

USE OF HIGH-RESOLUTION IMAGES TO ENABLE ARTIFICIAL INTELLIGENCE IDENTIFICATION OF SHIPS

TECHNICAL FIELD

The presently disclosed subject matter relates to the field of marine navigation.

BACKGROUND

Some key concerns in the field of navigation include avoiding collision of a
5 waterborne craft or other platform with other waterborne craft, platform or other obstacles
and objects, particularly in cases of inattention of the human lookout.

Acknowledgement of the above references herein is not to be inferred as meaning
that these are in any way relevant to the patentability of the presently disclosed subject
matter.

10

GENERAL DESCRIPTION

The following are example embodiments of the presently disclosed subject matter.

According to a first aspect of the presently disclosed subject matter there is
presented a computerized method of object detection, the method performed by a
15 processing circuitry of an object detection system, the method comprising:

A) receive one or more images captured by at least one imaging device mounted on a
waterborne platform,

the one or more images comprising one or more objects, which are located within
a body of water;

20 B) perform the following processes:

i. perform a low-resolution detection process, comprising:

1. receive one or more reduced resolution images, generated by performing
an image resolution reduction, the image resolution reduction comprising
reducing the spatial resolution of the one or more images; and

- 2 -

2. perform object detection, on the one or more reduced resolution images, utilizing a neural network, thereby obtaining one or more detected objects detected at reduced resolution;
- ii. perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:
 1. receive a set of image tiles, derived from at least one image of the one or more images, wherein each image tile of the set corresponds to a sub-section of the at least one image, wherein a spatial resolution of each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and
 2. perform a tile-based detection of objects, on image tiles of the set, utilizing at least one image tiles neural network, thereby obtaining one or more tile-based detected objects; and
- iii. merge the one or more detected objects detected at reduced resolution, and the one or more tiles-based detected objects, thereby obtaining a final set of detected objects.

In addition to the above features, the method according to this aspect of the presently disclosed subject matter can include one or more of features (i) to (lxiv) listed below, in any desired combination or permutation which is technically possible:

- (i) the one or more images comprise one or more high resolution images, wherein the at least one imaging device comprises at least one high resolution imaging device.
- (ii) the method further comprising:
 - C) deriving the set of image tiles.
 - (iii) the method further comprising:
 - D) performing the image resolution reduction.
 - (iv) the method further comprising:
 - E) for one or more LIAS-defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network, thereby obtaining one or more local-inquiry-based detected objects,

- 3 -

where the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile,

where each LIAS-defined area comprises portions of at least two image tiles; and

F) merge the one or more local-inquiry-based detected objects with the image-tiles-based set of detected objects,

thereby facilitating obtaining an updated final set of detected objects.

(v) the local inquiry and segmentation comprise generating at least one additional image tile, corresponding to the one or more defined areas, and performing the local inquiry and segmentation within the at least one additional image tile.

(vi) the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in an overlap area of image tiles.

(vii) the image(s) captured in real time, wherein the system is configured to perform the method in one of: real time; near real time; and online,

thereby facilitating an avoidance of collision between the waterborne platform and an object of the one or more objects.

(viii) the system is configured to perform the method within 0.01 to 1 second.

(ix) the system is configured to perform the method within 1 to 60 seconds.

(x) the method is repeated at least one additional time.

(xi) the system is configured to synchronize inputs from the at least one imaging device and at least one of an Inertial Measurement Unit (IMU) and an Inertial Navigation System (INS).

(xii) the method further comprising:

G) performing tracking of the one or more objects, based on the repetition.

(xiii) the waterborne platform is a waterborne craft.

(xiv) the waterborne craft is one of a ship and a boat.

(xv) the one or more objects comprise at least one of: another waterborne platform, another waterborne craft, a buoy, an obstacle, and a lighthouse.

(xvi) the final set of detected objects comprise at least one of: another waterborne platform, another waterborne craft, a buoy, an obstacle, and a lighthouse.

- 4 -

- (xvii) the at least one imaging device is configured to view the one or more objects with a viewing resolution associated with a human observer.
- (xviii) the one or more captured images having a resolution of at least 640 pixels in at least one dimension.
- 5 (xix) the one or more images having a horizontal resolution of at least 640 pixels.
- (xx) the one or more images having a vertical resolution of at least 640 pixels.
- (xxi) the one or more captured images having a resolution of at least 1080 pixels in at least one dimension.
- 10 (xxii) the one or more images having a horizontal resolution of at least 1080 pixels.
- (xxiii) the one or more images having a vertical resolution of at least 1080 pixels.
- (xxiv) the one or more images having a horizontal resolution of at least 1280 pixels.
- 15 (xxv) the one or more images having a vertical resolution of at least 1280 pixels.
- (xxvi) the one or more images having a horizontal resolution of at least 4000 pixels.
- (xxvii) the one or more images having a vertical resolution of at least 4000 pixels.
- (xxviii) the one or more images having a horizontal resolution of at least 6000 pixels.
- 20 (xxix) the one or more images having a vertical resolution of at least 6000 pixels.
- (xxx) the image resolution reduction comprises reducing at least one of a horizontal resolution and vertical resolution, of the one or more captured images.
- 25 (xxxi) at least some adjacent image tiles of the set of image tiles having at least partial horizontal overlap and/or at least partial vertical overlap.
- (xxxii) at least some adjacent image tiles of the set of image tiles having at least partial horizontal overlap.
- (xxxiii) at least some adjacent image tiles of the set having at least partial vertical overlap.
- 30 (xxxiv) at least some adjacent image tiles of the set of image tiles touch each other.
- (xxxv) at least one of the following is true:

- 5 -

- [a] at least some adjacent image tiles of the set of image tiles have at least partial horizontal overlap and/or at least partial vertical overlap; and
(xxxvi)[b] at least some adjacent image tiles of the set of image tiles touch each other. at least some adjacent image tiles of the set of image tiles have no
5 horizontal overlap and no vertical overlap.
- (xxxvii) each image tile comprising fewer horizontal pixels than a number of horizontal pixels of the at least one image.
- (xxxviii) each image tile comprising fewer vertical pixels than a number of vertical pixels of the at least one image.
- 10 (xxxix) the set of images tiles corresponds to all pixels of the one or more captured images.
- (xl) the low-resolution detection process is performed prior to the high-resolution detection process.
- (xli) the low-resolution detection process is performed in parallel to the high-resolution detection process.
15
- (xlii) the low-resolution detection process is performed after the high-resolution detection process.
- (xliii) the low-resolution detection process is performed prior to the LIAS-based detection process.
20
- (xliv) the low-resolution detection process is performed in parallel to the LIAS-based detection process.
- (xlv) the low-resolution detection process is performed after the LIAS-based detection process.
- 25 (xlvi) the neural network and the at least one images tiles neural network are of a same size.
- (xlvii) the neural network and the at least one image tiles neural network are the same.
- (xlviii) the at least one image tiles neural network and the at least one local-inquiry
30 neural network are of a same size.
- (xlix) the at least one image tiles neural network and the at least one local-inquiry neural network are the same.

