

SYSTEMS AND METHODS FOR DETECTING TRAFFIC ACCIDENTS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to Chinese Patent Application No. 202011314812.5 filed on November 20, 2020, the entire contents of which are hereby incorporated by reference.

TECHNICAL FIELD

[0002] The present disclosure generally relates to image processing, and in particular, to systems and methods for detecting a traffic accident by processing at least one image.

BACKGROUND

[0003] When a rescue is not timely or on-site protection is not adequate for a traffic accident, a road paralysis and/or a secondary accident may be easily caused. A current accident detection system mainly relies on manual observation, which is difficult to simultaneously monitor a large number of traffic scenes and efficiently detect traffic accidents therein. Therefore, it is desirable to provide systems and methods for automatically and efficiently detecting traffic accidents in traffic scenes, thereby improving the efficiency of traffic supervision.

SUMMARY

[0004] According to one aspect of the present disclosure, a system may be provided. The system may include: at least one storage device including a set of instructions; and at least one processor in communication with the at least one storage device, wherein when executing the set of instructions, the at least one processor is configured to cause the system to: obtain a video of a target scene that is captured by an image capture device; determine a target image associated with the target scene based on the video; determine a static region associated with the target scene in the target image; and determine a result indicating whether a traffic

accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image.

[0005] In some embodiments, wherein to determine a target image associated with the target scene based on the video, the at least one processor may be configured to cause the system to: obtain, from the video, a first image; obtain, from the video, one or more second images captured prior to the first image; and determine the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images.

[0006] In some embodiments, wherein to determine the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images, the at least one processor may be configured to cause the system to: generate a fusion result by fusing the first image and the one or more second images; and determine the target image based on the fusion result.

[0007] In some embodiments, a time interval between two adjacent images among the one or more second images and the first image may relate to a type of the target scene.

[0008] In some embodiments, wherein to determine a static region associated with the target scene in the target image, the at least one processor may be configured to cause the system to: determine one or more regions associated with one or more static objects of the target scene in the target image by processing the target image based on a second model; and determine the static region associated with the target scene in the target image based on the one or more regions associated with the one or more static objects of the target scene in the target image.

[0009] In some embodiments, the second model may include a detection model; and the one or more regions associated with the one or more static objects of the target scene in the target image may include one or more bounding boxes of the one or more static objects determined by the detection model.

[0010] In some embodiments, the second model may include a segmentation

model; and the one or more regions associated with the one or more static objects of the target scene in the target image may include one or more connected components of the one or more static objects determined by the segmentation model.

[0011] In some embodiments, wherein to determine the static region associated with the target scene in the target image based on the one or more regions associated with the one or more static objects of the target scene in the target image, the at least one processor may be configured to cause the system to: determine at least one of the one or more regions that satisfies a first condition; and generate a combination result by combining the at least one region; and determine the static region associated with the target scene in the target image based on the combination result.

[0012] In some embodiments, the first condition may relate to at least one distance between the one or more regions.

[0013] In some embodiments, wherein to determine a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image, the at least one processor may be configured to cause the system to: determine a confidence value associated with the traffic accident by processing, based on the first model, the static region associated with the target scene in the target image; and in response to determining that the confidence value satisfies a second condition, determine that the traffic accident happens in the static region.

[0014] In some embodiments, wherein to determine a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image, the at least one processor may be configured to cause the system to: in response to determining that the confidence value fails to satisfy the second condition, determine whether there are at least two of the one or more regions associated with the one or more static

objects in the static region that overlap; and in response to determining that there are the at least two of the one or more regions associated with the one or more static objects in the static region that overlap, modify the confidence value associated with the traffic accident.

[0015] In some embodiments, the one or more regions associated with the one or more static objects including one or more two-dimensional regions; and to modify the confidence value associated with the traffic accident, the at least one processor may be configured to cause the system to: determine one or more three-dimensional regions associated with the one or more static objects in the static region by performing a three-dimensional transformation on the one or more two-dimensional regions associated with the one or more static objects based on one or more coordinates associated with the one or more two-dimensional regions; and modify the confidence value associated with the traffic accident based on an overlap between the one or more three-dimensional regions associated with the one or more static objects in the static region.

[0016] According to another aspect of the present disclosure, a method may be provided. The method may be implemented on a computing device having at least one processor, at least one storage medium, and a communication platform connected to a network. The method may include: obtaining a video of a target scene that is captured by an image capture device; determining a target image associated with the target scene based on the video; determining a static region associated with the target scene in the target image; and determining a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image.

[0017] In some embodiments, wherein the determining a target image associated with the target scene based on the video may include: obtaining, from the video, a first image; obtaining, from the video, one or more second images captured prior to

the first image; and determining the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images.

[0018] In some embodiments, wherein the determining the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images may include: generating a fusion result by fusing the first image and the one or more second images; and determining the target image based on the fusion result.

[0019] In some embodiments, a time interval between two adjacent images among the one or more second images and the first image may relate to a type of the target scene.

[0020] In some embodiments, wherein the determining a static region associated with the target scene in the target image may include: determining one or more regions associated with one or more static objects of the target scene in the target image by processing the target image based on a second model; and determining the static region associated with the target scene in the target image based on the one or more regions associated with the one or more static objects of the target scene in the target image.

[0021] In some embodiments, the second model may include a detection model; and the one or more regions associated with the one or more static objects of the target scene in the target image may include one or more bounding boxes of the one or more static objects determined by the detection model.

[0022] In some embodiments, the second model may include a segmentation model; and the one or more regions associated with the one or more static objects of the target scene in the target image may include one or more connected components of the one or more static objects determined by the segmentation model.

[0023] In some embodiments, wherein the determining the static region associated

with the target scene in the target image based on the one or more regions associated with the one or more static objects of the target scene in the target image may include: determining at least one of the one or more regions that satisfies a first condition; and generating a combination result by combining the at least one region; and determining the static region associated with the target scene in the target image based on the combination result.

[0024] In some embodiments, the first condition may relate to at least one distance between the one or more regions.

[0025] In some embodiments, wherein the determining a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image may include: determining a confidence value associated with the traffic accident by processing, based on the first model, the static region associated with the target scene in the target image; and in response to determining that the confidence value satisfies a second condition, determining that the traffic accident happens in the static region.

[0026] In some embodiments, wherein the determining a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image may include: in response to determining that the confidence value fails to satisfy the second condition, determining whether there are at least two of the one or more regions associated with the one or more static objects in the static region that overlap; and in response to determining that there are the at least two of the one or more regions associated with the one or more static objects in the static region that overlap, modifying the confidence value associated with the traffic accident.

[0027] In some embodiments, the one or more regions associated with the one or more static objects may include one or more two-dimensional regions; and the modifying the confidence value associated with the traffic accident may include: determining one or more three-dimensional regions associated with the one or more

static objects in the static region by performing a three-dimensional transformation on the one or more two-dimensional regions associated with the one or more static objects based on one or more coordinates associated with the one or more two-dimensional regions; and modifying the confidence value associated with the traffic accident based on an overlap between the one or more three-dimensional regions associated with the one or more static objects in the static region.

[0028] According to another aspect of the present disclosure, a non-transitory computer readable medium may be provided. The non-transitory computer readable medium may include executable instructions that, when executed by at least one processor, directs the at least one processor to perform a method. The method may include: obtaining a video of a target scene that is captured by an image capture device; determining a target image associated with the target scene based on the video; determining a static region associated with the target scene in the target image; and determining a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image.

[0029] 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

[0030] The present disclosure is further described in terms of exemplary embodiments. These exemplary embodiments are described in detail with reference to the drawings. The drawings are not to scale. These embodiments are non-limiting schematic embodiments, in which like reference numerals represent

similar structures throughout the several views of the drawings, and wherein:

[0031] FIG. 1 is a schematic diagram illustrating an exemplary accident detection system according to some embodiments of the present disclosure;

[0032] FIG. 2 is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary computing device according to some embodiments of the present disclosure;

[0033] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary mobile device according to some embodiments of the present disclosure;

[0034] FIG. 4 is a block diagram illustrating an exemplary processing device according to some embodiments of the present disclosure;

[0035] FIG. 5 is a flowchart illustrating an exemplary process for determining a traffic accident result according to some embodiments of the present disclosure;

[0036] FIGs. 6A-6C are schematic diagrams illustrating exemplary traffic accidents according to some embodiments of the present disclosure;

[0037] FIG. 7 is a schematic diagram illustrating an exemplary target image according to some embodiments of the present disclosure;

[0038] FIG. 8 is a schematic diagram illustrating an exemplary process for adjusting a size of an image according to some embodiments of the present disclosure;

[0039] FIGs. 9A and 9B are schematic diagrams illustrating an exemplary process for determining a static region in a target image using a detection model according to some embodiments of the present disclosure;

[0040] FIGs. 10A-10C are schematic diagrams illustrating an exemplary process for determining a static region in a target image using a segmentation model according to some embodiments of the present disclosure;

[0041] FIG. 11 is a schematic diagram illustrating an exemplary 3D transformation result according to some embodiments of the present disclosure;

[0042] FIG. 12 is a schematic diagram illustrating an exemplary first model

according to some embodiments of the present disclosure;

