. The system may include an acquisition module, a detection module, an extraction module, a generation module and a determination module. The acquisition module may be configured to receive a plurality of video frames from a camera, and obtain a trained eye-blinking detection model and a trained yawning detection model. The detection module may be configured to detect a face of a driver in the plurality of video frames. The extraction module may be configured to extract the detected faces in the plurality of video frames. The generation module may be configured to generate an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model, generate a yawning detection result by inputting the extracted faces into the trained yawning detection model, and generate a notification based on a degree of drowsiness. The determination module may be configured to

determine the degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result.

[0016] According to another aspect of the present disclosure, a non-transitory computer-readable storage medium embodying a computer program product. The computer program product including instructions may be configured to cause a computing device to receive a plurality of video frames from a camera. The computer program product including instructions may be also configured to cause the computing device to detect a face of a driver in the plurality of video frames. The computer program product including instructions may be further configured to cause the computing device to extract the detected faces in the plurality of video frames. The computer program product including instructions may be further configured to cause the computing device to obtain a trained eye-blinking detection model and a trained yawning detection model. The computer program product including instructions may be further configured to cause the computing device to generate an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model, and generate a yawning detection result by inputting the extracted faces into the trained yawning detection model. The computer program product including instructions may be further configured to cause the computing device to determine a degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result. The computer program product including instructions may be further configured to cause the computing device to generate a notification based on the degree of drowsiness.

[0017] Additional features will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following and the accompanying drawings or may be learned by production or operation of the examples. The features of the present disclosure may be realized and attained by practice or use of various

aspects of the methodologies, instrumentalities, and combinations set forth in the detailed examples discussed below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The present disclosure is further described in terms of exemplary embodiments. These exemplary embodiments are described in detail with reference to the drawings. These embodiments are non-limiting exemplary embodiments, in which like reference numerals represent similar structures throughout the several views of the drawings, and wherein:

[0019] FIG. 1 is a schematic diagram illustrating an exemplary drowsiness detection system according to some embodiments of the present disclosure;

[0020] FIG. 2 is a schematic diagram illustrating exemplary hardware and/or software components of a computing device according to some embodiments of the present disclosure;

[0021] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of a mobile device on which a terminal may be implemented according to some embodiments of the present disclosure;

[0022] FIG. 4 is a block diagram illustrating an exemplary processing engine according to some embodiments of the present disclosure;

[0023] FIG. 5 is a flowchart illustrating an exemplary process for determining a degree of drowsiness of a driver according to some embodiments of the present disclosure;

[0024] FIG. 6 is a flowchart illustrating an exemplary process for determining a frequency of eye-blinking according to some embodiments of the present disclosure;

[0025] FIG. 7 is a flowchart illustrating an exemplary process for determining a frequency of yawning according to some embodiments of the present disclosure;

[0026] FIG. 8 is a schematic diagram illustrating an exemplary blink according to some embodiments of the present disclosure;

[0027] FIG. 9 is a schematic diagram illustrating an exemplary yawn according to some embodiments of the present disclosure;

[0028] FIG. 10 is a schematic diagram illustrating exemplary images of yawning according to some embodiments of the present disclosure;

[0029] FIG. 11 is a schematic diagram illustrating an exemplary model according to some embodiments of the present disclosure; and

[0030] FIG. 12 is a schematic diagram illustrating an exemplary automobile data recorder for detecting drowsiness of a driver according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0031] The following description is presented to enable any person skilled in the art to make and use the present disclosure and is provided in the context of a particular application and its requirements. Various modifications to the disclosed embodiments will be readily apparent to those skilled in the art, and the general principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the present disclosure. Thus, the present disclosure is not limited to the embodiments shown but is to be accorded the widest scope consistent with the claims.

[0032] The terminology used herein is for the purpose of describing particular example embodiments only and is not intended to be limiting. As used herein, the singular forms “a,” “an,” and “the” may be intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprise,” “comprises,” and/or “comprising,” “include,” “includes,” and/or “including,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[0033] These and other features, and characteristics of the present

disclosure, as well as the methods of operation and functions of the related elements of structure and the combination of parts and economies of manufacture, may become more apparent upon consideration of the following description with reference to the accompanying drawings, all of which form a part of this disclosure. It is to be expressly understood, however, that the drawings are for the purpose of illustration and description only and are not intended to limit the scope of the present disclosure. It is understood that the drawings are not to scale.

[0034] The flowcharts used in the present disclosure illustrate operations that systems implement according to some embodiments in the present disclosure. It is to be expressly understood, the operations of the flowchart may be implemented not in order. Conversely, the operations may be implemented in inverted order, or simultaneously. Moreover, one or more other operations may be added to the flowcharts. One or more operations may be removed from the flowcharts.

[0035] Moreover, while the systems and methods disclosed in the present disclosure are described primarily regarding driver drowsiness detection service, it should also be understood that this is only one exemplary embodiment. The system or method of the present disclosure may be applied to any other kind of drowsiness detection. For example, in order to raise the attention of a student or an audience that attends a class or a lecture, the systems or methods of the present disclosure may be used to detect the degree of drowsiness of the student or the audience. As another example, in order to reduce the risk of mechanical injury of a mechanical engineer or operator, the systems or methods of the present disclosure may be used to detect the degree of drowsiness of the mechanical engineer or the operator.

[0036] The term “user” in the present disclosure may refer to an individual, an entity or a tool that may be detected in the drowsiness detection system. For example, the user may be a driver, an operator, a student, a worker, or

the like, or a combination thereof. The terms “drowsiness” and “degree of drowsiness” may be used herein interchangeably.

[0037] An aspect of the present disclosure relates to systems and methods for detecting a drowsy driving of a driver. The drowsy driving of the driver may be judged based on the degree of the drowsiness of the driver. The degree of the drowsiness may be determined based on an eye-blinking detection result and a yawning detection result. The eye-blinking detection result may be generated by inputting face image(s) of the driver to a trained eye-blinking detection model. The yawning detection result may be generated by inputting the face image(s) of the driver to a trained yawning detection model. As used herein, the face image(s) of the driver may be extracted from a plurality of real-time video frames that include the face of the driver. After the determination of the degree of the drowsiness, a notification (e.g., sound, vibration, light) may be generated to warn the driver. The type and/or strength of the notification may depend on the degree of drowsiness.

[0038] FIG. 1 is a schematic diagram illustrating an exemplary drowsiness detection system 100 according to some embodiments of the present disclosure. For example, the drowsiness detection system 100 may be used to detect a drowsy driving of a driver. As another example, the drowsiness detection system 100 may be used to determine a degree of drowsiness of a mechanical engineer or operator who is working on a production line. As a further example, the drowsiness detection system 100 may be used to determine a degree of drowsiness of a student (or an audience) that attends a class (or a lecture).

[0039] The drowsiness detection system 100 may be mounted on a vehicle or a component thereof (e.g., an automobile data recorder in a vehicle). The drowsiness detection system 100 may include a processing device 110, a terminal 120, an image capturing device 130, a network 140 and a storage 150.

[0040] In some embodiments, the processing device 110 may be a single

processing device or a processing device group. The processing device group may be centralized, or distributed (e.g., the processing device 110 may be a distributed system). In some embodiments, the processing device 110 may be local or remote. For example, the processing device 110 may access information and/or data stored in the terminal 120, the image capturing device 130, and/or the storage 150 via the network 140. As another example, the processing device 110 may be directly connected to the terminal 120, the image capturing device 130, and/or the storage 150 to access stored information and/or data. In some embodiments, the processing device 110 may be implemented on a cloud platform. Merely by way of example, the cloud platform may include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, or the like, or a combination thereof. In some embodiments, the processing device 110 may be implemented on a computing device 200 having one or more components illustrated in FIG. 2 in the present disclosure.

[0041] In some embodiments, the processing device 110 may include a processing engine 112. The processing engine 112 may process information and/or data related to the drowsiness detection to perform one or more functions described in the present disclosure. For example, the processing engine 112 may determine a degree of drowsiness of the driver based on an eye-blinking detection result and a yawning detection result. In some embodiments, the processing engine 112 may include one or more processing engines (e.g., single-core processing engine(s) or multi-core processor(s)). Merely by way of example, the processing engine 112 may include a central processing unit (CPU), an application-specific integrated circuit (ASIC), an application-specific instruction-set processor (ASIP), a graphics processing unit (GPU), a physics processing unit (PPU), a digital signal processor (DSP), a field-programmable gate array (FPGA), a programmable logic device (PLD), a controller, a microcontroller unit, a reduced instruction-set computer (RISC), a microprocessor, or the like, or a

combination thereof.

[0042] In some embodiments, the terminal 120 may include a tablet computer 120-1, a laptop computer 120-2, a built-in device in a vehicle 120-3, a mobile device 120-4, or the like, or a combination thereof. In some embodiments, the mobile device 120-4 may include a smart home device, a wearable device, a smart mobile device, an augmented reality device, or the like, or a combination thereof. In some embodiments, the wearable device may include a smart bracelet, a smart footgear, smart glasses, a smart helmet, a smart watch, smart clothing, a smart backpack, a smart accessory, or the like, or a combination thereof. In some embodiments, the smart mobile device may include a smartphone, a personal digital assistant (PDA), a gaming device, a navigation device, a point of sale (POS) device, or the like, or a combination thereof. In some embodiments, the built-in device in the vehicle 120-3 may include an onboard computer, an automobile data recorder, an onboard human-computer interaction (HCI) system, an onboard television, etc. The terminal 120 may be a device with positioning technology for locating the position of the terminal 120. The terminal 120 may generate a notification based on the degree of drowsiness determined by the processing engine 112.

[0043] The image capturing device 130 may be configured to capture an image of one or more objects. The image may include a still picture, a motion picture, a video (offline or live streaming), a frame of a video (or referred to as a video frame), or a combination thereof. The one or more objects may be static or moving. The one or more objects may be an animal, a human being (a driver, an operator, a student, a worker) or a portion thereof (face), goods, or the like, or a combination thereof.

[0044] As shown in FIG. 1, the image capturing device 130 may include an automobile data recorder 130-1, a dome camera 130-2, a fixed camera 130-3, or the like, or a combination thereof. In some embodiments, the automobile data recorder 130-1 may be mounted in a vehicle and generally configured to

record a road condition around the vehicle when the driver is driving. The automobile data recorder 130-1 may include a camera orienting the face of a driver and capture an image of a face of a driver. In some embodiments, the imaging capturing device 130 may be combined with the terminal 120 (e.g., the mobile device 120-4). For example, a mobile device of a driver may include a camera that may capture images of the face of the driver.

[0045] In some embodiments, when a driver opens a navigation application (APP) in his or her mobile device for navigation during a journey, the camera of the mobile device may simultaneously capture images of the face of the driver. In some embodiments, the drowsiness detection method described in the present disclosure may be implemented by the mobile device or processors thereof to determine a degree of drowsiness of the driver based on the captured images. The mobile device may further include a component for generating a notification (e.g., a screen, a loudspeaker, a vibrating component) to generate a notification based on the determined degree of drowsiness of the driver. The notification may include a sound, a vibration, a light, or the like, or a combination thereof.

[0046] The network 140 may facilitate exchange of information and/or data. In some embodiments, one or more components of the drowsiness detection system 100 (e.g., the processing device 110, the terminal 120, the image capturing device 130 or the storage 150) may transmit information and/or data to another component (s) of the drowsiness detection system 100 via the network 140. For example, the processing device 110 (or the processing engine 112) may receive a plurality of video frames from the image capturing device 130 via the network 140. As another example, the processing device 110 (or the processing engine 112) may send a notification to the terminal 120 via the network 140. In some embodiments, the network 140 may be any type of wired or wireless network, or combination thereof. Merely by way of example, the network 140 may include a cable network, a wireline network, an optical fiber network, a telecommunications network, an intranet, an Internet,

a local area network (LAN), a wide area network (WAN), a wireless local area network (WLAN), a metropolitan area network (MAN), a public telephone switched network (PSTN), a Bluetooth network, a ZigBee network, a near field communication (NFC) network, or the like, or a combination thereof. In some embodiments, the network 140 may include one or more network access points. For example, the network 140 may include wired or wireless network access points such as base stations and/or internet exchange points 140-1, 140-2, through which one or more components of the drowsiness detection system 100 may be connected to the network 140 to exchange data and/or information.