- 6 -

- (l) the neural network and the at least one local-inquiry neural network are of a same size.
- (li) the neural network and the at least one local-inquiry neural network are the same.
- 5 (lii) the method further comprising:
 - H) classify the at least one detected object of the final set.
 - (liii) the method further comprising:
 - I) identify the at least one detected object of the final set.
 - (liv) the at least one imaging device is configured to cover an angle up to 360
 - 10 degrees around the waterborne platform.
 - (lv) the at least one imaging device is configured to cover at least 180 degrees around the waterborne platform.
 - (lvi) the at least one imaging device is configured to cover at least 270 degrees around the waterborne platform.
 - 15 (lvii) the at least one imaging device is configured to cover 360 degrees around the waterborne platform.
 - (lviii) the at least one imaging device is comprised in at least one payload, where each payload comprising one or more imaging devices, where the at least one payload is configured to cover an angle up to 360 degrees around the
 - 20 waterborne platform.
 - (lix) the at least one imaging device is comprised in a plurality of payloads, wherein each payload comprising a plurality of imaging devices.
 - (lx) the at least one imaging device is a camera.
 - (lxi) the object detection system comprising the at least one imaging device.
 - 25 (lxii) the method further comprising:
 - J) merge the final set of detected objects with an additional final set of detected objects associated with image capture in a different field of view (FOV) of the at least one imaging device.
 - (lxiii) the method further comprising:
 - 30 K) outputting information indicative of the final set of detected objects.
 - (lxiv) the outputted information comprises position information associated with at least one detected object of the final set.

- 7 -

According to a second aspect of the presently disclosed subject matter there is presented a computerized method of object detection, the method performed by a processing circuitry of an object detection system, the method comprising:

5 A) receive one or more captured images captured by at least one imaging device mounted on a waterborne platform,

the one or more images comprising one or more objects, which are located within a body of water;

B) perform the following processes:

- 10 i. performing a low-resolution detection process, comprising:
 1. receive one or more reduced resolution images, generated by performing an image resolution reduction, comprising reducing the spatial resolution of the one or more images; and
 2. performing object detection, on the one or more reduced resolution images, utilizing a neural network, thereby obtaining zero or more
15 detected objects detected at reduced spatial resolution;
- ii. perform a high-resolution detection process by splitting the one or more images to image tiles, comprising:
 1. receive a set of image tiles, derived from at least one image of the one or more images,
20 wherein each image tile of the set corresponds a sub-section of the at least one image,
wherein an image tile spatial resolution of each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and
 2. perform a tile-based detection of objects, on image tiles of the set, utilizing
25 at least one image tiles neural network, thereby obtaining one or more tile-based detected objects;
- iii. merge the zero or more detected objects detected at reduced spatial resolution, and the one or more tiles-based detected objects, thereby obtaining a final set of detected objects.

30 According to a third aspect of the presently disclosed subject matter there is presented a computerized method of object detection, the method performed by a processing circuitry of an object detection system, the method comprising:

- 8 -

A) receive one or more images captured by at least one imaging device mounted on a waterborne platform,

the one or more images comprising one or more objects, which are located within a body of water;

5 B) perform the following processes:

i) perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:

(1) receive a set of image tiles, derived from at least one image of the one or more images,

10 where each image tile of the set corresponds to a sub-section of the at least one image,

where a spatial resolution of each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and

(2) perform a tile-based detection of objects, on image tiles of the set, utilizing
15 at least one image tiles neural network, thereby obtaining one or more tile-based detected objects;

ii) for one or more LIAS-defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network,
20 thereby obtaining one or more local-inquiry-based detected objects,

where the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile,

where each defined area comprises portions of at least two image tiles; and

25 iii) merge the one or more local-inquiry-based detected objects with the image-tiles-based set of detected objects,

thereby obtaining a final set of detected objects.

According to a fourth aspect of the presently disclosed subject matter there is presented a computerized method of object detection, the method performed by a processing circuitry of an object detection system, the method comprising:

30 A) receive one or more images captured by at least one imaging device mounted on a waterborne platform,

the one or more images comprising one or more objects, which are located within a body of water;

B) perform the following processes:

- i. perform a low-resolution detection process, comprising:
 - 5 1. receive one or more reduced resolution images, generated by performing an image resolution reduction, the image resolution reduction comprising reducing the spatial resolution of the one or more images; and
 2. perform object detection, on the one or more reduced resolution images, utilizing a neural network, thereby obtaining one or more detected objects
10 detected at reduced resolution;
- ii. perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:
 - (1) receive a set of image tiles, derived from at least one image of the one or more images,
15 wherein each image tile of the set corresponds to a sub-section of the at least one image,
wherein a spatial resolution of each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and
 - (2) perform a tile-based detection of objects, on image tiles of the set, utilizing
20 at least one image tiles neural network, thereby obtaining one or more tile-based detected objects;
- iii. for one or more LIAS-defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural
25 network, thereby obtaining one or more local-inquiry-based detected objects,
where the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile,
where each LIAS-defined area comprises portions of at least two image tiles;
30 and

- 10 -

- iv. merge the one or more detected objects detected at reduced resolution, the image-tiles-based set of detected objects, and the one or more local-inquiry-based detected objects, thereby obtaining a final set of detected objects.

5 In addition to the above features, the method according to the second to fourth aspects of the presently disclosed subject matter can include one or more of features (i) to (lxiv) listed above, in any desired combination or permutation which is technically possible.

According to a fifth aspect of the presently disclosed subject matter there is
10 presented a computerized object detection system, comprising a processing circuitry, configured to perform the method of any one of the first to fourth aspects of the presently disclosed subject matter.

According to a sixth aspect of the presently disclosed subject matter there is
presented a non-transitory computer readable storage medium tangibly embodying a
15 program of instructions that, when executed by a processing circuitry of an object detection system, cause the processing circuitry to perform to perform the method of any one of the first to fourth aspects of the presently disclosed subject matter.

The fifth and sixth aspects of the disclosed subject matter can optionally include one or more of features (i) to (lxiv) listed above, mutatis mutandis, in any desired
20 combination or permutation which is technically possible.

BRIEF DESCRIPTION OF THE DRAWINGS

In order to understand the invention and to see how it can be carried out in practice, embodiments will be described, by way of non-limiting examples, with reference to the
25 accompanying drawings, in which:

Fig. 1 schematically illustrates an example generalized view of ship navigation, in accordance with some embodiments of the presently disclosed subject matter;

Fig. 2 schematically illustrates an example generalized schematic diagram of a waterborne platform, in accordance with some embodiments of the presently disclosed subject matter;

Figs. 3A and 3B schematically illustrate an example generalized schematic diagram of a computerized object detection system, in accordance with some
5 embodiments of the presently disclosed subject matter;

Fig. 4 schematically illustrates an example generalized an example generalized representation of an image, in accordance with some embodiments of the presently disclosed subject matter;

Fig. 5 schematically illustrates an example generalized example generalized representation of a reduced resolution image, in accordance with some embodiments of
10 the presently disclosed subject matter;

Figs. 6A and 6B schematically illustrate an example generalized representation of image tiles, in accordance with some embodiments of the presently disclosed subject
15 matter;

Fig. 7 schematically illustrates an example generalized representation of a data flow, in accordance with some embodiments of the presently disclosed subject matter;
and

Figs. 8A-8C schematically illustrate a generalized flow chart diagram, of a flow
20 of a process or method, for object detection, in accordance with some embodiments of the presently disclosed subject matter.

DETAILED DESCRIPTION

In the drawings and descriptions set forth, identical reference numerals indicate
25 those components that are common to different embodiments or configurations.