[0043] FIG. 13 is a block diagram illustrating an exemplary accident detection system according to some embodiments of the present disclosure;

[0044] FIG. 14 is a flowchart illustrating an exemplary process for determining a traffic accident result according to some embodiments of the present disclosure;

[0045] FIG. 15 is a flowchart illustrating an exemplary process for extracting a static region in a target image according to some embodiments of the present disclosure;

[0046] FIG. 16 is a flowchart illustrating an exemplary process for determining a static region in a target image according to some embodiments of the present disclosure;

[0047] FIG. 17 is a flowchart illustrating an exemplary process for determining a static region in a target image according to some embodiments of the present disclosure;

[0048] FIG. 18 is a flowchart illustrating an exemplary process for determining a static region in a target image according to some embodiments of the present disclosure;

[0049] FIG. 19 is a flowchart illustrating an exemplary process for determining a static region in a target image according to some embodiments of the present disclosure;

[0050] FIG. 20 is a flowchart illustrating an exemplary process for modifying a confidence value associated with a traffic accident result according to some embodiments of the present disclosure;

[0051] FIG. 21 is a flowchart illustrating an exemplary process for modifying a confidence value associated with a traffic accident result according to some embodiments of the present disclosure;

[0052] FIG. 22 is a block diagram illustrating an exemplary electronic device according to some embodiments of the present disclosure; and

[0053] FIG. 23 is a block diagram illustrating an exemplary computer readable

storage medium according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0054] In the following detailed description, numerous specific details are set forth by way of examples in order to provide a thorough understanding of the relevant disclosure. However, it should be apparent to those skilled in the art that the present disclosure may be practiced without such details. In other instances, well-known methods, procedures, systems, components, and/or circuitry have been described at a relatively high-level, without detail, in order to avoid unnecessarily obscuring aspects of the present disclosure. 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 to be accorded the widest scope consistent with the claims.

[0055] 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 understood that the terms “system,” “unit,” “module,” and/or “block” used herein are one method to distinguish different components, elements, parts, sections or assemblies of different levels in ascending order. However, the terms may be displaced by another expression if they achieve the same purpose.

[0056] The modules (or units, blocks, units) described in the present disclosure may be implemented as software and/or hardware modules and may be stored in any type of non-transitory computer-readable medium or other storage devices. In some embodiments, a software module may be compiled and linked into an executable program. It will be appreciated that software modules can be callable from other modules or from themselves, and/or can be invoked in response to

detected events or interrupts. Software modules configured for execution on computing devices can be provided on a computer readable medium, such as a compact disc, a digital video disc, a flash drive, a magnetic disc, or any other tangible medium, or as a digital download (and can be originally stored in a compressed or installable format that requires installation, decompression, or decryption prior to execution). Such software code can be stored, partially or fully, on a memory device of the executing computing device, for execution by the computing device. Software instructions can be embedded in a firmware, such as an EPROM. It will be further appreciated that hardware modules (e.g., circuits) can be included of connected or coupled logic units, such as gates and flip-flops, and/or can be included of programmable units, such as programmable gate arrays or processors. The modules or computing device functionality described herein are preferably implemented as hardware modules, but can be software modules as well. In general, the modules described herein refer to logical modules that can be combined with other modules or divided into units despite their physical organization or storage.

[0057] It will be understood that when a unit, engine, module or block is referred to as being “on,” “connected to,” or “coupled to,” another unit, engine, module, or block, it may be directly on, connected or coupled to, or communicate with the other unit, engine, module, or block, or an intervening unit, engine, module, or block may be present, unless the context clearly indicates otherwise. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0058] 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.

[0059] The flowcharts used in the present disclosure illustrate operations that systems implement according to some embodiments of the present disclosure. It is to be expressly understood, the operations of the flowcharts 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.

[0060] An aspect of the present disclosure relates to systems and methods for determining a traffic accident result based on a video of a target scene using a first model (also referred to as a traffic accident detection model). The systems may determine a target image associated with the target scene based on the video. A dynamic region (e.g., including one or more moving objects) in the target image may be blurred. The systems may also extract a static region (e.g., including one or more static objects) associated with the target scene in the target image based on a second model (e.g., a detection model, a segmentation model). Further, the systems may determine the traffic accident result by processing the static region in the target image based on the first model.

[0061] In some application scenarios, when a traffic accident happens in a traffic scene, one or more objects (e.g., a vehicle, a driver, a pedestrian) associated with the traffic accident in the traffic scene may be in a static state while one or more objects (e.g., a vehicle, a driver) irrelevant with the traffic accident in the traffic scene may be in a motion state. In some embodiments of the present disclosure, the systems may first extract the static region and then determine the traffic accident result by merely taking the static region in the target image into consideration, thereby not only ensuring the accuracy of the traffic accident detection but also improving the efficiency of the traffic accident detection.

[0062] FIG. 1 is a schematic diagram illustrating an exemplary accident detection

system according to some embodiments of the present disclosure. In some embodiments, the accident detection system 100 may include an image capture device 110, a processing device 120, one or more terminals 130, a network 140, and a storage device 150.

[0063] The image capture device 110 may be configured to acquire an image (e.g., a traffic image) or a video (e.g., a traffic video). The image or the video may be two-dimensional (2D), three-dimensional (3D), etc. In some embodiments, the image capture device 110 may include a camera, a video camera, a mobile phone, a tablet, an e-book reader, or the like, or any combination thereof. In some embodiments, a position and/or an orientation (e.g., an angle, a direction) of the image capture device 110 may be adjustable. In some embodiments, the image capture device 110 may acquire image(s) continuously, within a specific time interval, or based on a control command. The acquired image(s) may be stored in the storage device 150, or transmitted to other components via the network 140 for storage or further processing.

[0064] In some embodiments, the image capture device 110 may communicate with one or more components of the accident detection system 100 (e.g., the processing device 120, the terminal(s) 130, the storage device 150) via the network 140. In some embodiments, the image capture device 110 may be directly connected to one or more components of the accident detection system 100 (e.g., the processing device 120, the terminal(s) 130, the storage device 150).

[0065] In some embodiments, the processing device 120 may include a single server or a server group. The server group may be centralized or distributed (e.g., the processing device 120 may be a distributed system). In some embodiments, the processing device 120 may be local or remote. For example, the processing device 120 may access information and/or data stored in the image capture device 110, the terminal(s) 130, and/or the storage device 150 via the network 140. As another example, the processing device 120 may be directly connected to the image

capture device 110, the terminal(s) 130, and/or the storage device 150 to access stored information and/or data. In some embodiments, the processing device 120 may be implemented on a cloud platform or an onboard computer. 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 any combination thereof. In some embodiments, the processing device 120 may be implemented on a computing device 200 including one or more components illustrated in FIG. 2 of the present disclosure.

[0066] In some embodiments, the processing device 120 may process information and/or data associated with traffic accident detection to perform one or more functions described in the present disclosure. For example, the processing device 120 may determine a target image associated with a target scene based on a video of the target scene. The target image may include a static region associated with the target scene and a dynamic region associated with the target scene. The processing device 120 may determine a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image. In some embodiments, the processing device 120 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 device 120 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 any combination thereof. In some embodiments, the image capture device 110 may be part of the processing device 120.

[0067] In some embodiments, the processing device 120 may be connected to the

network 140 to communicate with one or more components (e.g., the image capture device 110, the terminal(s) 130, the storage device 150) of the accident detection system 100. In some embodiments, the processing device 120 may be directly connected to or communicate with one or more components (e.g., the image capture device 110, the terminal(s) 130, the storage device 150) of the accident detection system 100.

[0068] The terminal(s) 130 may be configured to receive/transmit data and/or information from the image capture device 110, the processing device 120, and/or the storage device 150. For example, the user terminal(s) 130 may receive a traffic accident result from the processing device 120. As another example, the user terminal(s) 130 may transmit an instruction to the image capture device 110 and/or the processing device 120. In some embodiments, the terminal(s) 130 may include a tablet computer 130-1, a laptop computer 130-2, a built-in device in a vehicle 130-3, a mobile device 130-4, or the like, or any combination thereof.

[0069] In some embodiments, the terminal(s) 130 may communicate with one or more components of the accident detection system 100 (e.g., the image capture device 110, the processing device 120, the storage device 150) via the network 140. In some embodiments, the terminal(s) 130 may be directly connected to one or more components of the accident detection system 100 (e.g., the image capture device 110, the processing device 120, the storage device 150).

[0070] The network 140 may facilitate exchange of information and/or data. In some embodiments, one or more components (e.g., the image capture device 110, the processing device 120, the terminal(s) 130, the storage device 150) of the accident detection system 100 may transmit information and/or data to other component(s) of the accident detection system 100 via the network 140. For example, the processing device 120 may obtain a video of a target scene from the storage device 150 via the network 140. In some embodiments, the network 140 may be any type of wired or wireless network, or a 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 any 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 (e.g., a point 140-1, a point 140-2), through which one or more components of the accident detection system 100 may be connected to the network 140 to exchange data and/or information.