[0047] The storage 150 may store data and/or instructions. In some embodiments, the storage 150 may store data obtained from the terminal 120 and/or the image capturing device 130. For example, the storage 150 may store a plurality of images of one or more objects captured from the image capturing device 130. In some embodiments, the storage 150 may store data and/or instructions that the processing device 110 may execute or use to perform exemplary methods described in the present disclosure. In some embodiments, storage 150 may include a mass storage, removable storage, a volatile read-and-write memory, a read-only memory (ROM), or the like, or a combination thereof. Exemplary mass storage may include a magnetic disk, an optical disk, solid-state drives, etc. Exemplary removable storage may include a flash drive, a floppy disk, an optical disk, a memory card, a zip disk, a magnetic tape, etc. Exemplary volatile read-and-write memory may include random-access memory (RAM). Exemplary RAM may include a dynamic RAM (DRAM), a double data rate synchronous dynamic RAM (DDR SDRAM), a static RAM (SRAM), a thyristor RAM (T-RAM), and a zero-capacitor RAM (Z-RAM), etc. Exemplary ROM may include a mask ROM (MROM), a programmable ROM (PROM), an erasable programmable ROM (EPROM), an electrically-erasable programmable ROM (EEPROM), a compact disk ROM (CD-ROM), and a digital versatile disk ROM, etc. In

some embodiments, the storage 150 may be implemented on a cloud platform. Merely by way of example, the cloud platform may include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, or the like, or a combination thereof.

[0048] In some embodiments, the storage 150 may be connected to the network 140 to communicate with one or more components of the drowsiness detection system 100 (e.g., the processing device 110, the terminal 120, or the image capturing device 130). One or more components of the drowsiness detection system 100 may access the data or instructions stored in the storage 150 via the network 140. In some embodiments, the storage 150 may be directly connected to or communicate with one or more components of the drowsiness detection system 100 (e.g., the processing device 110, the terminal 120, the image capturing device 130). In some embodiments, the storage 150 may be part of the processing device 110.

[0049] In some embodiments, one or more components of the drowsiness detection system 100 (e.g., the processing device 110, the terminal 120, the image capturing device 130) may have permissions to access the storage 150. In some embodiments, one or more components of the drowsiness detection system 100 may read and/or modify information when one or more conditions are met.

[0050] FIG. 2 is a schematic diagram illustrating exemplary hardware and/or software components of a computing device according to some embodiments of the present disclosure. In some embodiments, the processing device 110, the terminal 120, and/or the image capturing device 130 may be implemented on the computing device 200. For example, the processing engine 112 may be implemented on the computing device 200 and configured to perform functions of the processing engine 112 disclosed in this disclosure.

[0051] The computing device 200 may be used to implement any component of the drowsiness detection system 100 as described herein. For example, the processing engine 112 may be implemented on the computing device

200, via its hardware, software program, firmware, or a combination thereof. Although only one such computer is shown, for convenience, the computer functions relating to the drowsiness detection as described herein may be implemented in a distributed fashion on a number of similar platforms to distribute the processing load.

[0052] The computing device 200, for example, may include COM ports 250 connected to and from a network connected thereto to facilitate data communications. The computing device 200 may also include a processor 220, in the form of one or more processors (e.g., logic circuits), for executing program instructions. For example, the processor 220 may include interface circuits and processing circuits therein. The interface circuits may be configured to receive electronic signals from a bus 210, wherein the electronic signals encode structured data and/or instructions for the processing circuits to process. The processing circuits may conduct logic calculations, and then determine a conclusion, a result, and/or an instruction encoded as electronic signals. Then the interface circuits may send out the electronic signals from the processing circuits via the bus 210.

[0053] The computing device 200 may further include program storage and data storage of different forms including, for example, a disk 270, and a read-only memory (ROM) 230, or a random access memory (RAM) 240, for various data files to be processed and/or transmitted by the computing device. The exemplary computer platform may also include program instructions stored in the ROM 230, RAM 240, and/or another type of non-transitory storage medium to be executed by the processor 220. The methods and/or processes of the present disclosure may be implemented as the program instructions. The computing device 200 also includes an I/O component 260, supporting input/output between the computer and other components. The computing device 200 may also receive programming and data via network communications.

[0054] Merely for illustration, only one processor is described in FIG. 2.

Multiple processors are also contemplated. Thus operations and/or method steps performed by one processor as described in the present disclosure may also be jointly or separately performed by the multiple processors. For example, if in the present disclosure the processor of the computing device 200 executes both step A and step B, it should be understood that step A and step B may also be performed by two different CPUs and/or processors jointly or separately in the computing device 200 (e.g., the first processor executes step A and the second processor executes step B, or the first and second processors jointly execute steps A and B).

[0055] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary mobile device 300 on which the terminal 120 and/or the image capturing device 130 may be implemented according to some embodiments of the present disclosure. As illustrated in FIG. 3, the mobile device 300 may include a communication platform 310, a display 320, a graphics processing unit (GPU) 330, a central processing unit (CPU) 340, a I/O 350, a memory 360, a mobile operating system (OS) 370, and a storage 390. In some embodiments, any other suitable component, including but not limited to a system bus or a controller (not shown), may also be included in the mobile device 300.

[0056] In some embodiments, the mobile operating system 370 (e.g., iOS™, Android™, Windows Phone™, etc.) and one or more applications 380 may be loaded into the memory 360 from the storage 390 in order to be executed by the CPU 340. The applications 380 may include a browser or any other suitable mobile apps for receiving and rendering information relating to processing or other information from the drowsiness detection system 100. User interactions with the information stream may be achieved via the I/O 350 and provided to the processing engine 112 and/or other components of the drowsiness detection system 100 via the network 140.

[0057] In some embodiments, a camera (not shown in the figure) of the mobile device 300 may capture image(s) of a face of a driver. The CPU 340

of the mobile device 300 may determine a degree of drowsiness of the driver based on the captured video of the face of the driver. Then the mobile device 300 may generate a notification based on the degree of drowsiness of the driver. The notification may include a sound from a loudspeaker (not shown in the figure) of the mobile device 300, a vibration from a vibrator (not shown in the figure) of the mobile device 300, a light from the display 320 (or LEDs not shown in the figure), or the like, or a combination thereof.

[0058] It should be noted that one or more of the above functions of the mobile device 300 may be realized by a device outside the mobile device 300. For example, the mobile device 300 may work merely as a processing device and control another device to generate the notification. As another example, the mobile device 300 may receive an instruction of notification and work merely as a device for generating notifications.

[0059] To implement various modules, units, and their functionalities described in the present disclosure, computer hardware platforms may be used as the hardware platform(s) for one or more of the elements described herein. A computer with user interface elements may be used to implement a personal computer (PC) or any other type of work station or terminal device. A computer may also act as a system if appropriately programmed.

[0060] FIG. 4 is a block diagram illustrating an exemplary processing engine according to some embodiments of the present disclosure. The processing engine 112 may include an acquisition module 410, a detection module 420, an extraction module 430, a generation module 440 and a determination module 450.

[0061] The acquisition module 410 may be configured to receive information. For example, the acquisition module 410 may receive a plurality of images. The images may be video frames of a video captured by a camera via the network 140. For example, the processing engine 112 may receive a plurality of video frames from a camera. As used herein, the camera may be a network camera, a fixed camera, a dome camera, a covert camera, a Pan–

Tilt-Zoom (PTZ) camera, an infrared camera, a thermal camera, or the like, or a combination thereof. In some embodiments, the plurality of video frames may correspond to a driver, especially, a face of the driver. The acquisition module 410 may also receive a trained eye-blinking detection model and a trained yawning detection model from a storage device (e.g., the storage 150) via the network 140. In some embodiments, the trained eye-blinking detection model may be generated by training a preliminary eye-blinking detection model based on a plurality of training face samples. The trained yawning detection model may be generated by training a preliminary yawning detection model based on a plurality of training face samples. The training face samples used in training the preliminary eye-blinking detection model and the preliminary yawning detection model may be the same or different.

[0062] The detection module 420 may be configured to detect the face of a user in the plurality of images (e.g., video frames of a video). As used herein, the user may be anyone who needs to be monitored. For example, the user may be a driver, an operator, a worker, a student, an audience, or the like, or a combination thereof. In some embodiments, the face detection may be implemented based on a template matching, a skin color segmentation, a geometric rule confirmation, or the like, or a combination thereof. In some embodiments, the face detection may be implemented based on a model that employs a singular value algorithm, a binary wavelet transformation algorithm, an AdaBoost algorithm, etc.

[0063] The extraction module 430 may be configured to extract the detected faces from the plurality of video frames. In some embodiments, the detected faces may be in the form of image, or pixel value. The detected faces may include a plurality of face features including size, shape and/or a location of face outline, hair, lips, jaw, eyes, mouth, eyebrows, nose, or the like, or a combination thereof.

[0064] The generation module 440 may be configured to generate information associated with drowsiness detection. In some embodiments,

the generation module 440 may generate detection result associated with drowsiness detection. For example, the generation module 440 may generate an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detecting model. As another example, the generation module 440 may generate a yawning detection result by inputting the extracted faces into the trained yawning detecting model. In some embodiments, the generation module 440 may generate a notification (or an instruction for notification) based on the eye-blinking detection result and the yawning detection result.

[0065] The determination module 450 may be configured to determine a degree of drowsiness. In some embodiments, the degree of drowsiness of the user may be determined based on the frequency of eye-blinking and the frequency of yawning. To determine the frequency of eye-blinking, the determination module 450 may determine the number of blinks by the user detected in a plurality of video frames based on the eye-blinking detection result and a total time length of the plurality of video frames. To determine the frequency of yawning, the determination module 450 may determine the number of yawns by the user detected in a plurality of video frames and a total time length of the plurality of video frames.

[0066] In some embodiments, the processing engine 112 may further include a model training module (not shown in the figure). The model training module may be configured to train a preliminary eye-blinking detection model based on a plurality of training face samples to generate a trained eye-blinking detection model. The model training module may also train a preliminary yawning detection model based on a plurality of training face samples to generate a trained yawning detection model. The model training module may further train a preliminary angle determination model based on a plurality of training face samples to generate a trained angle determination model. The angle determination model may be used to determine whether an angle between the direction perpendicular to the face of a user and the

capturing direction of the camera is greater than a threshold (e.g., 30 degrees, 60 degrees, 100 degrees). The plurality of training face samples used in training the preliminary eye-blinking detection model, preliminary yawning detection model, and the preliminary angle determination model may be the same or different. Detailed descriptions of the exemplary process of training and using the model may be found elsewhere in this disclosure (e.g., FIG. 11 and the descriptions thereof).

[0067] The modules in the processing engine 112 may be connected to or communicate with each other via a wired connection or a wireless connection. The wired connection may include a metal cable, an optical cable, a hybrid cable, or the like, or a combination thereof. The wireless connection may include a Local Area Network (LAN), a Wide Area Network (WAN), a Bluetooth, a ZigBee, a Near Field Communication (NFC), or the like, or a combination thereof. Two or more of the modules may be combined into a single module, and any one of the modules may be divided into two or more units.