In the following detailed description, numerous specific details are set forth in order to provide a thorough understanding of the invention. However, it will be understood by those skilled in the art that the presently disclosed subject matter may be practiced without these specific details. In other instances, well-known methods,
5 procedures, components and circuits have not been described in detail so as not to obscure the presently disclosed subject matter.

It is to be understood that the invention is not limited in its application to the details set forth in the description contained herein or illustrated in the drawings. The invention is capable of other embodiments and of being practiced and carried out in
10 various ways. Hence, it is to be understood that the phraseology and terminology employed herein are for the purpose of description and should not be regarded as limiting. As such, those skilled in the art will appreciate that the conception upon which this disclosure is based may readily be utilized as a basis for designing other structures, methods, and systems for carrying out the several purposes of the presently disclosed
15 subject matter.

It will also be understood that the system according to the invention may be, at least partly, implemented on a suitably programmed computer. Likewise, the invention contemplates a computer program being readable by a computer for executing the method of the invention. The invention further contemplates a non-transitory computer-readable
20 memory tangibly embodying a program of instructions executable by the computer for executing the method of the invention.

Those skilled in the art will readily appreciate that various modifications and changes can be applied to the embodiments of the invention as hereinbefore described without departing from its scope, defined in and by the appended claims.

25 Unless specifically stated otherwise, as apparent from the following discussions, it is appreciated that throughout the specification discussions utilizing terms such as "providing", "receiving", "performing", "deriving", "merging", "selecting", "outputting", "classifying", "identifying", "computing", "re-computing", "calculating", "determining", "generating", or the like, refer to the action(s) and/or process(es) of a computer that
30 manipulate and/or transform data into other data, said data represented as physical, e.g. such as electronic or mechanical quantities, and/or said data representing the physical objects. The term "computer" should be expansively construed to cover any kind of

hardware-based electronic device with data processing capabilities including a personal computer, a server, a computing system, a communication device, a processor or processing unit (e.g. digital signal processor (DSP), a microcontroller, a microprocessor, a field programmable gate array (FPGA), an application specific integrated circuit (ASIC), etc.), and any other electronic computing device, including, by way of non-limiting example, computerized systems or devices **300** and processing circuitries such as e.g. **320** disclosed in the present application.

The operations in accordance with the teachings herein may be performed by a computer specially constructed for the desired purposes, or by a general-purpose computer specially configured for the desired purpose by a computer program stored in a non-transitory computer-readable storage medium.

Embodiments of the presently disclosed subject matter are not described with reference to any particular programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the presently disclosed subject matter as described herein.

The terms "non-transitory memory" and "non-transitory storage medium" used herein should be expansively construed to cover any volatile or non-volatile computer memory suitable to the presently disclosed subject matter.

As used herein, the phrase "for example," "such as", "for instance" and variants thereof describe non-limiting embodiments of the presently disclosed subject matter. Reference in the specification to "one case", "some cases", "other cases", "one example", "some examples", "other examples", or variants thereof, means that a particular described method, procedure, component, structure, feature or characteristic described in connection with the embodiment(s) is included in at least one embodiment of the presently disclosed subject matter, but not necessarily in all embodiments. The appearance of the same term does not necessarily refer to the same embodiment(s) or example(s).

Usage of conditional language, such as "may", "might", or variants thereof, should be construed as conveying that one or more examples of the subject matter may include, while one or more other examples of the subject matter may not necessarily include, certain methods, procedures, components and features. Thus, such conditional language is not generally intended to imply that a particular described method, procedure,

component or circuit is necessarily included in all examples of the subject matter. Moreover, the usage of non-conditional language does not necessarily imply that a particular described method, procedure, component or circuit is necessarily included in all examples of the subject matter.

5 It is appreciated that certain embodiments, methods, procedures, components or features of the presently disclosed subject matter, which are, for clarity, described in the context of separate embodiments or examples, may also be provided in combination in a single embodiment or examples. Conversely, various embodiments, methods, procedures, components or features of the presently disclosed subject matter, which are, for brevity,
10 described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination.

 It should also be noted that each of the figures herein, and the text discussion of each figure, describe one aspect of the presently disclosed subject matter in an informative manner only, by way of non-limiting example, for clarity of explanation only. It will be
15 understood that the teachings of the presently disclosed subject matter are not bound by what is described with reference to any of the figures or described in other documents referenced in this application.

 Bearing this in mind, attention is drawn to **Fig. 1**, schematically illustrating an example generalized view of ship navigation, in accordance with some embodiments of
20 the presently disclosed subject matter. View **100** depicts a conceptual and schematic example view of ship navigation. In some examples the waterborne platform **110** is a waterborne craft, e.g. one of a ship and a boat. Ship, boat or other waterborne craft **110** moves/sails **108** in a body of water (not shown explicitly), in a particular direction **108**. The ship **110** is turning **117** to the left (port). The waterborne craft **110** is referred to herein
25 also as a watercraft. In some examples, it is a marine craft **110**. In much of the disclosure herein, the waterborne platform **110** will be exemplified by the specific example of a waterborne craft. In other non-limiting examples, waterborne platform **110** is a drilling platform, e.g. an offshore oil drilling rig, or other sea-based platform.

 Aboard the craft are one or more human observers **115**, e.g. a lookout or other
30 member of the crew of the ship **110**, who is looking out for other platforms (e.g. drilling platforms), other ships, boats, obstacles and other objects on the body of water. Several example such objects are shown, for illustrative purposes only. In some examples the

- 15 -

observed object(s) comprises at least one of: another platform, e.g. another waterborne craft, a buoy, an obstacle, and a lighthouse.

Shown are a buoy **125**, which is relatively far from craft **110**, and a floating obstacle **127** located to the right of the craft **110**. Also shown are a relatively large ship **120**, and a smaller boat **123**, both positioned roughly ahead of the craft **110**. Also shown are a boat **140** to the rear left of the craft **110**, which is turning **144** such that it will soon be to the left of craft **110**. To the left of craft **110** is a relatively small boat **130**, which visually blocks part of a more distant large object **133**, e.g. an aircraft carrier, an ocean liner or a large tanker.

Examples of obstacle **127** are a floating shipping container which fell off of a container ship, a floating tree, and an iceberg. In some examples, the detected object is a lighthouse jutting into the body of water.

In some examples, there can be at least certain technical advantages to deploying an automated computerized object detection system **300**, which detects images captured in imaging devices such as cameras. Such a system can replace a trained human observer(s) **115**, e.g. lookout(s) **115**, or in other cases can provide assistance to the human lookout. For example, it can detect surrounding objects in cases where the human lookout is inattentive. As another example, it can detect in real time a relatively large number of objects, where a human may have problems detecting them all in a timely manner. In still other examples, the system can replace the human, thus saving staffing costs, or covering shifts where no human lookout is available. This detection can assist the vessel **110** when navigating, and/or can facilitate avoidance of collisions.

In some non-limiting examples, the imaging device(s) **210** is configured to capture an object of size 3-5 meters located at a distance of 3 kilometers (km). Such a requirement can drive the resolution of the imaging devices used.