[0071] The storage device 150 may store data and/or instructions. In some embodiments, the storage device 150 may store data obtained from the image capture device 110, the processing device 120, the terminal(s) 130, or an external storage device. For example, the storage device 150 may store a video of a target scene acquired by the image capture device 110. In some embodiments, the storage device 150 may store data and/or instructions that the processing device 120 may execute or use to perform exemplary methods described in the present disclosure. For example, the storage device 150 may store instructions that the processing device 120 may execute or use to determine a result indicating whether a traffic accident happens in a static region by processing, based on a first model, the static region associated with a target scene in a target image.

[0072] In some embodiments, the storage device 150 may include a mass storage, a removable storage, a volatile read-and-write memory, a read-only memory (ROM), or the like, or any combination thereof. Exemplary mass storage may include a magnetic disk, an optical disk, a solid-state drive, 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

a 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 device 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 any combination thereof.

[0073] In some embodiments, the storage device 150 may be connected to the network 140 to communicate with one or more components (e.g., the image capture device 110, the processing device 120, the terminal(s) 130) of the accident detection system 100. One or more components of the accident detection system 100 may access the data or instructions stored in the storage device 150 via the network 140. In some embodiments, the storage device 150 may be directly connected to or communicate with one or more components (e.g., the image capture device 110, the processing device 120, the terminal(s) 130) of the accident detection system 100. In some embodiments, the storage device 150 may be part of the processing device 120. For example, the storage device 150 may be integrated into the processing device 120.

[0074] In some embodiments, two or more components of the accident detection system 100 may be integrated in a same device. For example, the image capture device 110, the processing device 120, and the storage device 150 may be integrated into a same device (e.g., a camera, a smartphone, a computer, a workstation, a server). In some embodiments, one or more components of the accident detection system 100 may be remote from each other. For example, the

image capture device 110 may be remote from the processing device 120.

[0075] It should be noted that the accident detection system 100 is merely provided for the purposes of illustration, and is not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations or 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.

[0076] In some embodiments, the accident detection system 100 may also include a user device (not shown) configured to receive information and/or data from the image capture device 110, the processing device 120, the terminal(s) 130, and/or the storage device 150. The user device may provide a user interface via which a user may view information (e.g., a traffic accident result) and/or input data (e.g., a video of a target scene) and/or instructions to the accident detection system 100.

[0077] In some embodiments, the accident detection system 100 may also include an input/output (I/O) component. The I/O component may be configured to input or output signals, data, or information. In some embodiments, the I/O component may include an input device and an output device. Exemplary input devices may include a keyboard, a mouse, a touch screen, a microphone, or the like, or any combination thereof. Exemplary output devices may include a display device, a speaker, a printer, a projector, or the like, or any combination thereof.

[0078] FIG. 2 is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary computing device according to some embodiments of the present disclosure. The computing device 200 may be used to implement any component of the accident detection system 100 as described herein. For example, the processing device 120 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 traffic accident detection as described herein may be implemented in a

distributed fashion on a number of similar platforms to distribute the processing load.

[0079] 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 (e.g., 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.

[0080] The computing device 200 may further include one or more storages configured to store various data files (e.g., program instructions) to be processed and/or transmitted by the computing device 200. In some embodiments, the one or more storages may include a high speed random access memory (not shown), a non-volatile memory (e.g., a magnetic storage device, a flash memory, or other non-volatile solid state memories) (not shown), a disk 270, a read-only memory (ROM) 230, a random-access memory (RAM) 240, or the like, or any combination thereof. In some embodiments, the one or more storages may further include a remote storage corresponding to the processor 220. The remote storage may connect to the computing device 200 via the network 140. The computing device 200 may also include program instructions stored in the one or more storages (e.g., 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 may also include an I/O component 260, supporting input/output between the

computing device 200 and other components. The computing device 200 may also receive programming and data via network communications.

[0081] Merely for illustration, only one processor is illustrated in FIG. 2. Multiple processors 220 are also contemplated; thus, operations and/or method steps performed by one processor 220 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 220 of the computing device 200 executes both operation A and operation B, it should be understood that operation A and operation B may also be performed by two different processors 220 jointly or separately in the computing device 200 (e.g., a first processor executes operation A and a second processor executes operation B, or the first and second processors jointly execute operations A and B).

[0082] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary mobile device according to some embodiments of the present disclosure. In some embodiments, the processing device 120 or the user device may be implemented on the mobile device 300.

[0083] 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, an I/O 350, a memory 360, a mobile operating system (OS) 370, and a storage 390. In some embodiments, any other suitable components, including but not limited to a system bus or a controller (not shown), may also be in the mobile device 300.

[0084] In some embodiments, the mobile operating system 370 (e.g., iOS™, Android™, Windows Phone™) 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 traffic accident detection or other information from the accident detection system 100. User interactions with the

information stream may be achieved via the I/O 350 and provided to the processing device 120 and/or other components of the accident detection system 100 via the network 140.

[0085] FIG. 4 is a block diagram illustrating an exemplary processing device according to some embodiments of the present disclosure. The processing device 120 may include a video obtainment module 410, an image determination module 420, a region determination module 430, and a detection module 440.

[0086] In some embodiments, the video obtainment module 410 may be configured to obtain a video of a target scene (e.g., a traffic scene) that is captured by an image capture device (e.g., the image capture device 110 in FIG. 1). In some embodiments, the video may be acquired by the image capture device in real-time. In some embodiments, the video may include a plurality of images of the target scene.

[0087] In some embodiments, the image determination module 420 may be configured to determine a target image (e.g., image 710 in FIG. 7) associated with the target scene based on the video. In some embodiments, the image determination module 420 may obtain a first image and one or more second images from the video. The one or more second images may be captured prior to the first image. The image determination module 420 may determine the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images. In some embodiments, the dynamic region may include one or more moving objects, for example, a moving vehicle, a moving pedestrian, a moving driver, etc., in the target scene.

[0088] In some embodiments, the image determination module 420 may generate a fusion result by fusing the first image and the one or more second images and determine the target image based on the fusion result. In some embodiments, the image determination module 420 may blur the dynamic region in the target image based on the fusion result, that is, the target image may include the blurred dynamic

region (e.g., a region 711 in FIG. 7) in the target scene.

[0089] In some embodiments, the image determination module 420 may adjust a size (e.g., an area, a length, a width, an aspect ratio) of the first image and/or the one or more second images, and then determine the target image based on the adjusted first image and/or the adjusted one or more second images accordingly. In some embodiments, the image determination module 420 may adjust the size of the first image and/or the one or more second images using a size adjustment technique. For example, the size adjustment technique may include a secondary nonlinear interpolation technique, a linear interpolation technique, etc. In some embodiments, the image determination module 420 may adjust the size of the first image and/or the one or more second images to a reference size. The reference size may be predetermined according to practical demands, for example, an accuracy of traffic accident detection, a type of a first model used to determine the traffic accident, a type of a second model used to determine a static region, etc.

[0090] In some embodiments, the image determination module 420 may enlarge a length of a short side of an image (e.g., the first image, the one or more second images). In some embodiments, an enlarged portion of the image may be filled with a color whose pixel value is a specific value, e.g., 127.

[0091] In some embodiments, the region determination module 430 may be configured to may determine a static region associated with the target scene in the target image. In some embodiments, the region determination module 430 may determine one or more regions associated with one or more static objects (e.g., a vehicle, a pedestrian) of the target scene in the target image by processing the target image (e.g., a portion of the target image excluding the dynamic region) based on a second model (e.g., a neural network model). As used herein, a static object may be stationary relative to a road surface. In some embodiments, the region determination module 430 may also determine a position associated with one of the one or more regions (e.g., a coordinate of a center of a region), a size (e.g., a length,

a width, an area) of one of the one or more regions, etc., based on the second model.

[0092] In some embodiments, the second model may include a detection model (e.g., you-only-look-once (YOLO) neural network model). The one or more regions associated with the one or more static objects of the target scene in the target image may include one or more bounding boxes (e.g., rectangles 2, 3, or 4 in FIG. 9A) (e.g., a two-dimensional bounding box) of the one or more static objects determined by the detection model. In some embodiments, each of the one or more bounding boxes may correspond to a static object.

[0093] In some embodiments, the second model may include a segmentation model. The one or more regions associated with the one or more static objects of the target scene in the target image may include one or more connected components (e.g., regions 1030 and 1040 in FIG. 10B) of the one or more static objects determined by the segmentation model. In some embodiments, each of the one or more connected components may correspond to a type (e.g., a vehicle type (e.g., region 1030 in FIG. 10B), a person type (e.g., region 1040 in FIG. 10B)) of the one or more static objects.

[0094] In some embodiments, the region determination module 430 may determine the static region associated with the target scene in the target image based on the one or more regions (e.g., the one or more bounding boxes, the one or more connected components) associated with the one or more static objects of the target scene in the target image. In some embodiments, the region determination module 430 may designate each of the one or more regions as one static region associated with the target scene in the target image. As used herein, a static region may refer to a minimum region containing one of the one or more static objects.

[0095] In some embodiments, the region determination module 430 may determine at least one of the one or more regions that satisfies a first condition. The first condition may relate to at least one distance between the one or more regions (e.g.,

one or more centers thereof). For example, the first condition may include that at least one distance between the at least one region is smaller than or equal to a distance threshold. As another example, the first condition may include that a ratio of a distance between two regions of the at least one region and a width of an object (e.g., a vehicle) inside one of the two regions is smaller than or equal to a ratio threshold (e.g., 1.2).