[0068] FIG. 5 is a flowchart illustrating an exemplary process for determining a degree of drowsiness of a driver according to some embodiments of the present disclosure. In some embodiments, the process 500 may be implemented as a set of instructions (e.g., an application) stored in the storage 390, ROM 230 or RAM 240. The CPU 340, processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the CPU 340, the processor 220 and/or the modules in FIG. 4 may be configured to perform the process 500. The operations of the illustrated process present below are intended to be illustrative. In some embodiments, the process 500 may be accomplished with one or more additional operations not described and/or without one or more of the operations herein discussed. Additionally, the order in which the operations of the process as illustrated in FIG. 5 and described below is not intended to be limiting.

[0069] In 510, the processing engine 112 (e.g., the acquisition module 410) may receive a plurality of images. The images may be video frames of a video captured by a camera. For example, the processing engine 112 may receive a plurality of video frames from a camera. As used herein, the camera may include but is not limited to a network camera, a fixed camera, a dome camera, a covert camera, a Pan–Tilt–Zoom (PTZ) camera, an infrared camera, a thermal camera, or the like, or a combination thereof. In some embodiments, the camera may be a standalone camera or integrated into a vehicle-mounted device (e.g., an automobile data recorder). In some embodiments, the camera may be mounted on or part of a mobile device of a user.

[0070] The camera may capture a video of an object. The video may include a still picture, a motion picture, an offline video, a live streaming video, or a combination thereof. The object may be a user that is needed to be monitored (e.g., a driver, an operator, a student, a worker). For example, in order to reduce the risk of drowsy driving, the user may be a driver who is driving. As another example, in order to reduce the risk of mechanical injury of a mechanical engineer or operator, the user may be the mechanical engineer or operator. As a further example, in order to raise the attention of a student (or an audience), the user may be a student (or an audience) that attends a class (or a lecture). The plurality of video frames may be extracted continuously or discontinuously from the video. For example, the plurality of video frames may be extracted at fixed time intervals (e.g., 5 minutes, 10 minutes, 30 minutes, 1 hour, 2 hours, 5 hours, etc.) or certain time points (e.g., 4 a.m., 9 p.m.). As another example, the plurality of extracted video frames may correspond to a portion of the video wherein the user inside the video is at a certain condition or doing a certain thing (e.g., driving, teaching, working).

[0071] In 520, the processing engine 112 (e.g., the detection module 420) may detect a face of a user in the plurality of images (e.g., video frames of a

video). In some embodiments, the face detection may be implemented in various ways including a template matching, a skin color segmentation, a geometric rule confirmation, or the like, or a combination thereof. In some embodiments, the face detection may be implemented based on a model that employs a singular value algorithm, a binary wavelet transformation algorithm, an AdaBoost algorithm or the like, or a combination thereof. In some embodiments, any video frame from the plurality of video frames that doesn't contain a face of the desired user may be discarded.

[0072] In some embodiments, after the face has been detected in the plurality of video frames, the processing engine 112 may determine an angle between the direction perpendicular to the face of the user and the capturing direction of the camera in each of the plurality of video frames. The capturing direction of the camera refers to the orientation of lens of the camera. The processing engine 112 may further determine the angle is greater than a threshold (e.g., 30 degrees, 60 degrees, 100 degrees). If the processing engine 112 may determine that the angle between the direction perpendicular to the face of the user and the capturing direction of the camera in a video frame is greater than the threshold, the processing engine 112 may discard the video frame from the plurality of video frames. The discarded video frames may not be processed in further operations in FIG. 5 (e.g., determination of an eye-blinking result and/or a yawning result). The determination of the angle between the direction perpendicular to the face of the user and the capturing direction of the camera may be implemented based on a model. Detailed descriptions of the exemplary process of training and using the model may be found elsewhere in this disclosure (e.g., FIG. 11 and the descriptions thereof).

[0073] In 530, the processing engine 112 (e.g., the extraction module 430) may extract the detected faces from the plurality of video frames. In some embodiments, the extracted faces may be in the form of images (pixel values) or part thereof. The extracted faces may correspond to a plurality of face

features including size, shape and/or a location of face outline, hair, lips, jaw, eyes, mouth, eyebrows, nose, or the like, or a combination thereof. In some embodiments, the extracted faces may be stored in a storage device (e.g., the storage 150).

[0074] In 540, the processing engine 112 (e.g., the acquisition module 410) may obtain a trained eye-blinking detection model and a trained yawning detection model. In some embodiments, the trained eye-blinking detection model and the trained yawning detection model may be generated by training a preliminary eye-blinking detection model and a preliminary yawning detection model, respectively. Merely by way of example, the preliminary eye-blinking detection model and/or the preliminary yawning detection model may be a convolutional neural network (CNN), a deep belief network (DBN), Stacked Auto-Encoders (SAE), a logistic regression (LR) model, a support vector machine (SVM), a decision tree model, a Naive Bayesian Model, a random forest model, a Restricted Boltzmann Machine (RBM), a Q-learning Model, or the like, or a combination thereof. In some embodiments, the CNN model may include at least one of a convolutional layer, a Rectified Linear Unit (ReLU) layer, a fully connected layer or a pooling layer.

[0075] In some embodiments, the trained eye-blinking detection model may be generated by training a preliminary eye-blinking detection model based on a plurality of training face samples. The trained yawning detection model may be generated by training a preliminary yawning detection model based on a plurality of training face samples. The training face samples used in training the preliminary eye-blinking detection model and the preliminary yawning detection model may be the same or different. Detailed descriptions of the exemplary process of training the preliminary eye-blinking detection model and the preliminary yawning detection model may be found elsewhere in disclosure (e.g., FIG. 11 and the descriptions thereof). In some embodiments, the plurality of training face samples may be acquired from a storage device (e.g., the storage 150), the terminal 120 via the network 140,

the image capturing device 130 via the network 140, or the like, or a combination thereof.

[0076] In some embodiments, the plurality of training face samples may include a set of positive eye-blink training samples and a set of negative eye-blink training samples based on whether the eye(s) of the user in the training face samples blink. For example, an eye in each of the positive eye-blink training samples blink, and an eye in each of the negative eye-blink training samples does not blink. A blink refers to a series of eye actions in which an action of closing eyes (also referred to as eye-close) is followed by an action of opening eyes (also referred to as eye-open). The eye-open and/or eye-close may each last for at least one video frame. Detailed descriptions of the exemplary process of the determination of an eye-blink may be found elsewhere in this disclosure (e.g., FIG. 8, and the descriptions thereof).

[0077] In some embodiments, the plurality of training face samples may include a set of positive yawn training samples and a set of negative yawn training samples based on whether the face of the user in the training face samples yawns. For example, the face in each of the positive yawn training samples yawn and the face in each of the negative yawn training samples does not yawn. A yawn refers to a series of face actions including the changes of face outline, mouth, eye, eyebrow, cheek, nose, jaw, hair, or the like, or a combination thereof. For example, when a person is yawning, his or her mouth may open, eyes may become small, eyebrow may rise, and jaw may tend downward.

[0078] The plurality of training face samples may include faces in the form of images (pixels). In some embodiments, the face may be a segmented face with different portions or facial organs (e.g., eyes, mouth, nose, jaw, hair) segmented out. In some embodiments, the face may be an entire face without segmentation. In a case that an entire face is used to train the preliminary yawning detection model, the yawn detection model may learn the change of greyscales of the entire face during a yawn without the need of

knowing the default shape and size of each portion of the face. In some embodiments, the location of each portion of the face may be marked (e.g., eyes, mouth) to speed up the training of the yawn detection model. Alternatively, no portion of the face is marked, and the yawning detection model may learn the location and shape of the portions of the face during the training process. Detailed descriptions of the exemplary process of the determination of yawning and obtaining of the trained yawning detection model may be found elsewhere in this disclosure (e.g., FIGs. 9-11 and the descriptions thereof).

[0079] In 550, the processing engine 112 (e.g., the generation module 440) may generate an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detecting model. In some embodiments, the eye-blinking detection result may include a result that the eyes in the extracted faces blink or a result that the eyes in the extracted faces don't blink. In some embodiments, the result that the eyes blink may be determined according to a series of eye actions including an action of closing eyes (also referred to as eye-close) followed by an action of opening eyes (also referred to as eye-open). The eye-open and/or eye-close may each last for at least one video frame. Detailed descriptions of the exemplary process of the generation of the eye-blinking result may be found elsewhere in this disclosure (e.g., FIG. 8 and the descriptions thereof).

[0080] In some embodiments, the processing engine 112 (e.g., the generation module 440) may further determine the frequency of eye-blinking by the user based on the eye-blinking result. The frequency of eye-blinking may be in a range of 0 to 100 times per minute (e.g., zero, 5 times per minute, 10 times per minute, 20 times per minute, 50 times per minute, 100 times per minute). Detailed descriptions of the exemplary process of the determination of the frequency of eye-blinking may be found elsewhere in this disclosure (e.g., FIG. 6 and the relevant descriptions thereof).

[0081] In 560, the processing engine 112 (e.g., the generation module 440)

may generate a yawning detection result by inputting the extracted faces into the trained yawning detecting model. In some embodiments, the yawning detection result may include a result of yawning and/or a result of not being yawning. Detailed descriptions of the exemplary process of the determination of yawning result may be found elsewhere in this disclosure (e.g., FIG. 9, 10 and the descriptions thereof).

[0082] In some embodiments, the processing engine 112 (e.g., the generation module 440) may further determine the frequency of yawning by the user based on the yawning detection result. The frequency of yawning may be in a range of 0 to 30 times per minutes (e.g. zero, 1 time per minute, 2 times per minute, 5 times per minute, 10 times per minute, 30 times per minute). Detailed descriptions of the exemplary process of the determination of the frequency of yawning may be found elsewhere in this disclosure (e.g., FIG. 7 and the relevant descriptions thereof).

[0083] In 570, the processing engine 112 (e.g., the determination module 450) may determine a degree of drowsiness of the user based on the eye-blinking detection result and the yawning detection result. More particularly, the degree of drowsiness of the user may be determined based on the frequency of eye-blinking and the frequency of yawning. For example, a drowsiness score may be determined according to Equation (1) as below:

$$d = af_1 + bf_2, \quad (1)$$

where d represents the drowsiness score, a represents a weight coefficient of the frequency of eye-blinking, f_1 represents the frequency of eye-blinking, b represents a weight coefficient of the frequency of yawning, and f_2 represents the frequency of yawning. In some embodiments, the weight coefficient of the frequency of eye-blinking a and the weight coefficient of the frequency of yawning b be the same or different. The weight coefficients may be default parameters stored in a storage device (e.g., the storage device 150), or set or adjusted by an operator of the drowsiness detection system 100.

[0084] In some embodiments, the drowsiness score may be represented by a numeric value (e.g., 0, 1, 2, 5, 10, 20, 50, 100). The degree of drowsiness may be determined based on the drowsiness score. The degree of drowsiness may include a plurality of levels (e.g., low, medium, high). For example, when the drowsiness score is less than a first threshold (e.g., 10), the degree of drowsiness may be a low level. When the drowsiness score is less than a second threshold (e.g., 30) but greater than or equal to the first threshold, the degree of drowsiness may be a medium level. When the drowsiness score is greater than or equal to the second threshold, the degree of drowsiness may be determined at a high level.

[0085] In 580, the processing engine 112 (e.g., the generation module 440) may generate a notification based on the degree of drowsiness. In some embodiments, the processing engine 112 may generate an instruction for generating a notification based on the degree of drowsiness, which may be transmitted to a terminal of the user via a network (e.g., the network 140). The terminal may, after receiving the instruction, generate a notification according to the received instruction. The notification may include a sound, a vibration, a light, or the like, or a combination thereof.