In some examples, as disclosed further herein with reference to **Fig. 2**, the imaging device(s) is configured to capture a wide-angle field of view that is not merely ahead of the ship, e.g. the devices see also to the sides, and even perhaps to the rear – in some examples with a 360 degree field of view. This can provide example advantages in e.g. collision avoidance. As one example, if the ship **110** turns **117** hard to the left, it might collide with the other ship **133**, if the other ship is not seen. As another, If the boat **140**, currently behind ship **110**, turns **144** towards the left side of ship **110**, it might cause a

collision with ship **110** as **110** makes its turn **117**. Note also, that the objects/ships are not shown to any scale.

At least to address such technical problems and provide such advantages, there is disclosed herein a computerized method of object detection, as well as a computerized
5 object detection system **300** and software products to perform such a method. The method comprises, in some examples, the following:

A) receive one or more images **400**. In some examples, the images have a high resolution, e.g. a spatial resolution of at least 640 pixels in at least one dimension. These captured image(s) **400** comprise one or more objects **645**, **657**, which are
10 located within a body of water. The image(s) are captured by at least one imaging device **210** mounted on the waterborne platform **110**;

B) perform the following processes:

- i) perform a low-resolution detection process. This first process comprises:
 - 1. receive one or more reduced resolution images **500**, which are generated
15 by performing an image resolution reduction. The image resolution reduction comprise reducing the spatial resolution (e.g. $M \times N$) of the captured image(s) to a reduced resolution (e.g. $m \times n$); and
 - 2. perform object detection, on the reduced resolution image(s) **500**, utilizing a neural network **327**, thereby obtaining one or more detected objects **557**.
20 These objects **557** are detected at reduced resolution;
- ii) perform a high-resolution detection process, by splitting the captured image(s) **400** to image tiles **601**. This second process comprises:
 - 1. receive a set **600** of image tiles **601**, derived from at least one image
25 **400**. Each image tile of the set corresponds to a sub-section of the captured image **400**. The image tile spatial resolution (e.g. $M \times N$) of the image tile(s) is higher than the reduced spatial resolution (e.g. $m \times n$) **601** of the reduced resolution image(s) **500**; and
 - 2. performing a tile-based detection of objects, on image tiles **607**, **601** of the set **600**, utilizing at least one image tiles neural network **327**, thereby
30 obtaining one or more tile-based detected objects **680**, **670**; and

- iii) merging the one or more detected objects **557**, which are detected at the reduced spatial resolution, and the one or more tiles-based detected objects **680**, thereby obtaining a final set of detected objects.

Note that in some examples zero objects were detected, in either process.

- 5 In some examples the method further comprises deriving the set **600** of image tiles, and/or performing the image resolution reduction.

In some examples, the method further comprises (C) outputting information indicative of the final set of detected objects, for example position information associated with at least one detected object **680** of the final set.

- 10 In some examples, the method further comprises:

D) for one or more defined areas **612** associated with the set **600** of image tiles, perform a Local Inquiry and Segmentation (LIAS), at a spatial resolution higher than the reduced resolution ($m \times n$), utilizing at least one local-inquiry neural network **327**

These defined areas are referred to herein also as LIAS-defined areas.

- 15 The defined area(s) is indicative of detection of the tiles-based detected object(s) **670** in more than one image tile **601**, **602**. Each defined area **618** comprises portions of at least two image tiles **604**, **605**. In some implementation, the defined area(s) is indicative of detection of tiles-based detected object(s) **690** in an overlap area **663** of image tiles **604**, **602**, **607**, **608**. In some implementation, the defined area(s) is indicative of detection
20 of tiles-based detected object(s) **690** in a touching area of image tiles **605'**, **606'**. This step thereby obtains an updated final set of detected objects, in some implementations. The LIAS method is referred to herein also a local search. In some implementations, the local search comprises generating at least one additional image tile **614**, corresponding to the LIAS-defined area(s) **612**, and performing the local search within the at least one
25 additional image tile.

Also disclosed herein is another computerized method of object detection, as well as a computerized object detection system **300** and software products to perform such a method. This other method comprises, in some examples, the following:

- A) receive one or more images captured by at least one imaging device mounted on a
30 waterborne platform **110**. The one or more images comprising one or more objects, which are located within a body of water;
B) perform the following processes:

iv) perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:

(1) receive a set of image tiles, derived from at least one image of the one or more images,

5 where each image tile of the set corresponds to a sub-section of the at least one image,

 where a spatial resolution of each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and

 (2) perform a tile-based detection of objects, on image tiles of the set, utilizing
10 at least one image tiles neural network, thereby obtaining tile-based detected object(s);

 v) for one or more LIAS-defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network,
15 thereby obtaining one or more local-inquiry-based detected objects,

 where the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile,

 where each LIAS-defined area comprises portions of at least two image tiles; and

 vi) merge the one or more local-inquiry-based detected objects with the
20 image-tiles-based set of detected objects, thereby obtaining a final set of detected objects.

 In some implementations, the method is repeated one or more additional times. This can facilitate an implementation in which the method further comprises performing tracking of the object(s) **120, 123** based on the repetition.

25 Various figures illustrate the concepts of the methods disclosed herein. **Figs. 2-3** provide schematic diagrams of systems configured for performing the method of the presently disclosed subject matter. **Figs. 4-6** disclose illustrative examples of various types of images, **Fig. 7** discloses an example data flow associated with the method. **Figs. 8A-8C** disclose an example flow chart of a method for cross-validation using a coreset
30 tree.

 Additional advantages of the presently disclosed subject matter are disclosed further herein, with reference to one or more of the above-mentioned figures.

Attention is now drawn to **Fig. 2**, schematically illustrating an example generalized schematic diagram **200** of a waterborne platform **110**, e.g. a waterborne craft, in accordance with some embodiments of the presently disclosed subject matter. The craft in some examples comprises one or more Inertial Measurement Units (IMU) **242**, which
5 in some implementations is comprised in one or more Inertial Navigation Systems (INS) **240**. In some examples, the **110 110** comprises navigation system **243**. In some examples, the **110 110** comprises collision avoidance system **245**. In other examples, not shown in the figure, collision avoidance system **245** is comprised within navigation system **243**. Examples of such systems are known in the art.

10 In some examples, **110 110** comprises object detection system(s) **300**. More detail on examples of system **300** is disclosed further herein, with reference to **Figs. 3A** and **3B**.

In some examples, platform **110** comprises one or more imaging devices **210**, **224**, e.g. a camera. In some examples, there are a plurality of imaging devices. The figure shows references for only some of the devices shown in the figure, for ease of exposition.
15 In the non-limiting example of the figure, the craft comprises twelve (12) such imaging devices, where four payloads **220** each comprise three (3) imaging devices **210**.

Thus, in some implementations, the system comprises a plurality of payloads, and each payload comprises a plurality of imaging devices, each with its field of view. Several such payloads can be combined, to view, in combination, up to 360 degrees. In some
20 implementations, the cameras or other imaging devices sit on the same plane, e.g. the same horizontal plane.

In some implementations, the system comprises multiple payloads, each of which comprises one or more imaging devices.

In some implementations, the system comprises one payload, which comprises
25 more than one imaging device.

Thus, the at least one imaging device is comprised in at least one payload, where each payload comprising one or more imaging devices.