[0096] In some embodiments, the region determination module 430 may generate a combination result by combining the at least one region. For example, the combination result may include a combined bounding box (e.g., rectangle 6 in FIG. 9B, rectangle *f* in FIG. 10C) contained the at least one region. In some embodiments, the region determination module 430 may determine the static region associated with the target scene in the target image based on the combination result. In some embodiments, the region determination module 430 may designate a region (e.g., a minimum region containing the combined bounding box) corresponding to the combined bounding box as one static region. In some embodiments, the region determination module 430 may designate each region of a remaining portion of the one or more regions as one static region.

[0097] In some embodiments, since a distance between a region of the remaining portion of the one or more regions and one of the at least one region may be larger than the distance threshold, the remaining portion of the one or more regions may be probably irrelevant with a traffic accident. For example, the remaining portion of the one or more regions may include road plants 712 in FIG. 7 which are irrelevant with a traffic accident. In some embodiments, the region determination module 430 may exclude at least a portion of the remaining portion of the one or more regions from the static region. In some embodiments, the region determination module 430 may determine whether a distance between a region of the remaining portion of the one or more regions and one of the at least one region is larger than a second distance threshold. In response to determining that the distance between the region of the

remaining portion of the one or more regions and one of the at least one region is larger than the second distance threshold, the region determination module 430 may exclude the region from the static region, that is the region determination module 430 may not designate the region as one static region; in response to determining that the distance is smaller than or equal to the second distance threshold, the region determination module 430 may designate the region as one static region. In some embodiments, the second distance threshold may be larger than the distance threshold.

[0098] In some embodiments, the region determination module 430 may enlarge the bounding box (or the combined bounding box) by a preset ratio and designate a region corresponding to the enlarged bounding box (e.g., rectangle 6' in FIG. 9A) as the static region. In some embodiments, the ratio may be preset according to practical demands, for example, a type of the target scene, a size (e.g., a length, a width, an area, an aspect ratio) of the bounding box (or the combined bounding box), a size (e.g., a length, a width, an area, an aspect ratio) of an object inside the bounding box (or the combined bounding box). For example, the preset ratio may be 1.2. As another example, the preset ratio may be 0.5 of the width of a vehicle inside the bounding box.

[0099] In some embodiments, the detection module 440 may be configured to determine a result (also referred to as a traffic accident result) indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image. In some embodiments, the traffic accident result may also include a confidence value associated with the traffic accident, a type of the traffic accident, etc.

[0100] In some embodiments, the detection module 440 may modify the confidence value associated with the traffic accident. In some embodiments, in response to determining that the confidence value fails to satisfy a second condition (e.g., larger than the value threshold (e.g., 0.5)), the detection module 440 may

determine whether there are at least two of the one or more regions associated with the one or more static objects in the static region that overlap. In response to determining that there are at least two of the one or more regions associated with the one or more static objects in the static region that overlap, the detection module 440 may modify the confidence associated with the traffic accident; in response to determining that there are no at least two of the one or more regions associated with the one or more static objects in the static region that overlap, the detection module 440 may not modify the confidence value associated with the traffic accident and output a result indicating that there is no traffic accident happens in the static region. In response to determining that the confidence value satisfies the second condition, the detection module 440 may output a result indicating that a traffic accident happens in the static region.

[0101] In some embodiments, the one or more regions associated with the one or more static objects may be two-dimensional. The detection module 440 may determine one or more three-dimensional (3D) regions (e.g., rectangles 1110 and 1120 in FIG. 11) associated with the one or more static objects in the static region by performing a 3D transformation on the one or more 2D regions associated with the one or more static objects based on one or more coordinates associated with the one or more 2D regions (e.g., a coordinate of a vertex or a central point of a 2D region).

[0102] In some embodiments, the detection module 440 may modify the confidence value associated with the traffic accident based on an overlap (also referred to as an overlapping region) between the one or more 3D regions associated with the one or more static objects in the static region. The detection module 440 may determine a ratio between a volume of the overlap between the one or more 3D regions and a volume of one of the one or more 3D regions (e.g., a minimum volume of the one or more 3D regions) and modify the confidence value based on the ratio. In some embodiments, the detection module 440 may

determine the modified confidence value by performing a weighted summation on the ratio and the confidence value. For example, the detection module 440 may designate a value corresponding to the weighted summation result as the modified confidence value.

[0103] In some embodiments, in response to determining that the modified confidence value exceeds the value threshold, the detection module 440 may determine that the traffic accident happens in the static region. In some embodiments, in response to determining that the modified confidence value is smaller than or equal to the value threshold, the detection module 440 may determine there is no traffic accident in the static region.

[0104] The modules in the processing device 120 may be connected to or communicated 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 any 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 any 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. For example, the image determination module 420 and the region determination module 430 may be combined as a single module which may determine the target image associated with the target scene based on the video and determine the static region associated with the target scene in the target image. As another example, the processing device 120 may include a storage module (not shown) which may be used to store data generated by the above-mentioned modules.

[0105] FIG. 5 is a flowchart illustrating an exemplary process for determining a traffic accident result according to some embodiments of the present disclosure. FIGs. 6A-6C are schematic diagrams illustrating exemplary traffic accidents according to some embodiments of the present disclosure. FIG. 7 is a schematic

diagram illustrating an exemplary target image according to some embodiments of the present disclosure. FIG. 8 is a schematic diagram illustrating an exemplary process for adjusting a size of an image according to some embodiments of the present disclosure. FIGs. 9A and 9B are schematic diagrams illustrating an exemplary process for determining a static region in a target image using a detection model according to some embodiments of the present disclosure. FIGs. 10A-10C are schematic diagrams illustrating an exemplary process for determining a static region in a target image using a segmentation model according to some embodiments of the present disclosure. FIG. 11 is a schematic diagram illustrating an exemplary 3D transformation result according to some embodiments of the present disclosure.

[0106] In some embodiments, the process 500 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 500. The operations of the illustrated process presented 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 illustrated in FIG. 5 and described below is not intended to be limiting.

[0107] In 510, the processing device 120 (e.g., the video obtainment module 410) (e.g., the interface circuits of the processor 220) may obtain a video of a target scene (e.g., a traffic scene) that is captured by an image capture device (e.g., the image capture device 110 in FIG. 1). In some embodiments, the video may be acquired by the image capture device in real-time.

[0108] In some embodiments, the video may include a plurality of images of the target scene, for example, arranged in a time order. A time interval between two

adjacent images of the plurality of images may be set according to practical demands. In some embodiments, the time interval between two adjacent images may relate to a type of the target scene (e.g., a traffic flow of the target scene, a speed limit of the target scene), an accuracy of traffic accident detection, a speed of traffic accident detection, etc. For example, assuming that vehicles in a first road (e.g., a high-speed road) travels faster than vehicles in a second road (e.g., a city road of a speed limit lower than the high-speed road), a time interval (e.g., 5 frames) between two adjacent images corresponding to the first road may be smaller than or equal to a time interval (10 frames) between two adjacent images corresponding to the second road. As another example, a time interval (e.g., 5 frames) between two adjacent images corresponding to a tunnel may be smaller than or equal to a time interval (10 frames) between two adjacent images corresponding to a city road.

[0109] In 520, the processing device 120 (e.g., the image determination module 420) (e.g., the processing circuits of the processor 220) may determine a target image (e.g., image 710 in FIG. 7) associated with the target scene based on the video. In some embodiments, the processing device 120 may obtain a first image and one or more second images from the video. The one or more second images may be captured prior to the first image. The processing device 120 may determine the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images. In some embodiments, the dynamic region may include one or more moving objects, for example, a moving vehicle, a moving pedestrian, a moving driver, etc., in the target scene. The dynamic region may include a minimum region containing the one or more moving objects. In some embodiments, the processing device 120 may determine the target image using an exponential smoothing and filtering technique, a Gaussian filtering technique, etc. In some embodiments, the processing device 120 may perform traffic accident detection for each of the plurality images in the video, and the first image may be each of the plurality images captured in a current time point

by the image capture device.

[0110] In some embodiments, the processing device 120 may generate a fusion result by fusing the first image and the one or more second images and determine the target image based on the fusion result. In some embodiments, the processing device 120 may blur the dynamic region in the target image based on the fusion result, that is, the target image may include the blurred dynamic region (e.g., a region 711 in FIG. 7) in the target scene. In some embodiments, the processing device 120 may determine the target image according to formula (1) below:

$$B = (1 - \alpha) \times B' + \alpha \times P \quad (1)$$

where B refers to the target image, B' refers to the one or more second images before the blurred operation and/or the fusion operation, P refers to the first image before the blurred operation and/or the fusion operation, and α refers to an exponential smoothing weight associated with the first image, which is used to control a reserving degree of information of the one or more second images after the blurred operation and/or the fusion operation. In some embodiments, the processing device 120 may set the exponential smoothing weight associated with the first image according to practical demands. For example, the exponential smoothing weight may be 0.15.