[0086] In some embodiments, the processing engine 112 (e.g., the generation module 440) may determine the loudness or frequency of the sound, strength or frequency of the vibration, an intensity or twinkling frequency of the light based on the degree of drowsiness. For example, when the degree of drowsiness of the user is in the low level, the processing engine 112 or the terminal 120 may generate the notification that merely includes twinkling light. When the degree of drowsiness of the user is in the medium level, the processing engine 112 or the terminal 120 may generate the notification that includes twinkling light and a soft sound. When the degree of drowsiness of the user is in the high level, the processing engine 112 or the terminal 120 may generate the notification that includes a loud sound (e.g., with high decibels), bright light with a high-frequency twinkling,

and/or a strong vibration. It should be noted that the notification described above is provided merely as examples and shall not limit the scope of the present disclosure. People having ordinary skills in the art may have other types of notifications at the same or different degree of drowsiness.

[0087] In some embodiments, when a driver opens a navigation application (APP) in his or her mobile device for navigation during a journey, the camera of the mobile device may simultaneously capture images of the face of the driver. In some embodiments, the drowsiness detection method described in 500 may be implemented by the mobile device or processors thereof to determine a degree of drowsiness of the driver based on the captured images. The mobile device may further include a notification generating component (e.g., a screen, a loudspeaker, a vibrating component) to generate a notification based on the determined degree of drowsiness of the driver. The notification may include a sound, a vibration, a light, or the like, or a combination thereof.

[0088] It should be noted that the above description is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. In some embodiments, one or more other optional operations may be omitted in the exemplary process 500. For example, 570 may be omitted. The generation module 440 may generate a notification based on the eye-blinking result in 550 and the yawning detection result in 560. As another example, 550 may be omitted. The determination module 450 may determine the degree of drowsiness of the user based on the yawning detection result.

[0089] FIG. 6 is a flowchart illustrating an exemplary process for determining the frequency of eye-blinking according to some embodiments of the present disclosure. In some embodiments, the process 600 may be implemented as

a set of instructions (e.g., an application) stored in the storage 390, ROM 230 or RAM 240. The CPU 340, processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the CPU 340, the processor 220 and/or the modules in FIG. 4 may be configured to perform the process 600. The operations of the illustrated process present below are intended to be illustrative. In some embodiments, the process 600 may be accomplished with one or more additional operations not described and/or without one or more of the operations herein discussed. Additionally, the order in which the operations of the process as illustrated in FIG. 6 and described below is not intended to be limiting. In some embodiments, the determination of the frequency blinking by the user based on the eye-blinking result described above in connection with operation 550 of the process 500 may be performed according to the process 600.

[0090] In 610, the processing engine 112 (e.g., the determination module 450) may determine the number (or a number count) of blinks (also referred to as eye-blinks or eye-blinking) in the plurality of video frames based on an eye-blinking detection result. For example, the number of blinks may be zero times, 1 time, 2 times, 5 times, 10 times, 20 times. A blink refers to a series of eye actions in which an action of closing eyes (also referred to as eye-close) is followed by an action of opening eyes (also referred to as eye-open). The eye-open and/or eye-close may each last for at least one video frame. Detailed descriptions of the exemplary process of the determination of the number of blinks may be found elsewhere in this disclosure (e.g., FIG. 8 and the descriptions thereof).

[0091] In 620, the processing engine 112 (e.g., the determination module 450) may determine the total time length of the plurality of video frames. The total time length of the plurality of video frames may be determined by summing up the length of each of the plurality of video frames. In some embodiments, a video frame may correspond to a time point and a length of a video frame may be defined as a time length between the time point

corresponding to the video frame and the time point corresponding to the succeeding video frame. For example, assume the number of the plurality of video frames is 60, and interval time length between any two successive video frames is 0.8 seconds, and then the total time length of the plurality of video frames may be $60 \times 0.8 \text{ s} = 48 \text{ s}$. The total time length may be 5 seconds, 10 seconds, 20 seconds, 1 minute, 5 minutes. In some embodiments, at least one of the plurality of video frames may be discarded (or skipped). In order to determine a correct total time length of the plurality of video frames, the time length of the discarded or skipped video frame may be subtracted from the total time length of the plurality of video frames. For example, if two video frames of the plurality of video frames are discarded or skipped, the total time length of the plurality of video frames (e.g., remaining video frames) may be $48 \text{ s} - 2 \times 0.8 = 46.4 \text{ s}$. In some embodiments, the time lengths of the plurality of video frames may be same or different. The time length of the each of the plurality of video frames may be a default parameter stored in a storage device (e.g., the storage 150), or set or adjusted by an operator of the drowsiness detection system 100.

[0092] In 630, the processing engine 112 (e.g., the identification unit 630) may determine the frequency of eye-blinking based on the number of blinks and the total time length of the plurality of video frames. The frequency of eye-blinking may be 0, once per second, twice per second, once per minute, twice per minute, 10 times per minute, 20 times per minute, etc. For example, the number of blinks is 6 times, and the total time length of the plurality of video frames is 48 s, the frequency of eye-blinking may be $48 \text{ s} \div 6 \text{ times} = 8 \text{ s/times}$ or $\frac{1}{8} \text{ times/s}$.

[0093] It should be noted that the above description is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present

disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. For example, operation 610 and operation 620 may be combined as a single operation in which the processing engine 112 may determine both the number of blinks and the total time length of the plurality of video frames.

[0094] FIG. 7 is a flowchart illustrating an exemplary process for determining the frequency of yawning according to some embodiments of the present disclosure. In some embodiments, the process 700 may be implemented as a set of instructions (e.g., an application) stored in the storage 390, ROM 230 or RAM 240. The CPU 340, processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the CPU 340, the processor 220 and/or the modules in FIG. 4 may be configured to perform the process 700. The operations of the illustrated process present below are intended to be illustrative. In some embodiments, the process 700 may be accomplished with one or more additional operations not described and/or without one or more of the operations herein discussed. Additionally, the order in which the operations of the process as illustrated in FIG. 5 and described below is not intended to be limiting.

[0095] In 710, the processing engine 112 (e.g., the determination module 450) may determine the number (or a number count) of yawns in the plurality of video frames based on a yawning detection result. The number of yawns may be 0, 1 time, 5 times, 10 times, 20 times, etc. A yawn refers to a series of face actions including the changes of face outline, mouth, eyebrow, nose, hair, or the like, or a combination thereof. Detailed descriptions of the exemplary process of the determination of the number of yawns may be found elsewhere in this disclosure (e.g., FIGs. 9 and 10 and the descriptions thereof).

[0096] In 720, the processing engine 112 (e.g., the determination module 450) may determine the total time length of the plurality of video frames. The total time length of the plurality of video frames may be determined by

summing up the length of each of the plurality of video frames. The total time length may be 5 seconds, 10 seconds, 20 seconds, 1 minute, 5 minutes.

[0097] In 730, the processing engine 112 may determine a frequency of yawning based on the number of yawns and the total time length of the plurality of video frames. The frequency of yawning may be 0, once per minute, twice per minute, 3 times per minute, 5 times per minute, etc. For example, the number of yawns is 2 times, and the total time length of the plurality of video frames is 48 seconds, the frequency of eye-blinking may be $2 \text{ times} / 48 \text{ s} * 60 = 2.5 \text{ times} / \text{min}$.

[0098] It should be noted that the above description is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. For example, operation 710 and operation 720 may be combined as a single operation in which the processing engine 112 may determine both the number of yawns and the total time length of the plurality of video frames.

[0099] FIG. 8 is a schematic diagram illustrating an exemplary blink according to some embodiments of the present disclosure. In some embodiments, a blink may be determined according to a series of eye actions detected, which include an action of closing eyes (also referred to as eye-close) followed by an action of opening eyes (also referred to as eye-open). The eye-open and/or eye-close may each last for at least one video frame. In some embodiments, in order to detect a blink, the length of each video frame may be less than 0.2 s, which is a typical duration of a blink.

[0100] As illustrated in FIG. 8, the trained eye-blinking detection model may generate a result of “open” or “close” of eyes in response to each inputted video frame. The result of “open” or “close” of eyes may be arranged in a time sequence of the video frames. “Open” 810 may represent a video

frame with opened eyes, and “close” 820 may represent a video frame with closed eyes. “Open” 810 and “close” 820 in the continuous video frames may represent an action of closing eyes (eye-close). “close” 830 may represent a video frame with closed eyes, and “open” 840 may represent a video frame with opened eyes. “Close” 830 and “open” 840 in the continuous video frames may represent an action of opening eyes. The successive action of closing eyes and opening eyes may constitute a blink. In some embodiments, the trained eye-blinking detection model may directly generate a result that the eyes of the inputted face blink or not.

[0101] FIG. 9 is a schematic diagram illustrating an exemplary yawn according to some embodiment of the present disclosure. In some embodiments, yawning may be confirmed only when a positive (or affirmative) result of yawning is generated by the yawning detection model in a number of continuous video frames (or lasts longer than a time threshold). As illustrated in FIG. 9, “yes” frames 910, 920, 930, 940 may represent frames in which the face is determined to yawn by the yawning detection model (or referred to as frames with positive yawning result), while the face in “no” frames 950 and 960 are determined not to yawn by the yawning detection model (or referred to as frames with negative yawning result). If the requirement on the number of continuous video frames is three, then the successive video frames 910-930 may be confirmed to be a yawn. However, as the frames 940-960 only include one “yes” video frame, the frames 940-960 is not confirmed as a yawn. The positive yawning result (e.g., “yes”) in frame 940 may be caused due to other activities by the driver that includes a change of the entire face (e.g., talking, laughing).

[0102] FIG. 10 is a schematic diagram illustrating an exemplary process for detecting a yawning according to some embodiments of the present disclosure. In some embodiments, a yawn may cause the changes of shapes, sizes, and/or locations of any portion of the face (e.g., eyes, mouth, ears, lips, nose, eyebrow, hair). For example, when a person is yawning, his

or her mouth may open, his or her cheek may be stressed, his or her head may lean backward, his or her jaw may go down and/or his or her eyes may become small.

[0103] As shown in FIG. 10, video frames 1010-1080 may correspond to a change of the face during a yawn. In video frame 1010, the man's mouth is closed, and other portions of his face are in a usual condition. The yawning detection model may generate a result that the man in 1010 is not yawning. As illustrated in 1020 and/or 1030, the man's mouth begins to open or has opened slightly, and his head leans backward slightly. The yawning detection model may generate a result that the man has a low chance of yawning (but a high chance of, for example, deep breathing or talking). In video frames 1040 and 1050, the mouth of the man is widely opened, the cheek is stressed, the eyes are closed slightly, and the head leans backward. The yawning detection model may generate a result that the man has a high chance of yawning. In some embodiments, the processing engine 112 may use the method described in, e.g., FIG. 9 to confirm whether the man is yawning. In video frames 1050-1080, the face of the man is returning to its usual condition.

[0104] It should be noted the above description is provided merely as an example to ease the understanding of how the yawning detection model works, and the descriptions of the change of portions of the face are merely illustrative. The yawning detection model may not identify the portions of the faces to generate the yawning detection result but may generate the yawning detection result based on the greyscale change of pixels in the face. For example, the yawning detection model may not know that a bottom portion of the face of the man includes a mouth and the change of shape or size of the mouth. However, the yawning detection model may recognize that the average pixel value in a bottom section of the face increases (becomes darker when opening mouth) and determine a result that the man is yawning based on it.

[0105] FIG. 11 is a schematic diagram illustrating an exemplary model according to some embodiments of the present disclosure. The model 1110 may be obtained by operation 540 in process 500. Merely by way of example, the model 1110 may be convolutional neural network (CNN), deep belief network (DBN), Stacked Auto-Encoders (SAE), logistic regression (LR) model, support vector machine (SVM), decision tree model, Naive Bayesian Model, random forest model, or Restricted Boltzmann Machine (RBM), Q-learning Model, or the like, or a combination thereof.

[0106] In some embodiments, the model 1110 may be used to generate an eye-blinking result, a yawning result, an angle determination result, or the like, or a combination thereof. The angle determination result may include whether an angle between the direction perpendicular to the face of a user and the capturing direction of the camera is greater than a threshold (e.g., 30 degrees, 60 degrees, 100 degrees).