In some implementations, the imaging device(s) is configured, in aggregation across all of the devices, to cover an angle up to 360 degrees around the waterborne platform. Thus, in some implementations, the one or more payloads are configured, in
30 aggregation across all of the payloads, and devices, to cover an angle up to 360 degrees around the waterborne platform. In the particular example of the figure, the set of imaging

devices is configured to capture images which cover 360 degrees around the waterborne platform. For example, each of the 12 devices has an angular field of view of e.g. 38 degrees, and there is overlapping between their fields of view.

In other examples, the platform is configured with fewer cameras, e.g. in fewer
5 payloads, and the imaging device(s) sees a smaller overall field of view. Thus, in some example implementations, the set of imaging devices is configured to cover at least 180 degrees around the waterborne craft **110**. For example, if the craft comprises the three-device payload **222**, facing forward, along with two additional cameras **226** and **228**, the craft can detect objects in a field of view of over 180 degrees, centered around the
10 direction **108** of sailing. In still other example implementations, the set of imaging devices is configured to cover at least 270 degrees around the waterborne craft. Other possible configurations are possible.

In some examples, some, or all, of the imaging device fields of view (FOV) **A**, **B** of two adjacent imaging devices overlap, as shown. In other some other examples, some
15 or all of the adjacent fields of view **B**, **C** do not overlap. Thus, lines **280** and **284** show borders of fields of view which do not overlap, nor touch, and there are gaps in the field of view. Thus, in the figure it is seen that there is a gap between lines **280** and **284**.

In still other examples, some or all of the adjacent fields of view do not overlap, but rather touch. That is, the adjacent borders **280**, **284** of the fields of view touch, and
20 thus lines **280** and **284** are the same. This implementation is not shown in the figure, for simplicity of exposition.

Note also, that instead of covering a large field of view using multiple cameras, the craft can be configured with movable imaging device(s) **230**, e.g. one or more gimbaled cameras, instead of one or more of the fixed cameras **220**. The use of such a
25 movable camera can facilitate an increased total system field of view, as compared to the field of view of one individual imaging device. The dashed nature of gimbaled imaging device **230** indicates that it is an alternative option to the implementation shown in the figure as solid lines. In other non-limiting examples, a pan-tilt camera **230** or a pan-tilt-roll camera **230** is utilized.

30 Recall, as disclosed with reference to **Fig. 1**, that imaging a field of view around much or most of the platform **110** can provide at least certain advantages when attempting to avoid collision with e.g. boats that are on the side or the rear of ship **110**.

In some examples, the craft comprises communications network **247**, configured to communicate between the various components of platform **110**, e.g. systems **300**, **240**, **242**, **243**, **245**, **247** and the various imaging devices **210**, **226**, **227**, **224**, **228**.

In some examples, the detection system **300** is configured to synchronize inputs
5 from the imaging device(s) **210** with those of IMU **242** and/or INS **240**. Such synchronization can match a particular captured image, at a particular time *t*, with the IMU/INS data associated with that time *t* (or e.g. the closest time to *t*). This can provide advantageous, for example, when determining the craft's **110** orientation at the time of capture *t*.

10 For simplicity of exposition, the presently disclosed subject matter is with reference to a surface watercraft **110**. In another implementation, the watercraft is a submarine. In other example implementations, the imaging devices **210** are instead installed in airborne craft, e.g. an airplane, quadcopter, drone, or a balloon. In other example implementations, the imaging devices are instead installed on land, e.g. on shore
15 – that is with an optical view of the body of water. For example, the devices are mounted on a lighthouse or a tower.

Attention is now drawn to **Figs. 3A** and **3B**, schematically illustrating an example generalized schematic diagram **300** of a computerized object detection system **300**, in accordance with some embodiments of the presently disclosed subject matter. In some
20 non-limiting examples, computerized system **300** includes a computer. It may, by way of non-limiting example, comprise a processing circuitry **320**. This processing circuitry may comprise a processor **830** and a memory **325**.

This processing circuitry **320** may be, in non-limiting examples, general-purpose computer(s) specially configured for the desired purpose by a computer program stored
25 in a non-transitory computer-readable storage medium. They may be configured to execute several functional modules in accordance with computer-readable instructions. In other non-limiting examples, this processing circuitry **320** may be a computer(s) specially constructed for the desired purposes.

In some implementations, the processing circuitry **320** also comprises one or more
30 graphics processing units (GPU) **323**. In some examples, the GPU comprises one or more neural networks (NN) **327**. Various types of NN can be utilized, e.g. Deep Neural Network(s) (DNN) or other known *per se* networks. In some examples, there are three

different neural networks, or three different groups/sets of neural networks **327**, each configured for particular functionality. For example, there can be separate neural network(s) for each of:

- I. performing detection of objects, on one or more reduced resolution images **500**, at a reduced spatial resolution, as part of a low-resolution detection process. This neural network(s) is referred to herein, in some examples, also as a reduced-resolution, or low-resolution, neural network(s). More on this is disclosed with reference to **Figs. 5, 7 and 8**.
- II. performing a tile-based detection of objects, on image tiles, thereby obtaining one or more tile-based detected objects. This neural network(s) is referred to herein, in some examples, as an image tiles neural network(s). More on this is disclosed with reference to **Figs. 6, 7 and 8**.
- III. performing Local Inquiry and Segmentation (LIAS), at a spatial resolution higher than the reduced resolution. In some cases, this yields one or more local-inquiry-based detected objects, e.g. a local-inquiry-based set of detected objects. In some cases, this yields an updated final set of detected objects. This neural network(s) is referred to herein, in some examples, as a local-inquiry neural network(s). More on this is disclosed with reference to **Figs. 6, 7 and 8**.

In some example implementations, the reduced-resolution neural network(s) and the images tiles neural network(s) are of the same size. In some implementations, the reduced-resolution neural network(s) and the local-inquiry neural network(s) are of the same size. In some implementations, the images tiles neural network(s) and the local-inquiry neural network(s) are of the same size. In some implementations, the reduced-resolution neural network(s) and the images tiles neural network(s) are the same neural network(s). In some implementations, the reduced-resolution neural network(s) and the local-inquiry neural network(s) are the same neural network(s). In some implementations, the images tiles neural network(s) and the local-inquiry neural network(s) are the same neural network(s). Note that in some cases the sizing and sharing of NNs are engineering decisions, to optimize system resources and the time to perform the method.

Thus, for example, using the same neural network for both the reduced-resolution detection and image tiles detection can decrease the use of computer resources, but can

require performing certain tasks sequentially rather than in parallel, thus requiring more time. Note also, that also within a particular function of the process, the number of neural networks utilized is an engineering decision. Thus, for example, if fifteen image tiles are used in **Fig. 6**, fifteen NN's will allow a more fully parallel processing in image-tiles-based detection, but at the cost of more resources - as compared to running a single neural network fifteen times. Similarly, a number of NN more than one and less than fifteen can be utilized.

In other examples, other machine learning techniques, e.g. decision trees or random forests, are used in place of neural networks to perform the detection.

10 The processor(s) **330**, and/or the GPU(s) **323**, can run various functional modules, as disclosed further herein with reference to **Fig. 3B**.

In some examples, memory **325** of processing circuitry **320** is configured to store data associated with the object detection process, e.g. comparatively transitory data. Non-limiting examples of data stored include orientation of the craft **110**, estimated sizes and distances of objects, pixels associated with each detected object etc. The memory can also store, for example, viewing angles of a gimbaled camera **230** for each image capture, and synchronized time of capture, t.