[0111] In some embodiments, the one or more second images may include all images captured before the first image and included in the video. In some embodiments, the one or more second images may include a preset count (e.g., 5, 10, 15) of images captured before the first image and included in the video. A time interval between two adjacent second images may be multiple of the time interval between two adjacent images of the plurality of images included in the video. In some embodiments, the processing device 120 may determine the target image by blurring the dynamic region in the first image directly using all of the one or more second images. In some embodiments, the processing device 120 may determine the target image by blurring a dynamic region in an intermediate image using at least

one of the one or more second images captured before the intermediate image (e.g., one of the one or more second images captured before the first image and next to the first image) and then blurring the dynamic region in the first image using the blurred intermediate image.

[0112] In some embodiments, the one or more second images and the first image may be arranged in a time order. A time interval between two adjacent images among the one or more second images and the first image may be the same as the time interval between two adjacent images among the plurality of images. In some embodiments, the time interval between two adjacent images among the one or more second images and the first image may be multiple of the time interval between two adjacent images among the plurality of images. In some embodiments, similar to the time interval between two adjacent images among the plurality of images, the time interval between two adjacent images among the one or more second images and the first image may relate to a type of the target scene (e.g., a traffic flow of the target scene, a speed limit of the target scene), an accuracy of traffic accident detection, a speed of traffic accident detection, etc.

[0113] In some embodiments, the processing device 120 may adjust a size (e.g., an area, a length, a width, an aspect ratio) of the first image and/or the one or more second images, and then determine the target image based on the adjusted first image and/or the adjusted one or more second images accordingly. In some embodiments, the processing device 120 may adjust the size of the first image and/or the one or more second images using a size adjustment technique. For example, the size adjustment technique may include a secondary linear interpolation technique, a nonlinear interpolation technique, etc. In some embodiments, the processing device 120 may adjust the size of the first image and/or the one or more second images to a reference size. The reference size may be predetermined according to practical demands, for example, an accuracy of traffic accident detection, a type of a first model used to determine the traffic accident, a type of a

second model used to determine a static region, etc.

[0114] In some embodiments, the processing device 120 may enlarge a length of a short side of an image (e.g., the first image, the one or more second images). In some embodiments, an enlarged portion of the image may be filled with a color whose pixel value is a specific value within a range of 0 and 255, e.g., 0, 127, 255. The specific value may be default of the accident detection system 100 or predetermined according to practical demands. In some embodiments, the processing device 120 may enlarge the length of the short side of the image to be the same as the length of the long side of the image. Take an image 810 in FIG. 8 as an example, a length of a long side of the image 810 is a , a length of a short side of the image 810 is b smaller than a . The processing device 120 may enlarge the length of the short side of the image 810 to c (same as or different from a) by filling an enlarged portion of the image 810 with a color whose pixel value is 127. Since the target image that is directly determined based on the first image and/or the one or more second images after the size adjustment may deform, affecting the accuracy of the traffic accident detection, the processing device 120 may first enlarge the length of the short side of the image and then perform the size adjustment of the image.

[0115] In 530, the processing device 120 (e.g., the region determination module 430) (e.g., the processing circuits of the processor 220) may determine a static region associated with the target scene in the target image. In some embodiments, the processing device 120 may determine one or more regions associated with one or more static objects (e.g., a vehicle, a pedestrian) of the target scene in the target image by processing the target image (e.g., a portion of the target image excluding the dynamic region) based on a second model (e.g., a neural network model). As used herein, a static object may be stationary relative to a road surface. Since the dynamic region is blurred before the static region is determined, the processing device 120 may determine the static region without identifying objects inside the

dynamic region, thereby reducing a computing load for determining the static region and improving an accuracy for determining the static region. In some embodiments, the processing device 120 may also determine a position associated with one of the one or more regions (e.g., a coordinate of a center of a region), a size (e.g., a length, a width, an area) of one of the one or more regions, etc., based on the second model.

[0116] In some embodiments, the second model may include a detection model (e.g., you-only-look-once (YOLO) neural network model). The one or more regions associated with the one or more static objects of the target scene in the target image may include one or more bounding boxes (e.g., rectangles 2, 3, or 4 in FIG. 9A) (e.g., a two-dimensional bounding box) of the one or more static objects determined by the detection model. In some embodiments, each of the one or more bounding boxes may correspond to a static object.

[0117] In some embodiments, the second model may include a segmentation model. The one or more regions associated with the one or more static objects of the target scene in the target image may include one or more connected components (e.g., regions 1030 and 1040 in FIG. 10B) of the one or more static objects determined by the segmentation model. In some embodiments, each of the one or more connected components may correspond to a type (e.g., a vehicle type (e.g., region 1030 in FIG. 10B), a person type (e.g., region 1040 in FIG. 10B)) of the one or more static objects. In some embodiments, a connected component of a static object determined based on the segmentation model may be a contour of the static object. A bounding box of a static object determined based on the detection model may be a maximum bounding rectangle of the static object. Therefore, the extraction of the static region based on the detection model may be less time-consuming than the extraction of the static region based on the segmentation model. An accuracy for extracting the static region based on the segmentation model may be higher than an accuracy for extracting the static region based on the

detection model.

[0118] In some embodiments, the second model (e.g., the detection model, the segmentation model) may be trained using a plurality of sample images of various sample traffic accidents, for example, a front collision between sample vehicles, a front-end collision between sample vehicles (e.g., as shown in FIG. 6A), a side collision between sample vehicles, a collision between a sample vehicle and a sample fixture (e.g., as shown in FIG. 6B), a side flip of a sample vehicle (e.g., 620 shown in FIG. 6C), a disintegration of a sample vehicle, or the like, or any combination thereof.

[0119] In some embodiments, each of the plurality of sample images may have a label as a known output (e.g., a ground truth). A label of a sample image may indicate a bounding box or a connected component of a static object in the sample image, and/or a type of the static object in the sample image.

[0120] In some embodiments, in the training process, the values of the second model's parameters may be determined or updated by performing an iteration of a backpropagation training procedure, e.g., a stochastic gradient descent backpropagation training technique. The training process may backpropagate the error (e.g., indicated by a loss function) determined for the output of the second model by comparing the output of the second model with the label (e.g., the ground truth). The training process may terminate until the error satisfies a condition (e.g., the error smaller than or equal to the threshold).

[0121] In some embodiments, the processing device 120 may determine the static region associated with the target scene in the target image based on the one or more regions (e.g., the one or more bounding boxes, the one or more connected components) associated with the one or more static objects of the target scene in the target image. In some embodiments, the processing device 120 may designate each of the one or more regions as one static region associated with the target scene in the target image. As used herein, a static region may refer to a minimum

region containing one of the one or more static objects.

[0122] In some embodiments, the processing device 120 may determine at least one of the one or more regions that satisfies a first condition. The first condition may relate to at least one distance between the one or more regions (e.g., one or more centers thereof). For example, the first condition may include that at least one distance between the at least one region is smaller than or equal to a distance threshold. As another example, the first condition may include that a ratio of a distance between two regions of the at least one region and a width of an object (e.g., a vehicle) inside one of the two regions is smaller than or equal to a ratio threshold (e.g., 1.2).

[0123] In some embodiments, the processing device 120 may generate a combination result by combining the at least one region. For example, the combination result may include a combined bounding box (e.g., rectangle 6 in FIG. 9B, rectangle *f* in FIG. 10C) contained the at least one region. In some embodiments, the processing device 120 may determine the static region associated with the target scene in the target image based on the combination result. In some embodiments, the processing device 120 may designate a region (e.g., a minimum region containing the combined bounding box) corresponding to the combined bounding box as one static region. In some embodiments, the processing device 120 may designate each region of a remaining portion of the one or more regions as one static region. In some embodiments, the processing device 120 may determine the combined bounding box based on a disjoint set union (DSU) structure.

[0124] In some embodiments, since a distance between a region of the remaining portion of the one or more regions and one of the at least one region may be larger than the distance threshold, the remaining portion of the one or more regions may be probably irrelevant with a traffic accident. For example, the remaining portion of the one or more regions may include road plants 712 in FIG. 7 which are irrelevant with a traffic accident. In some embodiments, the processing device 120 may exclude at

least a portion of the remaining portion of the one or more regions from the static region. In some embodiments, the processing device 120 may determine whether a distance between a region of the remaining portion of the one or more regions and one of the at least one region is larger than a second distance threshold. In response to determining that the distance between the region of the remaining portion of the one or more regions and one of the at least one region is larger than the second distance threshold, the processing device 120 may exclude the region from the static region, that is the processing device 120 may not designate the region as one static region; in response to determining that the distance is smaller than or equal to the second distance threshold, the processing device 120 may designate the region as one static region. In some embodiments, the second distance threshold may be larger than the distance threshold. In such cases, the blurred dynamic region and the at least a portion of the remaining portion may be unnecessary for further traffic accident detection, thereby not only reserving information that needs for the following traffic accident detection process but also reducing a computing load for the following traffic accident detection process.

[0125] Merely by way of example, the target image may include five static objects, i.e., a first object, a second object, a third object, a fourth object, and a fifth object, bounding boxes of which are a first bounding box, a second bounding box, a third bounding box, a fourth bounding box, and a fifth bounding box determined by the detection model. Assuming that a distance between the second bounding box and the third bounding boxes is smaller than or equal to the distance threshold, and a distance between the third bounding box and the fourth bounding box is smaller than or equal to the distance threshold, the processing device 120 may generate a sixth bounding box by combining the second bounding, the third bounding box, and the fourth bounding box. The target image may include a static region corresponding to the first bounding box, a static region corresponding to the fifth bounding box, and a static region corresponding to the sixth bounding box.