[0107] In some embodiments, the model 1110 may be trained in a plurality of iterations. During each of the plurality of iterations, a plurality of training face samples 1120 may be input into the model 1110 to generate a preliminary result 1130. In some embodiments, the plurality of training face samples 1120 may be labeled as positive and/or negative samples. For example, the plurality of training samples may be labelled as yawn or not yawn, eye-blinking or not eye-blinking, appropriate angle (an angle between the direction perpendicular to face of a user and the capturing direction of the camera less than or equal to the threshold) or inappropriate angle (an angle between the direction perpendicular to face of a user and the capturing direction of the camera greater than the threshold). The one or more internal parameters of the model 1110 may include a weight factor, a bias term, etc. In some embodiments, a loss function may be obtained based on the label of the training face sample and the preliminary result 1130. For example, if the preliminary result 1130 and the label of the training face sample is the same, the loss function may be small. If the preliminary result 1130 and the label of

the training face sample is different, the loss function may be large. The loss function may be used to update the one or more parameters of the model 1110.

[0108] In some embodiments, the iterations may terminate when a preset condition is satisfied. The preset condition may include that a loss function reaches a minimum value (convergence) in the training process of the model. In some embodiments, the preset condition may include the number of iterations (for example, two hundred times) performed reaches a threshold number. If the number of iterations performed reaches the threshold number (e.g., the preset condition is satisfied), the iterations may terminate. The threshold number may be set by an operator or according to default settings of drowsiness detection system 100, or a combination thereof.

[0109] In some embodiments, different training face samples and labels may be used to generate different models. For example, the plurality of training face samples 1120 may be classified into a set of positive eye-blink training samples and a set of negative eye-blink training samples based on whether the eye(s) of the user in the training face samples blink. The set of positive eye-blink training samples and the set of negative eye-blink training samples may be used to train an eye-blinking detection model.

[0110] As another example, the plurality of training face samples 1120 may be classified into a set of positive yawn training samples and a set of negative yawn training samples based on whether the face of the user in the training face samples 1120 yawns. The set of positive yawn training samples and the set of negative yawn training samples may be used to train a yawning detection model.

[0111] As yet a further example, the plurality of training face samples 1120 may be classified into a set of positive angle training samples and a set of negative angle training samples based on whether the angle between the direction perpendicular to face of a user and the capturing direction of the camera is less than or equal to the threshold. More particularly, the angle

between the direction perpendicular to face of a user and the capturing direction of the camera may be less than or equal to the threshold in the set of positive angle training samples while the angle between the direction perpendicular to face of a user and the capturing direction of the camera may be greater than the threshold in the set of negative angle training samples. The set of positive angle training samples and the set of negative angle training samples may be used to train an angle determination model.

[0112] When the training iterations terminate, a trained model may be generated. Detected faces 1140 may be inputted to the trained model, and the trained model may generate detecting results 1150 in response to the detected faces 1140. For example, a trained eye-blinking detection model may output a detection result 1150 as to whether one or more eyes in the detected faces 1140 blink. A trained yawning detection model may output a detection result 1150 as to whether the detected faces 1140 yawn. A trained angle determination model may generate a result as to whether an angle between the direction perpendicular to detected faces 1140 and the capturing direction of the camera is less than or equal to a threshold (e.g., 30 degrees, 60 degrees). In some embodiments, the model 1110 (and/or its one or more internal parameters) may be further updated based on the detecting results 1150.

[0113] FIG. 12 is a schematic diagram illustrating an exemplary automobile data recorder according to some embodiment of the present disclosure. As illustrated in FIG. 12, a camera in the automobile data recorder 1210 may be orientated toward a driver to acquire a plurality of video frames of the face of the driver. The automobile data recorder 1210 may be connected to the vehicle via an electric wire 1220. In some embodiments, an onboard HCI system, a mobile device of the driver, or the automobile data recorder 1210 itself may process the acquired plurality of video frames to determine a degree of drowsiness of the driver. A notification may be generated by the onboard HCI system, the mobile device of the driver or the automobile data

recorder 1210 based on the degree of drowsiness of the driver to warn the driver of drowsy driving and remind the driver to take a break.

[0114] Having thus described the basic concepts, it may be rather apparent to those skilled in the art after reading this detailed disclosure that the foregoing detailed disclosure is intended to be presented by way of example only and is not limiting. Various alterations, improvements, and modifications may occur and are intended to those skilled in the art, though not expressly stated herein. These alterations, improvements, and modifications are intended to be suggested by this disclosure and are within the spirit and scope of the exemplary embodiments of this disclosure.

[0115] Moreover, certain terminology has been used to describe embodiments of the present disclosure. For example, the terms “one embodiment,” “an embodiment,” and/or “some embodiments” mean that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present disclosure. Therefore, it is emphasized and should be appreciated that two or more references to “an embodiment,” “one embodiment,” or “an alternative embodiment” in various portions of this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures or characteristics may be combined as suitable in one or more embodiments of the present disclosure.

[0116] Further, it will be appreciated by one skilled in the art, aspects of the present disclosure may be illustrated and described herein in any of a number of patentable classes or context including any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof. Accordingly, aspects of the present disclosure may be implemented entirely hardware, entirely software (including firmware, resident software, micro-code, etc.) or combining software and hardware implementation that may all generally be referred to herein as a “block,” “module,” “engine,” “unit,” “component,” or “system.” Furthermore, aspects

of the present disclosure may take the form of a computer program product embodied in one or more computer-readable media having computer readable program code embodied thereon.

[0117] A computer readable signal medium may include a propagated data signal with computer readable program code embodied therein, for example, in baseband or as part of a carrier wave. Such a propagated signal may take any of a variety of forms, including electromagnetic, optical, or the like, or any suitable combination thereof. A computer readable signal medium may be any computer readable medium that is not a computer readable storage medium and that may communicate, propagate, or transport a program for use by or in connection with an instruction execution system, apparatus, or device. Program code embodied on a computer readable signal medium may be transmitted using any appropriate medium, including wireless, wireline, optical fiber cable, RF, or the like, or any suitable combination of the foregoing.

[0118] Computer program code for carrying out operations for aspects of the present disclosure may be written in a combination of one or more programming languages, including an object-oriented programming language such as Java, Scala, Smalltalk, Eiffel, JADE, Emerald, C++, C#, VB. NET, Python or the like, conventional procedural programming languages, such as the "C" programming language, Visual Basic, Fortran 1703, Perl, COBOL 1702, PHP, ABAP, dynamic programming languages such as Python, Ruby, and Groovy, or other programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer or in a cloud computing environment or offered as a service such as a

software as a service (SaaS).

[0119] Furthermore, the recited order of processing elements or sequences, or the use of numbers, letters, or other designations, therefore, is not intended to limit the claimed processes and methods to any order except as may be specified in the claims. Although the above disclosure discusses through various examples what is currently considered to be a variety of useful embodiments of the disclosure, it is to be understood that such detail is solely for that purpose and that the appended claims are not limited to the disclosed embodiments, but, on the contrary, are intended to cover modifications and equivalent arrangements that are within the spirit and scope of the disclosed embodiments. For example, although the implementation of various components described above may be embodied in a hardware device, it may also be implemented as a software-only solution—e.g., an installation on an existing server or mobile device.

[0120] Similarly, it should be appreciated that in the foregoing description of embodiments of the present disclosure, various features are sometimes grouped together in a single embodiment, figure, or description thereof for the purpose of streamlining the disclosure aiding in the understanding of one or more of the various embodiments. This method of disclosure, however, is not to be interpreted as reflecting an intention that the claimed subject matter requires more features than are expressly recited in each claim. Rather, claimed subject matter may lie in less than all features of a single foregoing disclosed embodiment.

WE CLAIM:

1. A system, comprising:

at least one computer-readable storage device including a set of instructions for determining a degree of drowsiness of a driver; and

at least one processor in communication with the at least one computer-readable storage device, wherein when executing the set of instructions, the at least one processor is configured to cause the system to:

receive, from a camera, a plurality of video frames;

detect a face of a driver in the plurality of video frames;

extract the detected faces in the plurality of video frames;

obtain a trained eye-blinking detection model and a trained yawning detection model;

generate an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model;

generate a yawning detection result by inputting the extracted faces into the trained yawning detection model;

determine a degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result; and

generate a notification based on the degree of drowsiness.

2. The system of claim 1, wherein the trained eye-blinking detection model is generated by a process for training an eye-blinking detection model, the process comprising:

obtaining a preliminary eye-blinking detection model;

obtaining a plurality of training face samples;

classifying the plurality of training face samples into a set of positive eye-blink training samples and a set of negative eye-blink training samples, wherein eyes in each of the positive eye-blink training samples blink and eyes in each of the negative eye-blink training samples do not blink; and

training the preliminary eye-blinking detection model based on the set of

positive eye-blink training samples and the set of negative eye-blink training samples to generate the trained eye-blinking detection model.

3. The system of claim 1, wherein the trained yawning detection model is generated by a process for training a yawning detection model, the process comprising:

- obtaining a preliminary yawning detection model;

- obtaining a plurality of training face samples;

- classifying the plurality of training face samples into a set of positive yawn training samples and a set of negative yawn training samples, wherein a person of the face in each of the positive yawn training samples yawns and a person of the face in each of the negative yawn training samples does not yawn; and

- training the preliminary yawning detection model based on the set of positive yawn training samples and the set of negative yawn training samples to generate the trained yawning detection model.

4. The system of any one of claims 1-3, wherein the at least one processor is further configured to cause the system to:

- for at least one video frame of the plurality of video frames:

- determine whether an angle between a direction perpendicular to a face of the driver in the at least one video frame of the plurality of video frames and a capturing direction of the camera is greater than a threshold;
- and

- in response to the determination that the angle between the direction perpendicular to the face of the driver and the capturing direction of the camera is greater than the threshold, discard the at least one video frame of the plurality of video frames from the plurality of video frames to be inputted into the trained eye-blinking detection model or the trained yawning detection model.

5. The system of claim 4, wherein to determine the angle between the direction perpendicular to the face of the driver and the capturing direction of the camera, the at least one processor is configured to cause the system to input the extracted face in each of the plurality of video frames into a trained angle determination model to generate a result that whether the angle between the direction perpendicular to the face of the driver and the capturing direction of the camera is greater than the threshold.

6. The system of claim 5, wherein the trained angle determination model is generated by a process for training an angle determination model, the process comprising:

- obtaining a preliminary angle determination model;

- obtaining a plurality of training face samples;

- classifying the plurality of training face samples into a set of positive angle training samples and a set of negative angle training samples, wherein

- the angle between the direction perpendicular to a face of a driver and the capturing direction of a camera in each positive angle training sample is greater than 60, and

- the angle between the direction perpendicular to a face of the driver and the capturing direction of the camera in each negative angle training sample is less than or equal to 60 degrees; and

- training the preliminary angle determination model based on the set of positive training samples and the set of negative training samples to generate the trained angle determination model.

7. The system of any one of claims 1-3, wherein to determine the degree of drowsiness based on the eye-blinking detection result and the yawning detection result, the at least one processor is further configured to cause the system to:

determine a frequency of eye-blinking by the driver based on the eye-blinking detection result;

determine a frequency of yawning by the driver based on the yawning detection result; and

determine the degree of drowsiness based on the frequency of eye-blinking and the frequency of yawning.

8. The system of claim 7, wherein to determine the frequency of eye-blinking of the driver based on the eye-blinking detection result, the at least one processor is further configured to cause the system to:

determine, based on the eye-blinking detection result, a number count of one or more blinks by the driver detected in the plurality of video frames;

determine a total time length of the plurality of video frames; and

determine the frequency of eye-blinking based on a number count of the one or more blinks and the total time length of the plurality of video frames.