In some examples, object detection system **300** comprises data store **390**. In some examples, more long-term and persistent data is stored in the data store. Non-limiting examples shown include data which is tracked over time, view directions and angular fields of view of each fixed camera **210**, and classification and/or identification information associated with detected objects **120**.

Turning now to **Fig. 3B**, possible modules of processor(s) **330**, and or of GPU(s) **323**, of processing circuitry **320**, are shown. The processor(s) **330**, and/or the GPU(s) **323**, can run various functional modules. All of the modules can run on processor(s) **330**, all of them can run on the GPU(s) **323**, or a division of the modules among the processor(s) **330** and the GPU(s) **323** is possible, in any appropriate combination. These choices are in some cases an engineering decision. In some examples, the functional modules are software modules stored on datastore **390**, and when executed they are loaded to memory **325** and run by the processor(s) **330** and/or the GPU(s) **323**.

In some examples, component **330**, **320** comprises input/output (I/O) module **380**. In some examples, module **380** is configured to receive input from imaging devices **210**,

and possibly from IMU/INS **240, 242**. Example outputs are detected objects, and various alerts (e.g. of problem situations detected), sent to e.g. navigation system **243** and/or collision system **245**. In other implementations, there are separate input and output modules (not shown). These modules can communicate, in some implementations, with
5 I/O interfaces (not shown) comprised in system **300** – which can be e.g. various possible types of physical interface known in the art. The communication can be done e.g. over communications network(s) **247**.

In some examples, component **330, 320** comprises captured images interface and control module **333**. This module is configured to e.g. receive captured images via I/O
10 module **380**, including metadata such as time of capture, and/or to send commands to control the imaging devices **210** – e.g. telling them when to capture an image, telling a gimbaled device **230** to rotate to another field of view etc. In other implementations, there are separate modules for receiving captured images and for controlling imaging devices (not shown).

15 In some examples, component **330, 320** comprises resolution reduction module **335**. This module is configured to reduce the spatial resolution of captured image(s) **400**, e.g. as disclosed with reference to **Fig. 5**.

In some examples, component **330, 320** comprises reduced resolution detection module **337**. This module is configured to perform object detection on reduced resolution
20 image(s) **500**, using a neural network **327**, e.g. as disclosed with reference to **Fig. 5**.

In some examples, component **330, 320** comprises tiles creation module **340**. This module is configured to derive image tiles from captured image(s) **400**, e.g. as disclosed with reference to **Fig. 6**.

In some examples, component **330, 320** comprises tiles-based detection module
25 **345**. This module is configured to detect objects in image tiles, e.g. using image-tiles neural network(s) **327**, e.g. as disclosed with reference to **Fig. 6**.

In some examples, component **330, 320** comprises local search module **348**, referred to herein also as local inquiry and segmentation (LIAS) module **348**. This module is configured to perform a local inquiry on image tiles, e.g. where objects are detected in
30 more than one tile, e.g. using local inquiry neural network(s) **327**, e.g. as disclosed with reference to **Fig. 6**.

In some examples, component **330, 320** comprises results merging module **345**. This module is configured to merge objects detected e.g. in reduced resolution image(s), in image tiles, and/or in the local inquiry/segmentation, e.g. as disclosed with reference to **Figs. 7 and 8**. In other example implementations, the method involves several different merges, and each uses a different merging module.

In some examples, component **330, 320** comprises repetition and tracking module **355**. This module is configured to control repetitions of the detection process performed by e.g. modules **335, 337, 340, 345, 348, 350** etc. In some examples, the module also tracks the movement of objects in the body of water, e.g. as disclosed with reference to **Figs. 8**.

In some examples, component **330, 320** comprises classification and identification module **358**. This module is configured to perform classification and/or identification of detected objects, e.g. as disclosed with reference to **Figs. 8**. In some other implementations, there are separate modules for each of classification and identification.

Figs. 3A, 3B illustrates only general schematics of the system architecture, describing, by way of non-limiting example, certain aspects of the presently disclosed subject matter in an informative manner, merely for clarity of explanation. It will be understood that the teachings of the presently disclosed subject matter are not bound by what is described with reference to **Figs. 3A, 3B**.

Only certain components are shown, as needed, to exemplify the presently disclosed subject matter. Other components and sub-components, not shown, may exist. Systems such as those described with respect to the non-limiting examples of **Figs. 3A, 3B** may be capable of performing all, some, or part of the methods disclosed herein.

Each system component and module in **Figs. 3A, 3B** can be made up of any combination of software, hardware and/or firmware, as relevant, executed on a suitable device or devices, which perform the functions as defined and explained herein. The hardware can be digital and/or analog. Equivalent and/or modified functionality, as described with respect to each system component and module, can be consolidated or divided in another manner. Thus, in some embodiments of the presently disclosed subject matter, the system may include fewer, more, modified and/or different components, modules and functions than those shown in **Figs. 3A, 3B**. To provide one non-limiting example of this, in some examples, resolution reduction module **335** and reduced

resolution detection module **337** are combined into one module. Similarly, in some cases module **380** can instead be two modules – one for input and one for output. Similarly, in some cases module **355** is split into two modules – one for repetition and one for tracking.

One or more of these components and modules can be centralized in one location, or dispersed and distributed over more than one location, as is relevant. In some examples, certain components utilize a cloud implementation, e.g. implemented in a private or public cloud.

Each component in **Figs. 3A, 3B** may represent a plurality of the particular component, possibly in a distributed architecture, which are adapted to independently and/or cooperatively operate to process various data and electrical inputs, and for enabling operations related to computerized object detection. In some cases, multiple instances of a component may be utilized for reasons of performance, redundancy and/or availability. Similarly, in some cases, multiple instances of a component may be utilized for reasons of functionality or application. For example, different portions of the particular functionality may be placed in different instances of the component.

Communication between the various components of the systems of **Figs. 3A, 3B**, in cases where they are not located entirely in one location or in one physical component, can be realized by any signaling system or communication components, modules, protocols, software languages and drive signals, and can be wired and/or wireless, as appropriate. The same applies to interfaces such as I/O module **380**, and the external interface (not shown) of system **300**.

Attention is now drawn to **Fig. 4**, schematically illustrating an example generalized representation **400** of an image **400**, in accordance with some embodiments of the presently disclosed subject matter. The image was captured by an imaging device **210**, and thus is referred to herein as a captured image **400**. The captured image has a spatial resolution of $M \times N$ pixels (horizontal by vertical). The example image includes several objects. Objects **660** and **665** are relatively small on the image, due to e.g. their being small and/or distant objects, e.g. a buoy or a far-away fishing boat. Object **645** is of a medium size, comparatively. Objects **657** and **630** are relatively large objects, e.g. an ocean liner. Note that a smaller boat or other object **655** is in front of object **657**. Object **653** partially blocks or obscures object **650** on the captured image. Objects **640** and **643** appear, on the image, to touch.

In some examples, the one or more images **400** comprise one or more high resolution images **400**, and the at least one imaging device **210** comprises at least one high resolution imaging device **210**. Non-limiting examples of "high resolution" are disclosed herein. In some examples, the image **400** has a spatial resolution of at least 640 pixels in at least one dimension, e.g. horizontally and/or vertically. This level of resolution is considered, for purposes of the presently disclosed subject matter, as a "high" resolution. In some examples, one or more of the images have a resolution of at least 1080 pixels in at least one dimension. In some examples, one or more of the images have a resolution of at least 4000 pixels in at least one dimension. In some examples, one or more of the images have a resolution of at least 6000 pixels in at least one dimension. In some examples, one or more of the images have a resolution of at least 8000 pixels in at least one dimension. Examples include 6000 x 3376 pixels and 4000 x 2250 pixels. These are non-limiting examples.