[0126] Take a target image in FIGs. 10A-10C as an example, the target image may include two static vehicles 1010 and a static person 1020. Region 1030 indicating a first connected component corresponding to the two static vehicles 1010 and region 1040 indicating a second connected component corresponding to the static person 1020 may be determined based on the segmentation model. Since the first connected component and the second connected component satisfy the first condition, a combined bounding box f may be determined by combining the first connected component and the second connected component. The processing device 120 may designate a region corresponding to the combined bounding box f as one static region.

[0127] In some embodiments, the processing device 120 may enlarge the bounding box (or the combined bounding box) by a preset ratio and designate a region corresponding to the enlarged bounding box (e.g., rectangle 6' in FIG. 9A) as the static region. In some embodiments, the ratio may be preset according to practical demands, for example, a type of the target scene, a size (e.g., a length, a width, an area, an aspect ratio) of the bounding box (or the combined bounding box), a size (e.g., a length, a width, an area, an aspect ratio) of an object inside the bounding box (or the combined bounding box). For example, the preset ratio may be 1.2. As another example, the preset ratio may be 0.5 of the width of a vehicle inside the bounding box.

[0128] In 540, the processing device 120 (e.g., the detection module 440) (e.g., the processing circuits of the processor 220) may determine a result (also referred to as a traffic accident result) indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image. In some embodiments, the traffic accident result may also include a confidence value associated with the traffic accident, a type of the traffic accident, etc. In some embodiments, the larger the confidence value is, the larger probability that the traffic accident exists in the static region may be. In

some embodiments, if the confidence value exceeds a value threshold (e.g., 0.5), the confidence value may be encoded as "1"; if the confidence value is smaller than or equal to the value threshold, the confidence value may be encoded as "0". As used herein, "1" may indicate that there is a traffic accident in the target image; "0" may indicate that there is no traffic accident in the target image.

[0129] In some embodiments, the type of the traffic accident may include a collision between vehicles, a collision between a vehicle and a preset fixture, a traffic accident of by a single vehicle, or the like, or any combination thereof. In some embodiments, the type of the traffic accident may include a front collision between vehicles, a front-end collision between vehicles (e.g., as shown in FIG. 6A), a side collision between vehicles, a collision between a vehicle and a preset fixture (e.g., as shown in FIG. 6B), a side flip of a vehicle (e.g., 620 shown in FIG. 6C), a disintegration of a vehicle, or the like, or any combination thereof. For example, the preset fixture may include a road facility such as a road barrier (e.g., 610 shown in FIG. 6B), a lamp rod, a marker rod, a building, a green plant, etc. In some embodiments, the side flip or the disintegration of the vehicle may be caused by a flat tire of the vehicle, aging of a mechanical component of the vehicle, malfunction of a mechanical component of the vehicle, etc.

[0130] In some embodiments, the processing device 120 may adjust a size of the target image according to the first model, and determine the traffic accident result based on the adjusted target image accordingly. For example, the processing device 120 may adjust the size of the target image to be the same as a size of an image that the first model can process. The process for adjusting a size of the target image according to the first model may be the same as the process for adjusting the size of the first image and/or the one or more second images, the descriptions of which are not repeated.

[0131] In some embodiments, the first model may be trained using a plurality of sample images of various sample traffic accidents, for example, a front collision

between sample vehicles, a front-end collision between sample vehicles (e.g., as shown in FIG. 6A), a side collision between sample vehicles, a collision between a sample vehicle and a sample fixture (e.g., as shown in FIG. 6B), a side flip of a sample vehicle (e.g., 620 shown in FIG. 6C), a disintegration of a sample vehicle, or the like, or any combination thereof. More descriptions of a structure of the first model may be found elsewhere in the present disclosure, for example, FIG. 12 and the descriptions thereof.

[0132] In some embodiments, the first model may be trained using the plurality of sample images each having a label as a known output (e.g., a ground truth). A label of a sample image may indicate a confidence value associated with the traffic accident of a static region in the sample image, and/or a type of the traffic accident of the sample image.

[0133] In some embodiments, in the training process, the values of the first model's parameters may be determined or updated by performing an iteration of a backpropagation training procedure, e.g., a stochastic gradient descent backpropagation training technique. The training process may backpropagate the error (e.g., indicated by a loss function) determined for the output of the first model by comparing the output of the first model with the label (e.g., the ground truth). The training process may terminate until the error satisfies a condition (e.g., the error smaller than or equal to the threshold).

[0134] In some embodiments, the processing device 120 may modify the confidence value associated with the traffic accident. In some embodiments, in response to determining that the confidence value fails to satisfy a second condition (e.g., larger than the value threshold (e.g., 0.5)), the processing device 120 may determine whether there are at least two of the one or more regions associated with the one or more static objects in the static region that overlap. In response to determining that there are at least two of the one or more regions associated with the one or more static objects in the static region that overlap, the processing device 120

may modify the confidence associated with the traffic accident; in response to determining that there are no regions associated with the one or more static objects in the static region that overlap, the processing device 120 may not modify the confidence value associated with the traffic accident and output a result indicating that there is no traffic accident happens in the static region. In response to determining that the confidence value satisfies the second condition, the processing device 120 may output a result indicating that a traffic accident happens in the static region.

[0135] In some embodiments, the one or more regions associated with the one or more static objects may be two-dimensional. The processing device 120 may determine one or more three-dimensional (3D) regions (e.g., rectangles 1110 and 1120 in FIG. 11) associated with the one or more static objects in the static region by performing a 3D transformation on the one or more 2D regions associated with the one or more static objects based on one or more coordinates associated with the one or more 2D regions (e.g., a coordinate of a vertex or a central point of a 2D region). In some embodiments, the processing device 120 may determine the 3D regions according to formula (2) below:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = K[R|t] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = p \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2)$$

where p refers to a projection transformation matrix, K refers to an intrinsic parameter (e.g., a focal length, a distortion parameter, etc.) of the image capture device, (x, y) refers to a coordinate of a point (e.g., a vertex, a central point) of the one or more regions in a 2D coordinate system (e.g., an image coordinate system), (X, Y, Z) refers to a coordinate of the point of the one or more regions in a 3D coordinate system (e.g., a world coordinate system), R refers to a rotation transformation vector between the 2D coordinate system and the 3D coordinate system, and t refers to a translation transformation vector between the 2D coordinate system and the 3D coordinate system. In some embodiments, the

processing device 120 may determine the projection transform matrix based on a process for determining a mapping relation between the 2D coordinate system and the 3D coordinate system, e.g., a least-squares technique, a Hough transformation technique, or a combination thereof.

[0136] In some embodiments, the processing device 120 may modify the confidence value associated with the traffic accident based on an overlap (also referred to as an overlapping region) between the one or more 3D regions associated with the one or more static objects in the static region. The processing device 120 may determine a ratio between a volume of the overlap between the one or more 3D regions and a volume of one of the one or more 3D regions (e.g., a minimum volume of the one or more 3D regions) and modify the confidence value based on the ratio. In some embodiments, the processing device 120 may determine the modified confidence value by performing a weighted summation on the ratio and the confidence value. For example, the processing device 120 may designate a value corresponding to the weighted summation result as the modified confidence value.

[0137] In some embodiments, in response to determining that the modified confidence value exceeds the value threshold, the processing device 120 may determine that the traffic accident happens in the static region. In some embodiments, in response to determining that the modified confidence value is smaller than or equal to the value threshold, the processing device 120 may determine there is no traffic accident in the static region.

[0138] In some embodiments, since the one or more regions associated with the one or more static objects determined by the second model may be two-dimensional, the confidence value may be determined by considering 2D information of the target scene, causing inaccurate traffic accident detection (e.g., failing to detect a slight collision between vehicles). Information of the target scene may be enriched (e.g., 3D information compared to the 2D information) by determining the one or more 3D

regions associated with the one or more static objects in the static region, such that the modified confidence value may be more accurate than the confidence value, thereby improving an accuracy for the traffic accident detection.

[0139] It should be noted a count of the traffic accident happens in the static region may be non-limiting, for example, 1, 2, 3, etc. For example, in response to determining that a confidence value (or a modified confidence value) that a rear-end collision happens in the static region exceeds the value threshold, the processing device 120 may determine that the rear-end collision happens in the static region. As another example, in response to determining that a confidence value (or a modified confidence value) that a collision between a vehicle and a preset fixture happens in the static region exceeds the value threshold, and a confidence value (or a modified confidence value) that a side flip of a vehicle happens in the static region exceeds the value threshold, the processing device 120 may determine that both the collision between the vehicle and the preset fixture and the side flip of the vehicle happen in the static region.