9. The system of claim 7, wherein to determine the frequency of yawning of the driver based on the plurality of yawning detection results, the at least one processor is further configured to cause the system to:

determine, based on the plurality of yawning detection results, the number of one or more yawns by the driver detected in the plurality of successive frames;

determine a total time length of the plurality of video frames; and

determine the frequency of yawning based on a number count of the one or more yawns of the driver and the total time length of the plurality of video frames.

10. The system of any one of claims 1-3, wherein the notification includes at least one of a sound, a vibration, or a light.

11. The system of claim 10, wherein a loudness or frequency of the sound, a strength of the vibration, or an intensity or frequency of the light depends on the degree of drowsiness.

12. A method for determining a degree of drowsiness of a driver, comprising:

receiving, from a camera, a plurality of video frames;

detecting a face of a driver in the plurality of video frames;

extracting the detected faces in the plurality of video frames;

obtaining a trained eye-blinking detection model and a trained yawning detection model;

generating an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model;

generating a yawning detection result by inputting the extracted faces into the trained yawning detection model;

determining a degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result; and

generating a notification based on the degree of drowsiness.

13. The method of claim 12, wherein the trained eye-blinking detection model is generated by a process for training an eye-blinking detection model, the process comprising:

obtaining a preliminary eye-blinking detection model;

obtaining a plurality of training face samples;

classifying the plurality of training face samples into a set of positive eye-blink training samples and a set of negative eye-blink training samples, wherein eyes in each of the positive eye-blink training samples blink and eyes in each of the negative eye-blink training samples do not blink; and

training the preliminary eye-blinking detection model based on the set of positive eye-blink training samples and the set of negative eye-blink training samples to generate the trained eye-blinking detection model.

14. The method of claim 12, wherein the trained yawning detection model is generated by a process for training a yawning detection model, the process comprising:

obtaining a preliminary yawning detection model;

obtaining a plurality of training face samples;

classifying the plurality of training face samples into a set of positive yawn training samples and a set of negative yawn training samples, wherein a person of the face in each of the positive yawn training samples yawns and a person of the face in each of the negative yawn training samples does not yawn; and

training the preliminary yawning detection model based on the set of positive yawn training samples and the set of negative yawn training samples to generate the trained yawning detection model.

15. The method of any one of claims 12-14, wherein the method further comprises:

for at least one video frame of the plurality of video frames:

determining whether an angle between a direction perpendicular to a face of the driver in the at least one video frame of the plurality of video frames and a capturing direction of the camera is greater than a threshold; and

in response to the determination that the angle between the direction perpendicular to the face of the driver and the capturing direction of the camera is greater than the threshold, discarding the at least one video frame of the plurality of video frames from the plurality of video frames to be inputted into the trained eye-blinking detection model or the trained yawning detection model.

16. The method of claim 15, wherein the determining the angle between the

direction perpendicular to the face of the driver and the capturing direction of the camera includes inputting the extracted face in each of the plurality of video frames into a trained angle determination model to generate a result that whether the angle between the direction perpendicular to the face of the driver and the capturing direction of the camera is greater than the threshold.

17. The method of claim 16, wherein the trained angle determination model is generated by a process for training an angle determination model, the process comprising:

- obtaining a preliminary angle determination model;

- obtaining a plurality of training face samples;

- classifying the plurality of training face samples into a set of positive angle training samples and a set of negative angle training samples, wherein

 - the angle between the direction perpendicular to a face of a driver and the capturing direction of a camera in each positive angle training sample is greater than 60, and

 - the angle between the direction perpendicular to a face of the driver and the capturing direction of the camera in each negative angle training sample is less than or equal to 60 degrees; and

- training the preliminary angle determination model based on the set of positive training samples and the set of negative training samples to generate the trained angle determination model.

18. The method of any one of claims 12-14, wherein the determining the degree of drowsiness based on the eye-blinking detection result and the yawning detection result includes:

- determine a frequency of eye-blinking by the driver based on the eye-blinking detection result;

- determine a frequency of yawning by the driver based on the yawning detection result; and

determine the degree of drowsiness based on the frequency of eye-blinking and the frequency of yawning.

19. The method of claim 18, wherein the determining the frequency of eye-blinking of the driver based on the eye-blinking detection result includes:

determine, based on the eye-blinking detection result, a number count of one or more blinks by the driver detected in the plurality of video frames;

determine a total time length of the plurality of video frames; and

determine the frequency of eye-blinking based on a number count of the one or more blinks and the total time length of the plurality of video frames.

20. The method of claim 18, wherein the determining the frequency of yawning of the driver based on the plurality of yawning detection results includes:

determine, based on the plurality of yawning detection results, the number of one or more yawns by the driver detected in the plurality of successive frames;

determine a total time length of the plurality of video frames; and

determine the frequency of yawning based on a number count of the one or more yawns of the driver and the total time length of the plurality of video frames.

21 The method of any one of claims 12-14, wherein the notification includes at least one of a sound, a vibration, or a light.

22. The method of claim 21, wherein a loudness or frequency of the sound, a strength of the vibration, or an intensity or frequency of the light depends on the degree of drowsiness.

23. A system for determining a degree of drowsiness of a driver, comprising:

an acquisition module, configured to receive a plurality of video frames from a camera, and obtain a trained eye-blinking detection model and a trained yawning detection model;

a detection module, configured to detect a face of a driver in the plurality of video frames;

an extraction module, configured to extract the detected faces in the plurality of video frames;

a generation module, configured to generate an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model, generate a yawning detection result by inputting the extracted faces into the trained yawning detection model, and generate a notification based on a degree of drowsiness; and

a determination module, configured to determine the degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result.

24. A non-transitory computer-readable storage medium embodying a computer program product, the computer program product comprising instructions configured to cause a computing device to:

receive, from a camera, a plurality of video frames;

detect a face of a driver in the plurality of video frames;

extract the detected faces in the plurality of video frames;

obtain a trained eye-blinking detection model and a trained yawning detection model;

generate an eye-blinking detection result by inputting the extracted faces into the trained eye-blinking detection model;

generate a yawning detection result by inputting the extracted faces into the trained yawning detection model;

determine a degree of drowsiness of the driver based on the eye-blinking detection result and the yawning detection result; and

generate a notification based on the degree of drowsiness.