It can be technically advantageous to detect images using the highest available device resolutions. In cases where cost is of concern, it can be technically advantageous to detect images using the highest available resolutions of standard cameras of the desired price range. Note that as camera or other device technology improves, the resolutions achievable by camera **210**, at various price points, increase. Detection on a high-resolution image has at least the example technical advantage of detecting also small and/or far away objects such as **660** and **665**. As will be shown with reference to **Fig. 6**, detection on image tiles can facilitate such a result.

More generally, in some implementation the imaging device(s) **210** is configured to view the objects with a viewing resolution associated with a human observer **115**, so as to enable the device(s) to replace or supplement the services of the human observer.

Attention is now drawn to **Fig. 5**, schematically illustrating an example generalized representation **500** of a reduced resolution image(s) **500**, in accordance with some embodiments of the presently disclosed subject matter. The image(s) **500** is generated by reducing the resolution of the captured image(s) **400**, from a resolution of $M \times N$ pixels to a lower resolution of $m \times n$ pixels. For example, $m < M$ and/or $n < N$.

Thus, the image resolution reduction process comprises reducing the horizontal resolution and/or the vertical resolution of the captured image(s) **400**. In some examples,

image **500** is referred to herein also as degraded image **500** or degraded resolution image **500**, having a degraded resolution relative to the original image **400**.

In one non-limiting example, a captured image of 6000 x 3376 pixels is reduced to 1280 x 1280 pixels. Thus, in some examples, the two dimensions are reduced by different
5 factors. In other examples, the two dimensions are reduced by the same factor. The factor can be an integer factor, or a non-integer factor.

This reduced resolution image(s) is sent to one or more neural networks **327**, referred to herein also as a reduced-resolution image neural network(s). Detection of objects, on the reduced resolution image(s) **500**, is performed utilizing this neural
10 network(s). Objects are detected at reduced resolution. In some examples, each pixel of the image is represented as a feature in the NN.

In the non-limiting example of the figure, the physical objects **630**, **640**, **643**, **645**, **650**, **653**, **655**, **657**, located on or near the body of water and captured in image **400**, are detected as detected objects **530**, **540**, **543**, **545**, **550**, **553**, **555**, **557**. Note that captured
15 objects **660** and **665** of captured image **400** are not detected at the reduced resolution. Since they appear as small objects, comprising a relatively small number of pixels of image **400**, once the resolution is reduced to derive image **500**, these smaller objects cannot be detected using the low-resolution neural network(s) **327**. Similarly, in some cases, relatively small and/or overlapping objects **650** and **653** are detected as a single
20 object **550** using the low-resolution neural network(s), due to the relatively low resolution. Similarly, in some cases, adjacent (or very close) objects **640** and **643** are detected as a single object **543** using the low-resolution neural network(s), due to the relatively low resolution. The low-resolution detection process is unable to distinguish them as two objects. The same applies to objects **655** and **657**.

25 The process disclosed above is a low-resolution detection process, at least in that the reduced resolution image **500**, in which this detection is performed, is of a resolution lower than the resolution of the original image **400**.

The detected objects **530**, detected at low resolution, are also referred to herein as low-resolution-detection detected objects, or simply as low-resolution detected objects. In
30 some examples, zero objects are detected (e.g. when the ocean or sea is empty in the area, or the only objects present are too small to be detected on image **500**).

Some example technical advantages of performing a low resolution (i.e. reduced resolution) detection process are that this facilitates detection using standard, inexpensive and/or relatively quick neural networks (NN). The use of inexpensive components can be advantageous in e.g. commercial / civilian navigation applications. The image **400** was captured at a comparatively high resolution – at least for reasons disclosed further herein with reference to **Fig. 6**. However, detection of objects on the high-resolution image **400** would require a high-resolution NN. In some examples, a NN of sufficient size to support such an image does not exist in the then-current technology, or it requires special development effort (rather than being readily available as standard and less expensive products). In some examples, such a NN supporting high resolution is expensive, and perhaps power-consuming, e.g. it uses a comparatively large number of computer resources. In addition, in some examples, detection using such a large NN requires a relatively long time, in some cases too long for practical use in navigation and collision avoidance applications.

Thus, reducing the image **400** to image **500**, and performing detection on an appropriately sized NN of comparatively small size, can enable some or all of these advantages. Since the detection is done on the entire captured image **400**, it can detect relatively large ships or other objects **657**, which span multiple image tiles **601** (per **Fig. 6**, further herein). That is, image **500** provides “the big picture” of the surroundings of the craft **110**, although at a lower resolution than the original image **400**.

In many implementations, the resolution is reduced by a factor of multiples of two (2) in each dimension, e.g. by a factor of 2, 4, 16 etc.

Attention is now drawn to **Figs. 6A and 6B**, schematically illustrating an example generalized representation **600** of image tiles **600**, in accordance with some embodiments of the presently disclosed subject matter. The figure depicts performing a high-resolution detection process, i.e. an image-tiles-based detection process, by splitting the one or more captured images **400** to image tiles. A set of image tiles **601**, **602** etc. is derived or generated from one or more of the image(s) **400**.

The **Fig. 6A** shows nine image tiles, **601**, **602**, **603**, **604**, **605**, **606**, **607**, **608**, **609**, arranged as 3 x 3, purely for illustration purposes. In other non-limiting examples, the tiles run five across (horizontal) by three tiles vertical, for a total of 15 tiles.

Each image tile **601** of the set corresponds to a sub-section of the captured image(s) **400**, that is the tile does not comprise the entire captured image. For example, an image tile such as **601** comprises fewer vertical pixels than the number of vertical pixels of the captured image(s) **400**, and/or fewer horizontal pixels than the number of horizontal pixels of the captured image. That is, in one or both dimensions the image tile has a second number of pixels, while the captured image has a first number of pixels, and the second number is smaller than the first number. That is, each image tile does not comprise all of the pixels of the full captured image **400**.

As one purely illustrative example of this, the image **400** is of resolution 6000 x 3900 pixels, while each image tile **601** is composed of 2000 x 1300 pixels. The image tile **601**'s pixel density in this case is the same as for the image **400**, but the image tile comprises fewer pixels, since it represents only a portion of the image **400**, that is a sub-section of the image -- and not the entire image **400**. The image tile represents an incomplete portion of image **400**. The numbers are presented here for illustration purposes only.

The resolution of at least some of the image tiles **601** (referred to herein also as an image tile resolution) is higher than the reduced resolution of the reduced resolution image(s) **500**. In some implementations, the resolution of at least some of the image tiles (e.g. expressed in terms of pixel density, that is pixels per inch or per centimeter) is the same as the resolution $M \times N$ of the original captured image(s) **400**. This special case can be referred to as "full resolution" image tiles.