[0140] In some embodiments, the processing device 120 may transmit a notification of the traffic accident result to a third party (e.g., a transportation surveillance department). The third party may dispatch staff to process the traffic accident in real-time. In some embodiments, the processing device 120 may transmit a notification of the traffic accident result to a terminal of a user (e.g., a driver) associated with the traffic accident. The notification may be used to confirm whether the traffic accident happens, whether the user needs a rescue, etc. In some embodiments, the processing device 120 may transmit a notification of the traffic accident result to a monitor screen (e.g., implemented on the terminal 130). The monitor screen may present (e.g., display, broadcast, etc.) the notification of the traffic accident result in a form of text, image, video, voice, etc.

[0141] 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 or 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, one or more other optional operations (e.g., a storing operation) may be added elsewhere in the process 500. In the storing operation, the processing device 120 may store information and/or data (e.g., the traffic accident result) associated with the traffic accident detection in a storage device (e.g., the storage device 150, the ROM 230, the RAM 240, and/or the storage 390) disclosed elsewhere in the present disclosure.

[0142] FIG. 12 is a schematic diagram illustrating an exemplary first model according to some embodiments of the present disclosure.

[0143] As described in connection with FIG. 5, a first model may be configured to output a traffic accident result in a static region by processing the static region associated with a target scene in a target image. For example, the traffic accident result may include a result indicating whether a traffic accident happens in the static region, a confidence value associated with the traffic accident, a type of the traffic accident, etc.

[0144] As shown in FIG. 12, the first model may include a first structure including a convolutional layer and a nonlinear layer, a second structure including a max-pooling layer, a convolutional layer, and a nonlinear layer, two linear layers, and a softmax layer. A count of the first structure may be 3. A count of the second structure may be 2.

[0145] In some embodiments, the first model (e.g., the softmax layer thereof) may determine the confidence value associated with the traffic accident and encode the confidence value. If the confidence value exceeds a value threshold (e.g., 0.5), the confidence value may be encoded as "1"; if the confidence value is smaller than or equal to the value threshold, the confidence value may be encoded as "0". As shown in FIG. 12, the first model may output traffic accident results (e.g., 0, 1, 1, 1,

1, 0) of six images, where “1” indicates that there is a traffic accident in an image; “0” indicates that there is no traffic accident in an image.

[0146] FIG. 13 is a block diagram illustrating an exemplary accident detection system according to some embodiments of the present disclosure. The accident detection system 1300 may be an example of the accident detection system 100 in FIG. 1.

[0147] As shown in FIG. 13, the accident detection system 1300 may include an image capture device 1310 (e.g., the image capture device 110 in FIG. 1) and a detection device 1320 (e.g., the processing device 120 in FIG. 1). In some embodiments, the image capture device 1310 may be pre-mounted in a region where a traffic condition needs to be monitored. The image capture device 1310 may acquire video data (e.g., a video) of the region in real-time and transmit the video data to the detection device 1320 automatically or under a request of the detection device 1320. The detection device 1320 may determine whether a traffic accident happens in the region by analyzing the video data, for example, based on processes 500, 1400-2100.

[0148] FIG. 14 is a flowchart illustrating an exemplary process for determining a traffic accident result according to some embodiments of the present disclosure. In some embodiments, the process 1400 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 1400. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 1400 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 illustrated in FIG. 14 and described below is not intended to be limiting.

[0149] In 1410, the processing device 120 may determine a target image (e.g., the target image in FIG. 5) from a video.

[0150] In 1420, the processing device 120 may extract a static region (e.g., the static region in FIG. 5) of the target image.

[0151] In 1430, the processing device 120 may determine a traffic accident result by performing traffic accident detection on the static region using a traffic accident detection model (also referred to as a first model). More descriptions regarding the process 1400 may be found elsewhere in the present disclosure, for example, the process 500 and the descriptions thereof.

[0152] FIG. 15 is a flowchart illustrating an exemplary process for extracting a static region in a target image according to some embodiments of the present disclosure. In some embodiments, the process 1500 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 1500. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 1500 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 illustrated in FIG. 15 and described below is not intended to be limiting.

[0153] In 1510, the processing device 120 may determine a target image by blurring a dynamic region of a first image (e.g., the first image in FIG. 5) using one or more images (also referred to as one or more second images, e.g., the one or more second images in FIG. 5) captured before the first image. More descriptions regarding operation 1510 may be found elsewhere in the present disclosure, for example, operation 520 and the descriptions thereof.

[0154] In 1520, the processing device 120 may extract a static region of the target

image. More descriptions regarding operation 1520 may be found elsewhere in the present disclosure, for example, operation 530 and the descriptions thereof.

[0155] FIG. 16 is a flowchart illustrating an exemplary process for determining a static region in a target image according to some embodiments of the present disclosure. In some embodiments, the process 1600 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 1600. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 1600 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 illustrated in FIG. 16 and described below is not intended to be limiting.

[0156] In 1610, the processing device 120 may determine one or more bounding boxes corresponding to one or more static objects in a target image (e.g., the target image in FIG. 5) by performing object detection on the target image using a detection model (e.g., the detection model in FIG. 5).

[0157] In 1620, the processing device 120 may determine a static region of the target image based on the one or more bounding boxes corresponding to one or more static objects. More descriptions regarding the process 1600 may be found elsewhere in the present disclosure, for example, operation 530 and the descriptions thereof.

[0158] FIG. 17 is a flowchart illustrating an exemplary process for determining a static region in a target image according to some embodiments of the present disclosure. In some embodiments, the process 1700 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions,

and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 1700. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 1700 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 illustrated in FIG. 17 and described below is not intended to be limiting.

[0159] In 1710, the processing device 120 may determine a new bounding box (e.g., the combined bounding box in FIG. 17) by combining at least one of one or more bounding boxes (e.g., the one or more bounding boxes in FIG. 16) in a target image. At least one distance between the at least one bounding box may be smaller than or equal to a distance threshold.

[0160] In 1720, the processing device 120 may determine a static region of the target image based on the new bounding box. More descriptions regarding the process 1600 may be found elsewhere in the present disclosure, for example, operation 530 and the descriptions thereof.

[0161] FIG. 18 is a flowchart illustrating an exemplary process for determining a static region in a target image according to some embodiments of the present disclosure. In some embodiments, the process 1800 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 1800. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 1800 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 illustrated in FIG. 18 and described below is not intended to be limiting.

[0162] In 1810, the processing device 120 may determine one or more connected components corresponding to one or more static objects in a target image by performing object segmentation on the target image using a segmentation model.

[0163] In 1820, the processing device 120 may determine a static region of the target image based on the one or more connected components corresponding to the one or more static objects. More descriptions regarding the process 1800 may be found elsewhere in the present disclosure, for example, operation 530 and the descriptions thereof.

[0164] FIG. 19 is a flowchart illustrating an exemplary process for determining a static region in a target image according to some embodiments of the present disclosure. In some embodiments, the process 1900 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 1900. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 1900 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 illustrated in FIG. 19 and described below is not intended to be limiting.

[0165] In 1910, the processing device 120 may determine one or more bounding boxes corresponding to one or more connected components (e.g., the one or more connected components in FIG. 18) in a target image (e.g., the target image in FIG. 18).

[0166] In 1920, the processing device 120 may determine a new bounding box (e.g., the combined bounding box in FIG. 5) by combining at least one of the one or more bounding boxes in the target image that satisfies a condition (also referred to as a first condition). In some embodiments, the condition may relate to at least one

distance between the one or more bounding boxes, more descriptions of which may be found in operation 530.

[0167] In 1930, the processing device 120 may determine a static region of the target image based on the new bounding box. More descriptions regarding the process 1900 may be found elsewhere in the present disclosure, for example, operation 530 and the descriptions thereof.

[0168] FIG. 20 is a flowchart illustrating an exemplary process for modifying a confidence value associated with a traffic accident result according to some embodiments of the present disclosure. In some embodiments, the process 2000 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 2000. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 2000 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 illustrated in FIG. 20 and described below is not intended to be limiting.

[0169] In 2010, the processing device 120 may determine one or more 3D bounding boxes of one or more static objects by performing a 3D transformation on the one or more static objects. The one or more static objects may be partially overlapped.

[0170] In 2020, the processing device 120 may determine an overlapping region of the one or more 3D bounding boxes and determine a ratio between a volume of the overlapping region and a volume of one of the one or more 3D bounding boxes.

[0171] In 2030, the processing device 120 may modify a confidence value associated with a traffic accident result based on the ratio. More descriptions regarding the process 2000 may be found elsewhere in the present disclosure, for example, operation 540 and the descriptions thereof.

[0172] FIG. 21 is a flowchart illustrating an exemplary process for modifying a confidence value associated with a traffic accident result according to some embodiments of the present disclosure. In some embodiments, the process 2100 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform the process 2100. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 2100 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 illustrated in FIG. 21 and described below is not intended to be limiting.

[0173] In 2110, the processing device 120 may determine a weighted summation result by performing a weighted summation on a ratio (e.g., the ratio in FIG. 20) and a confidence value (e.g., the confidence value in FIG. 20).

[0174] In 2120, the processing device 120 may modify the confidence value based on the weighted summation result. In some embodiments, the processing device 120 may designate a value corresponding to the weighted summation result as the modified confidence value. More descriptions regarding the process 2100 may be found elsewhere in the present disclosure, for example, operation 540 and the descriptions thereof.

[0175] FIG. 22 is a block diagram illustrating an exemplary electronic device according to some embodiments of the present disclosure.