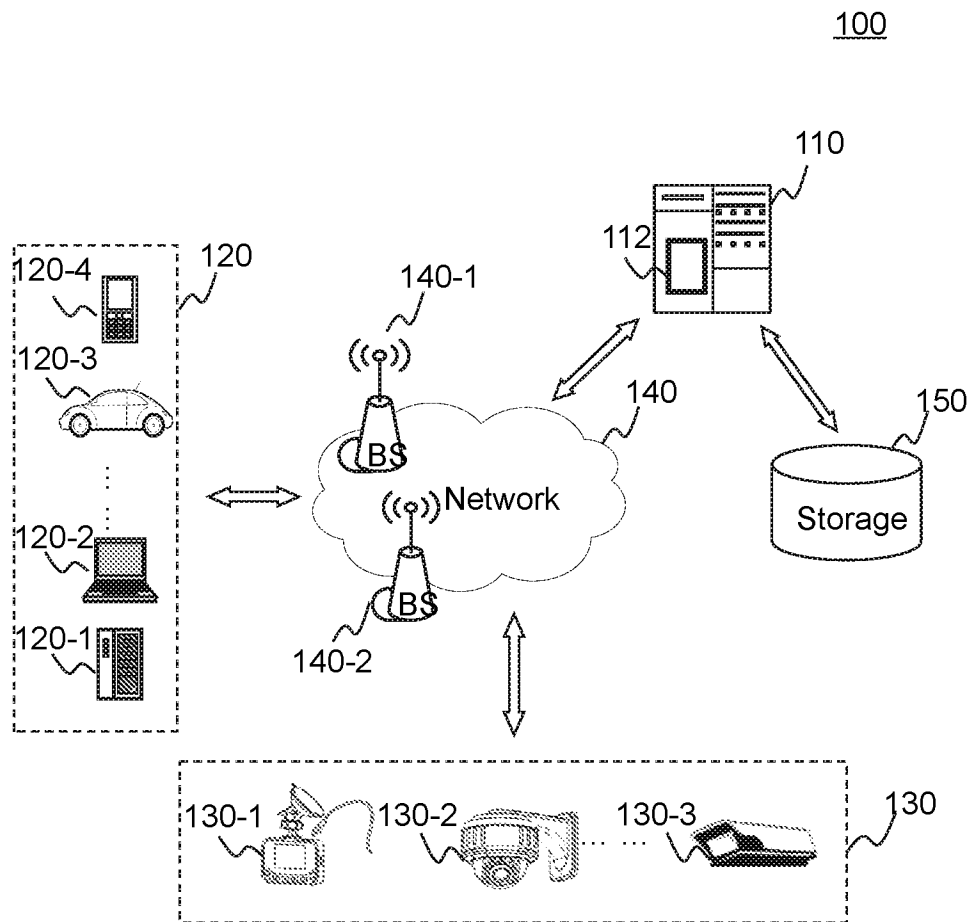


FIG. 1

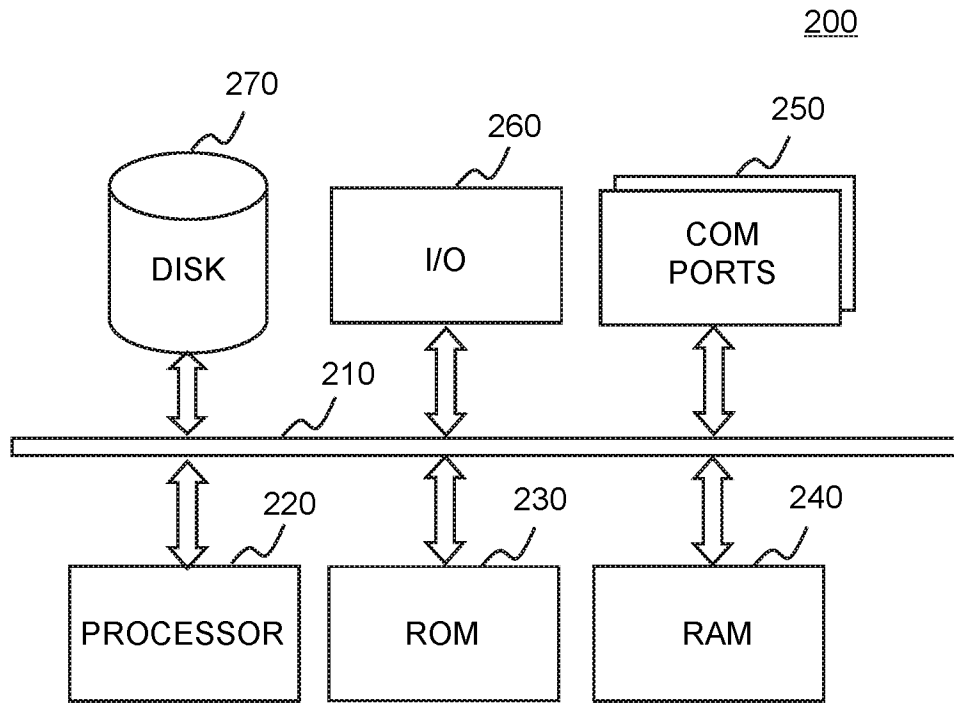


FIG. 2

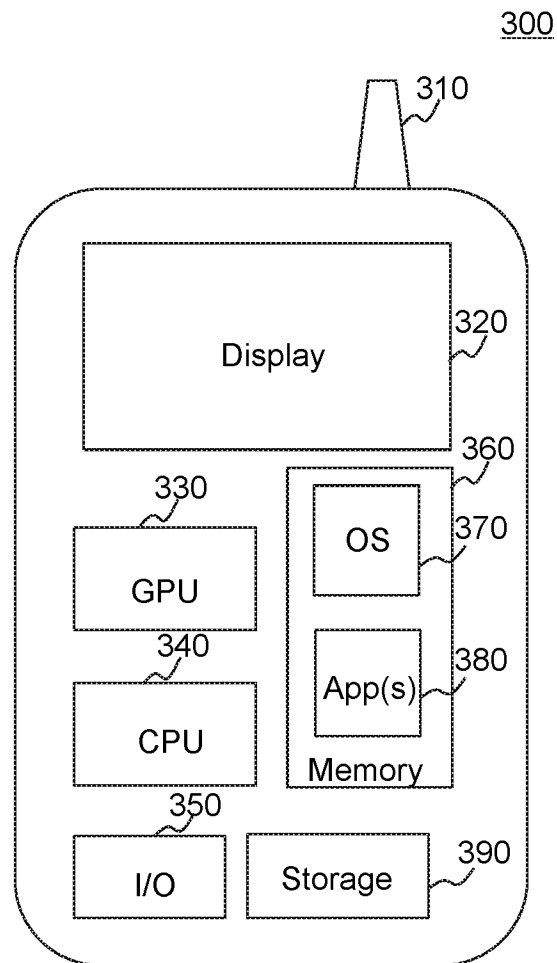


FIG. 3

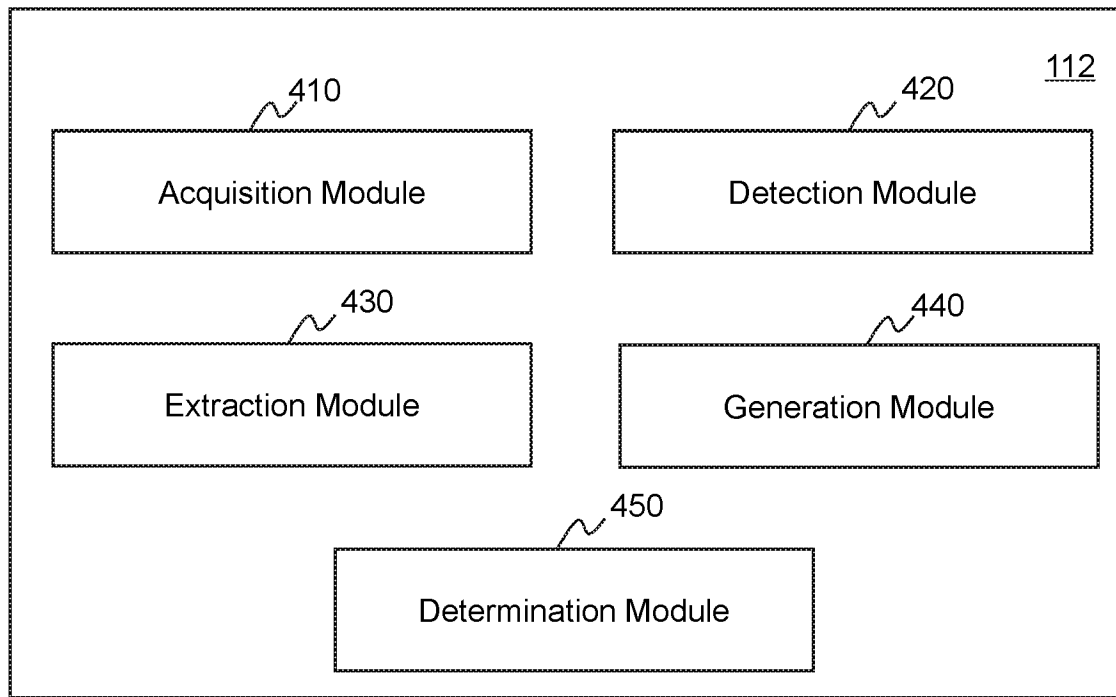


FIG. 4

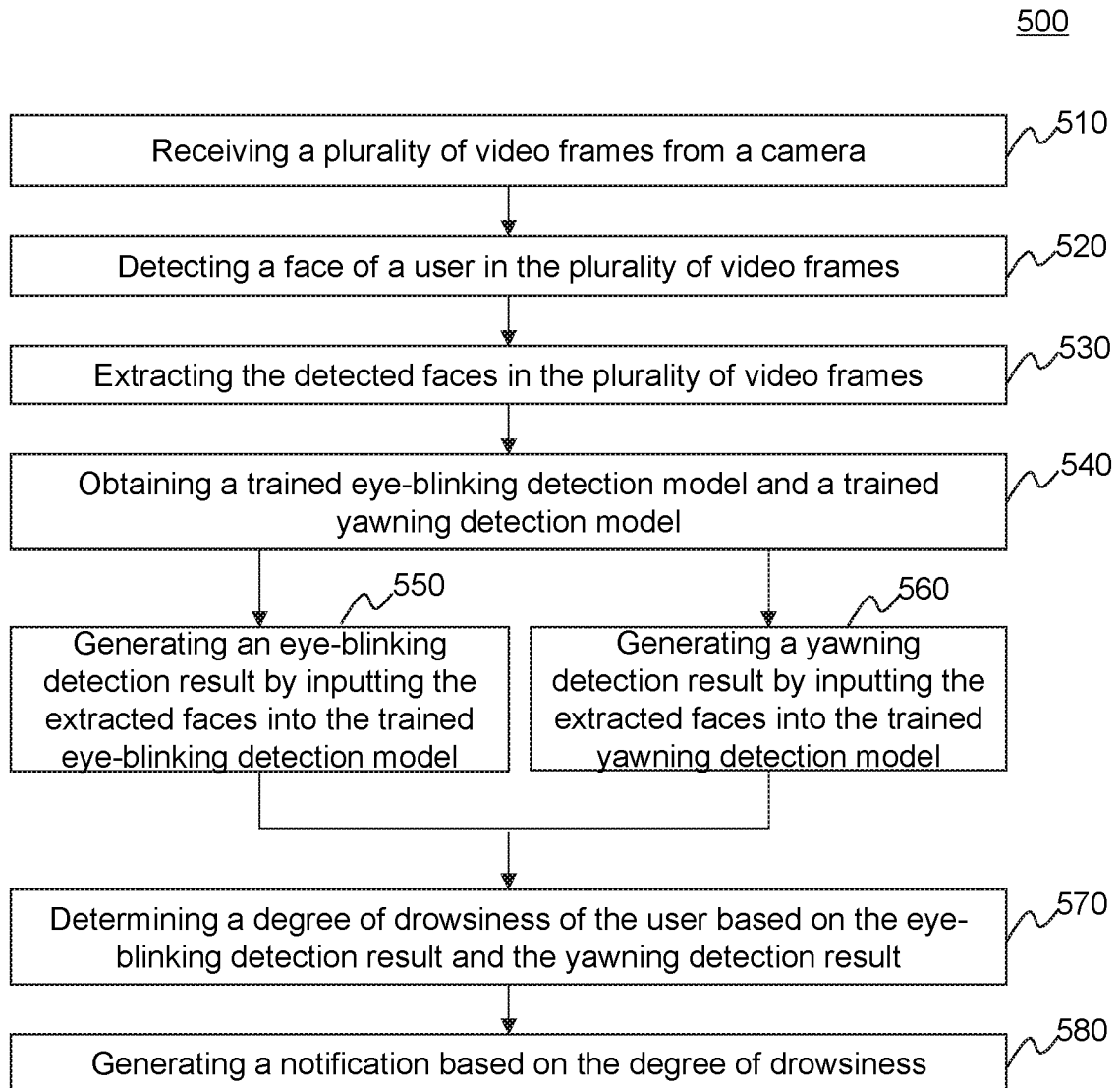


FIG. 5

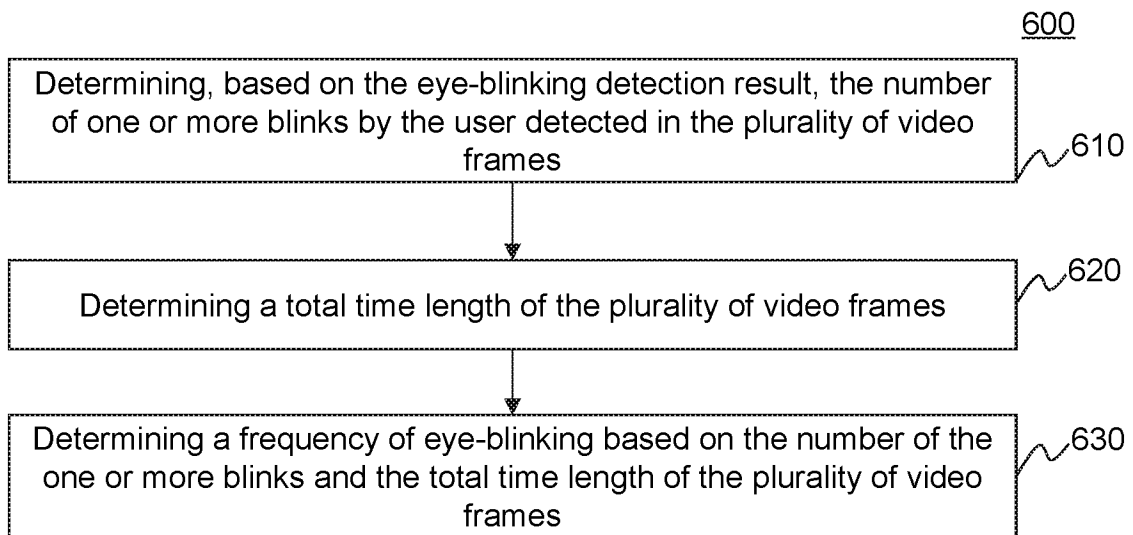


FIG. 6

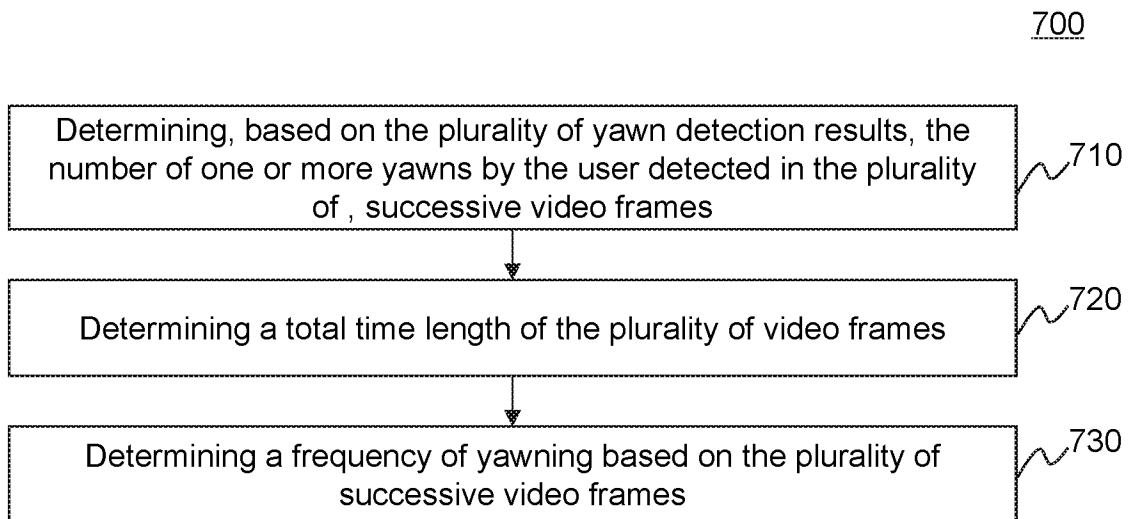


FIG. 7

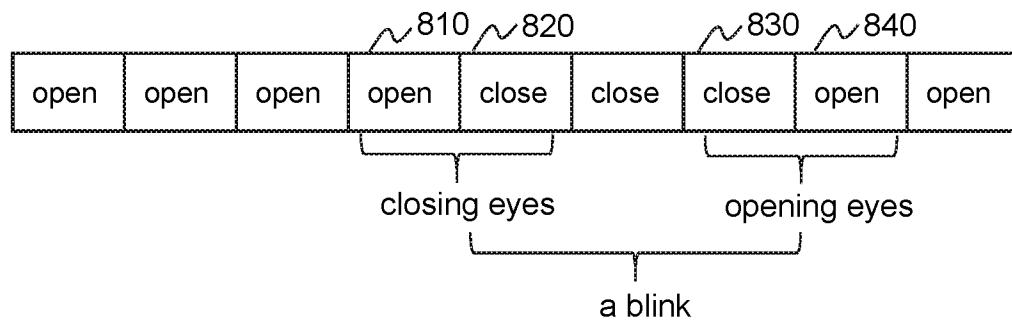


FIG. 8

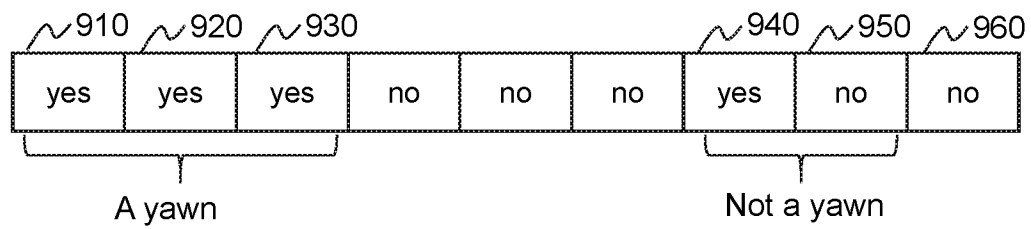


FIG. 9

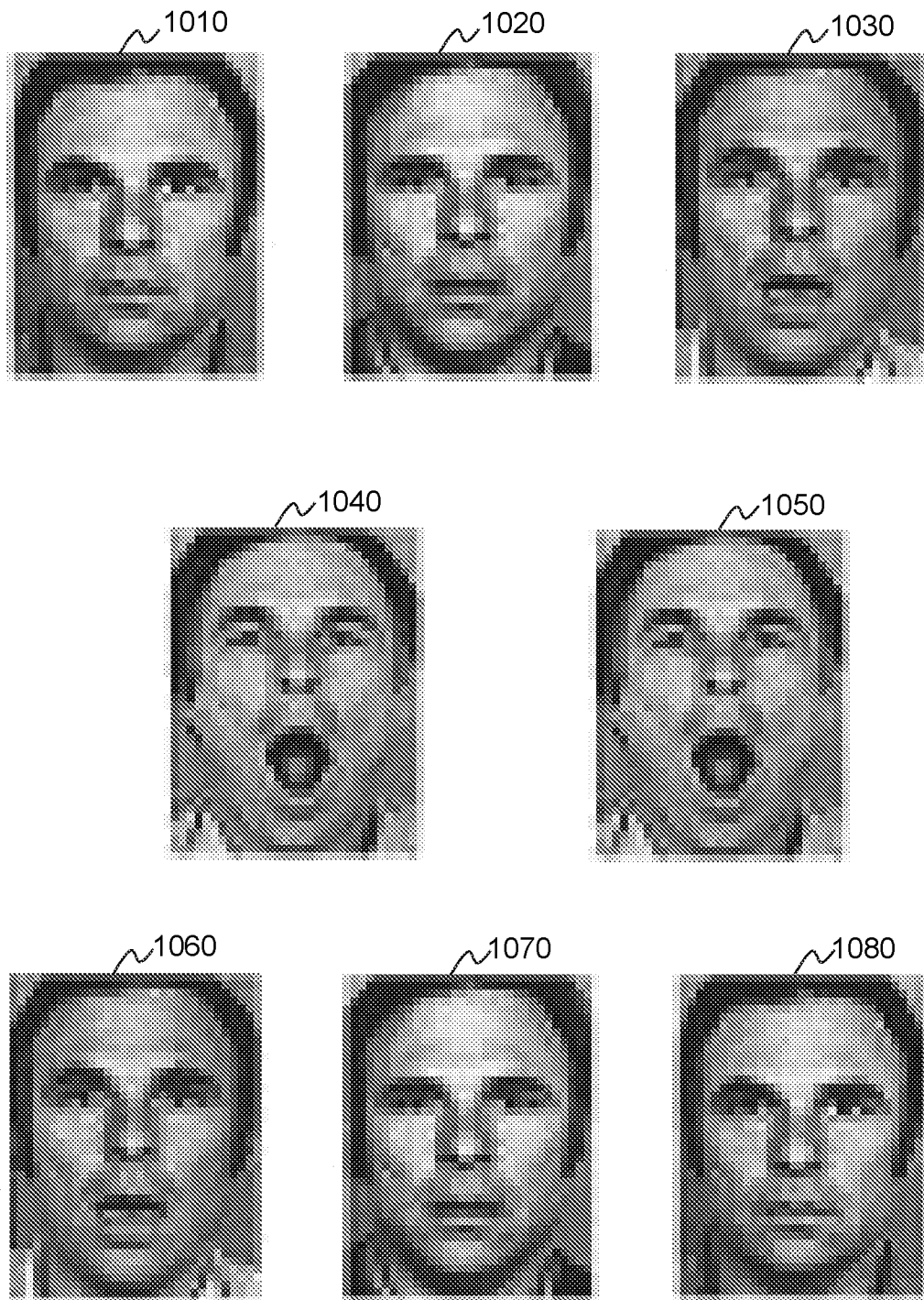


FIG. 10

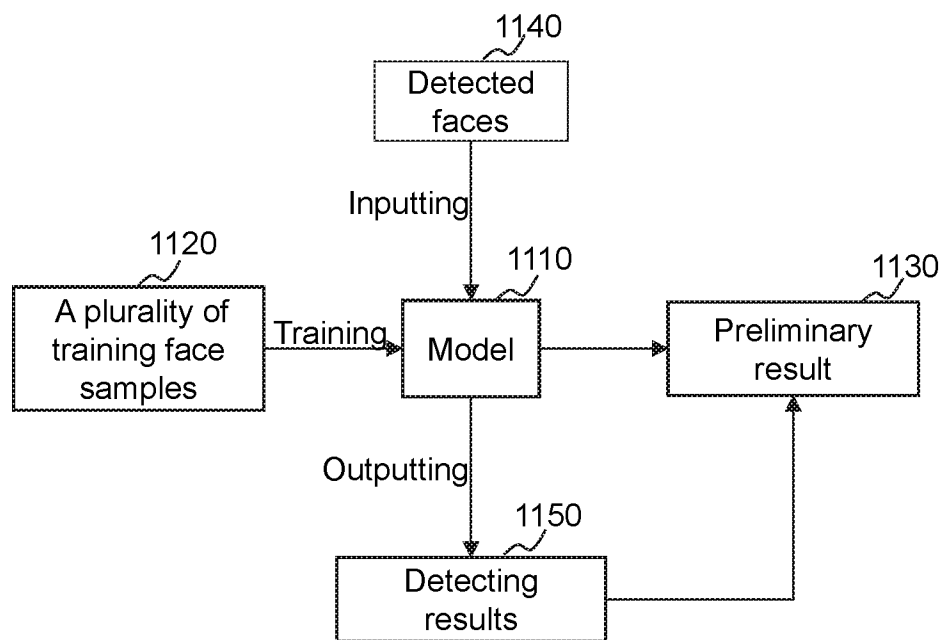


FIG. 11

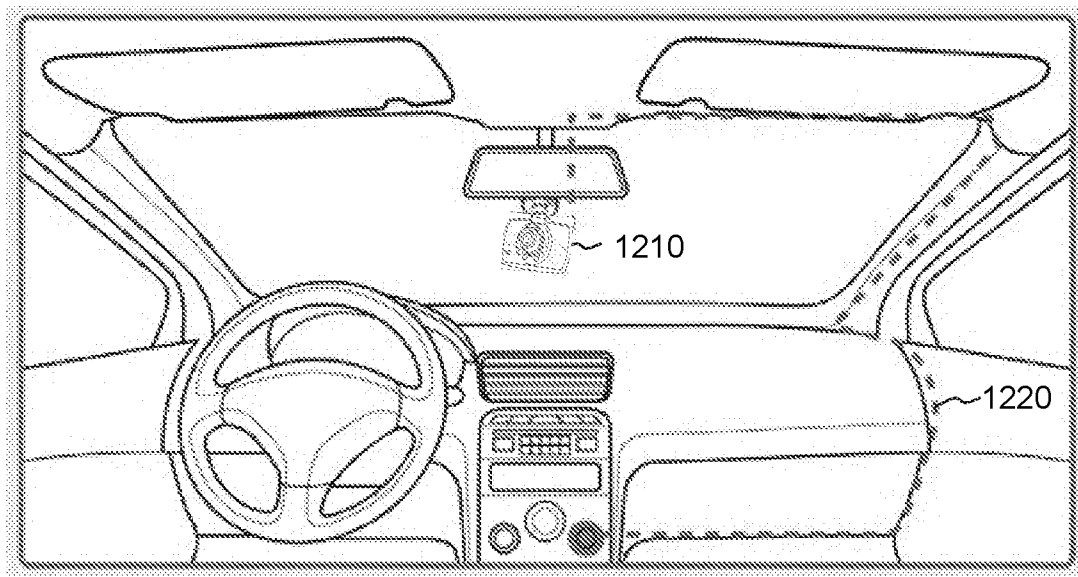


FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/105132

A. CLASSIFICATION OF SUBJECT MATTER

G01C 21/00(2006.01)i; G06K 9/00(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)

G01C21;G06K9

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)

CNPAT,CNKI,WPI,EPODOC:drows+,driver,video,image,face,eye,blink+,yawn+,notif+,alert+,model+,fatigue,detect+,monit+,train+,yaw+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 106446811 A (BEIJING ICE TECH. CO., LTD.) 22 February 2017 (2017-02-22) paragraphs [0060]-[0145], figures 1-2, in the description	1-3, 7-14, 18-24
Y	CN 107491769 A (UNIVERSITY CHINA GEOSCIENCES WUHAN) 19 December 2017 (2017-12-19) paragraphs [0008]-[0071], figures 1-5, in the description	1-3, 7-14, 18-24
A	CN 107697069 A (SAIC MOTOR CO., LTD.) 16 February 2018 (2018-02-16) the whole document	1-24