[0176] As shown in FIG. 21, the electronic device 2200 may include a processor 2210 and a storage 2220 operably coupled to the processor 2210. In some embodiments, the storage 2220 may be configured to store a program instruction for implementing any traffic accident detection process (e.g., the process 500, 1400-2100) described in the present disclosure. The processor 2210 may be configured

to execute the program instruction stored by the storage 2220 to implement any traffic accident detection process (e.g., the process 500, 1400-2100) described in the present disclosure. In some embodiments, the processor 2210 may also be referred to as a central processing unit (CPU). The processor 2210 may be an integrated circuit chip with signal processing capability. The processor 2210 may also be a general-purpose processor, a digital signal processor (DSP), a dedicated integrated circuit (ASIC), a field programmable gate array (FPGA), or other programmable logic devices, a discrete logic device, a transistor logic device, a discrete hardware component, or the like, or any combination thereof. The general-purpose processor may include a microprocessor or any general processor.

[0177] FIG. 23 is a block diagram illustrating an exemplary computer readable storage medium according to some embodiments of the present disclosure.

[0178] As shown in FIG. 23, the computer readable storage medium 2300 may store a program instruction 2310 used to execute any traffic accident detection process (e.g., the process 500, 1400-2100) described in the present disclosure. In some embodiments, the program instruction 2310 may form a program file and be stored in the computer readable storage medium 2300 in a form of a software product, such that a computer device (e.g., a personal computer, a server, a network device) or a processor may execute any traffic accident detection process (e.g., the process 500, 1400-2100) described in the present disclosure. In some embodiments, the computer readable storage medium 2300 may include a medium that can store a program instruction, for example, a universal serial bus (USB) flash drive, a mobile hard disk, a read-only memory (ROM), a random access memory (RAM), a magnetic disk, or a terminal device such as an optical disk, a computer, a server, a mobile phone, a tablet computer, etc.

[0179] 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.

[0180] 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 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” or “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.

[0181] 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 “unit,” “module,” 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.

[0182] 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 electro-magnetic, 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.

[0183] Computer program code for carrying out operations for aspects of the present disclosure may be written in any 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 2003, Perl, COBOL 2002, 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 (for example, through the Internet using an Internet Service Provider) or in a cloud computing environment or offered as a service such as a Software as a Service (SaaS).

[0184] 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.

[0185] 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 smaller than all features of a single foregoing disclosed embodiment.

WHAT IS CLAIMED IS:

1. A system, comprising:

at least one storage device including a set of instructions; and
at least one processor in communication with the at least one storage device,
wherein when executing the set of instructions, the at least one processor is
configured to cause the system to:

- obtain a video of a target scene that is captured by an image capture device;
- determine a target image associated with the target scene based on the video;
- determine a static region associated with the target scene in the target image; and
- determine a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image.

2. The system of claim 1, wherein to determine a target image associated with the target scene based on the video, the at least one processor is configured to cause the system to:

obtain, from the video, a first image;

obtain, from the video, one or more second images captured prior to the first image; and

determine the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images.

3. The system of claim 2, wherein to determine the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images, the at least one processor is configured to cause the

system to:

generate a fusion result by fusing the first image and the one or more second images; and

determine the target image based on the fusion result.

4. The system of claim 2 or claim 3, wherein a time interval between two adjacent images among the one or more second images and the first image relates to a type of the target scene.

5. The system of any one of claims 1-4, wherein to determine a static region associated with the target scene in the target image, the at least one processor is configured to cause the system to:

determine one or more regions associated with one or more static objects of the target scene in the target image by processing the target image based on a second model; and

determine the static region associated with the target scene in the target image based on the one or more regions associated with the one or more static objects of the target scene in the target image.

6. The system of claim 5, wherein

the second model includes a detection model; and

the one or more regions associated with the one or more static objects of the target scene in the target image includes one or more bounding boxes of the one or more static objects determined by the detection model.

7. The system of claim 5, wherein

the second model includes a segmentation model; and

the one or more regions associated with the one or more static objects of the

target scene in the target image includes one or more connected components of the one or more static objects determined by the segmentation model.

8. The system of any one of claims 5-7, wherein to determine the static region associated with the target scene in the target image based on the one or more regions associated with the one or more static objects of the target scene in the target image, the at least one processor is configured to cause the system to:

- determine at least one of the one or more regions that satisfies a first condition; and
- generate a combination result by combining the at least one region; and
- determine the static region associated with the target scene in the target image based on the combination result.

9. The system of claim 8, wherein the first condition relates to at least one distance between the one or more regions.

10. The system of any one of claims 5-9, wherein to determine a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image, the at least one processor is configured to cause the system to:

- determine a confidence value associated with the traffic accident by processing, based on the first model, the static region associated with the target scene in the target image; and
- in response to determining that the confidence value satisfies a second condition, determine that the traffic accident happens in the static region.

11. The system of claim 10, wherein to determine a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the

static region associated with the target scene in the target image, the at least one processor is configured to cause the system to:

in response to determining that the confidence value fails to satisfy the second condition, determine whether there are at least two of the one or more regions associated with the one or more static objects in the static region that overlap; and

in response to determining that there are the at least two of the one or more regions associated with the one or more static objects in the static region that overlap, modify the confidence value associated with the traffic accident.

12. The system of claim 11, wherein

the one or more regions associated with the one or more static objects including one or more two-dimensional regions; and

to modify the confidence value associated with the traffic accident, the at least one processor is configured to cause the system to:

determine one or more three-dimensional regions associated with the one or more static objects in the static region by performing a three-dimensional transformation on the one or more two-dimensional regions associated with the one or more static objects based on one or more coordinates associated with the one or more two-dimensional regions; and

modify the confidence value associated with the traffic accident based on an overlap between the one or more three-dimensional regions associated with the one or more static objects in the static region.

13. A method implemented on a computing device having at least one processor, at least one storage medium, and a communication platform connected to a network, the method comprising:

obtaining a video of a target scene that is captured by an image capture device;
determining a target image associated with the target scene based on the video;

determining a static region associated with the target scene in the target image;
and

determining a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image.

14. The method of claim 13, wherein the determining a target image associated with the target scene based on the video includes:

obtaining, from the video, a first image;

obtaining, from the video, one or more second images captured prior to the first image; and

determining the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images.

15. The method of claim 14, wherein the determining the target image by processing a dynamic region associated with the target scene in the first image based on the one or more second images includes:

generating a fusion result by fusing the first image and the one or more second images; and

determining the target image based on the fusion result.

16. The method of claim 14 or claim 15, wherein a time interval between two adjacent images among the one or more second images and the first image relates to a type of the target scene.

17. The method of any one of claims 13-16, wherein the determining a static region associated with the target scene in the target image includes:

determining one or more regions associated with one or more static objects of

the target scene in the target image by processing the target image based on a second model; and

determining the static region associated with the target scene in the target image based on the one or more regions associated with the one or more static objects of the target scene in the target image.

18. The method of claim 17, wherein

the second model includes a detection model; and

the one or more regions associated with the one or more static objects of the target scene in the target image includes one or more bounding boxes of the one or more static objects determined by the detection model.

19. The method of claim 17, wherein

the second model includes a segmentation model; and

the one or more regions associated with the one or more static objects of the target scene in the target image includes one or more connected components of the one or more static objects determined by the segmentation model.

20. The method of any one of claims 17-19, wherein the determining the static region associated with the target scene in the target image based on the one or more regions associated with the one or more static objects of the target scene in the target image includes:

determining at least one of the one or more regions that satisfies a first condition; and

generating a combination result by combining the at least one region; and

determining the static region associated with the target scene in the target image based on the combination result.

21. The method of claim 20, wherein the first condition relates to at least one distance between the one or more regions.

22. The method of any one of claims 17-21, wherein the determining a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image includes:

determining a confidence value associated with the traffic accident by processing, based on the first model, the static region associated with the target scene in the target image; and

in response to determining that the confidence value satisfies a second condition, determining that the traffic accident happens in the static region.

23. The method of claim 22, wherein the determining a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image includes:

in response to determining that the confidence value fails to satisfy the second condition, determining whether there are at least two of the one or more regions associated with the one or more static objects in the static region that overlap; and

in response to determining that there are the at least two of the one or more regions associated with the one or more static objects in the static region that overlap, modifying the confidence value associated with the traffic accident.

24. The method of claim 23, wherein

the one or more regions associated with the one or more static objects including one or more two-dimensional regions; and

the modifying the confidence value associated with the traffic accident includes:
determining one or more three-dimensional regions associated with the one or

more static objects in the static region by performing a three-dimensional transformation on the one or more two-dimensional regions associated with the one or more static objects based on one or more coordinates associated with the one or more two-dimensional regions; and

modifying the confidence value associated with the traffic accident based on an overlap between the one or more three-dimensional regions associated with the one or more static objects in the static region.

25. A non-transitory computer readable medium, comprising executable instructions that, when executed by at least one processor, directs the at least one processor to perform a method, the method comprising:

obtaining a video of a target scene that is captured by an image capture device;
determining a target image associated with the target scene based on the video;
determining a static region associated with the target scene in the target image;

and

determining a result indicating whether a traffic accident happens in the static region by processing, based on a first model, the static region associated with the target scene in the target image.