A	CN 106295600 A (NINGBO OSIGHT OPTOELECTRONIC CO., LTD.) 04 January 2017 (2017-01-04) the whole document	1-24
A	CN 108294759 A (UNIVERSITY TIANJIN POLYTECHNIC) 20 July 2018 (2018-07-20) the whole document	1-24
A	EP 2557549 A1 (TOYOTA JIDOSHA K.K.) 13 February 2013 (2013-02-13) the whole document	1-24
A	US 2004170304 A1 (AGILENT TECHNOLOGIES, INC. ET AL.) 02 September 2004 (2004-09-02) the whole document	1-24

☐ 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

09 May 2019

Date of mailing of the international search report

29 May 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088
China

Authorized officer

XU,Jianying

Facsimile No. (86-10)62019451

Telephone No. 86-(10)53962421

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/105132

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106446811	A	22 February 2017	None			
CN	107491769	A	19 December 2017	None			
CN	107697069	A	16 February 2018	None			
CN	106295600	A	04 January 2017	None			
CN	108294759	A	20 July 2018	None			
EP	2557549	A1	13 February 2013	JP	5447657	B2	19 March 2014
				US	8866896	B2	21 October 2014
				WO	2011125166	A1	13 October 2011
				JP	WO2011125166	A1	08 July 2013
				US	2013021463	A1	24 January 2013
US	2004170304	A1	02 September 2004	US	7280678	B2	09 October 2007
				DE	60324763	D1	02 January 2009
				JP	4593942	B2	08 December 2010
				EP	1452127	B1	19 November 2008
				JP	2004261598	A	24 September 2004
				EP	1452127	A1	01 September 2004