In some other example implementations, the image tiles resolution of one or more image tiles is reduced somewhat from that of image **400**. For example, the resolution can be 2%, 5% or 10% reduced. These are non-limiting examples. This relatively small resolution reduction, of image tiles **601**, is by an amount less than the reduction performed to obtain reduced resolution image(s) **500**. The resolution of image tile(s) **601** is still high enough to achieve the desired result. For example, the reduced-resolution image tile is still of sufficiently high resolution to still detect distant and/or small objects such as buoy **665**.

In some examples, at least some adjacent image tiles, of the set **600** of image tiles, having at least partial horizontal overlap and/or at least partial vertical overlap. Thus, for example, tile **608** (shown by the dashed line) and tile **609** (shown by the dot-dash line)

have partial horizontal overlap, as shown by region **667**. Similarly, tile **606** (shown by the heavy dash and double dot line) and tile **609** have partial vertical overlap, as shown by region **627**. Similarly, region **663** shows an area of overlap of the four overlapping image tiles **604** (shown by the dot-dash line), **605** (shown by the heavy dashed line), **607** (shown by the heavy dash double-dot line) and **608**. Thus, the image tiles **600** cover part of the image **400**, or the whole image **400**, with optional overlap of tiles boundaries.

In some examples, the image tiles of the set have at least 5% overlap. In some examples, the tiles have at least 10% overlap.

In some examples, at least some adjacent image tiles of the set of image tiles have no horizontal overlap and no vertical overlap.

In some examples, at least some adjacent image tiles of the set are separated by gaps in the horizontal and/or the vertical direction. For example, there is a "gap" of a relatively small number of pixels between adjacent tiles, within a horizontal row and/or a vertical column. Turning to **Fig. 6B**, for example, an alternate case **600'** of tile arrangement is illustrated. Image tile **603'** (shown by the dashed line) and tile **602'** (shown by the dot-dash line) are separated by a horizontal gap, as shown by region **623**. Similarly, tile **601'** (shown by the dashed line) and tile **604'** are separated by a vertical gap, as shown by region **620**.

Note that only some tiles are shown in **Fig. 6B**, for ease of exposition.

In one non-limiting illustrative example, the rightmost pixel of the left tile corresponds to pixel #500, of the horizontal 1280 pixels in the captured image **400**, and the left pixel of right tile is pixel #505 of captured image **400**. Pixels #501-504 in the captured image are thus not in any derived image tile of set **600**. This in some cases can still provide a good enough detection of objects, of sufficient accuracy etc., despite the tiles "skipping" some pixels.

Similarly, some examples, at least some adjacent image tiles "touch" in the horizontal and/or the vertical direction. Thus, for example, in **Fig. 6B** image tiles **605'** and **606'** touch in the horizontal direction. That is, the right pixel of the left tile and the left pixel of the right tile are two horizontally adjacent pixels on the original captured image **400**. The tiles are of "consecutive" pixels. For example, the rightmost pixel of tile **605'** is tile i of captured image **400**, while the leftmost pixel of adjacent tile **606'** is tile $i+1$ of

image **400**. An area in the region of the border between the touching tiles **605'**, **606'** is referred to herein as a touching area (not shown in the figure, for ease of exposition).

Note that in the non-limiting example of the figures, the image tiles are not all of the same size. In other examples, they are of the same size.

5 Thus, there are multiple possible implementations of the image tiling, including those not disclosed herein.

In some examples, the set **600** of images tiles correspond to all $M \times N$ pixels of the one or more captured images **400**. That is, the set **600** of image tiles cover the one or more captured images **400**. In other examples, not all of the pixels of the captured image **400** are represented in image tiles. One example of this is the existence of gaps **620**, **623** 10 **400** are represented in image tiles. One example of this is the existence of gaps **620**, **623** in the example of **Fig. 6B**. The pixels associated with those gaps are not represented in any image tile. In such an implementation, the set **600'** of image tiles covers only a part of the captured image(s) **400**.

Similarly, in some implementations the image tile(s) **601** does not comprise every 15 pixel within the "area" framed by that particular tile. As one non-limiting example, presented purely for illustration, it can be that if every eighth column, and/or every eighth row, of pixels of captured image **400**, is not represented in image tile **601**, sufficient visual information is preserved to enable a high-quality object detection on the tile, while using a NN of relatively smaller size (compared to the size of a NN for an image tile which 20 skips no pixels/rows/columns). Again, all of the numerical examples are provided for illustrative purposes only.

Note that in a case where multiple images are captured, e.g. at time t by multiple cameras **226** and **228**, and/or by the same camera at multiple times t , $t+1$, etc., in some such implementations the image tiling process of **Fig. 6** is not performed for every image 25 on which detection is performed, but rather only for certain images.

In the second part of the high-resolution detection process, after splitting the captured image(s) **400** to image tiles **607**, system **300** performs a tile-based detection of objects **660**, on image tiles **607** of the set **600**, utilizing image tiles neural network(s) **327**. The system thereby obtains one or more tile-based detected objects **680**.

30 Note that, because the detection on tile **607** was done at a resolution higher than that of image **500**, e.g. at or close to the resolution $M \times N$ of captured image **400**, the system is able to detect the small buoy **660**, **680**. This was not possible to do on the

reduced-resolution image **500**. On the other hand, because the detection **680** of buoy **660** was done on a tile **607** which comprises only a portion of image **400**, rather than on the entire captured image **400**, it is possible to use smaller, cheaper, quicker and/or off-the-shelf neural network(s) **327**, sized for the relatively smaller number of pixels in tile **607**.

5 In some examples, it is not feasible to detect on the full image **400**, using a NN sized for it. Thus, the small buoy was detected, but in a relatively efficient manner.

Note that in some cases, detection of tile-based detected objects does not occur in every image tile. In one example, no detection is made in a particular, since the area is empty of ships/objects at that moment in time, or perhaps that tile covers a portion of the captured image **400** which corresponds to empty sky.

10

A final set of detected objects (not shown in a figure) can be obtained by merging the zero or more detected objects **540**, **557**, detected at reduced resolution, and the zero or more tiles-based detected objects **670**, **677**. That is, the system **300** has items of information (detected objects) obtained from both the process of **Fig. 5** and the process of **Fig. 6**, and it creates a coherent overall picture based on the output of the two steps, without duplications and without missing items.

15

For example, the merging determines that detected objects **557** and **667** both correspond to a single actual object **657**, and thus only one version of the image (or other data) of the object **657** is e.g. stored in a list, displayed to a user, and/or forwarded to another system or process. Thus, the merging process removes redundancies, multiple detections of what is in fact the same object.

20

Similarly, small object **660** was not detected in image **500**, but the final set of detected objects includes it, since it was detected as tile-based detected object **680**. The objects detected in each of **Figs. 5** and **6** can be referred to herein also as candidate detected objects, while the final set of detected objects, e.g. selected from the candidates, is a result of merging processes.

25

Note that some objects **657**, **630**, **602** are large enough, such that they appear in multiple tiles of **Fig. 6**. In some examples, the detection in each individual image tile detects these as separate objects. Thus, for instance, parts of object **630** are detected in each of the four tiles **604**, **605**, **607**, **608**, as four different tile-based objects **690'**, **690''**, **690'''**, **690''''**.

30

In some cases, it is easier to accurately detect an object in a single image **500** than to detect portions of the object in multiple images **690'**, **690''**, **690'''**, **690''''**. In some cases, it is easier to accurately performing segmentation of nearby, touching and/or overlapping objects such as **640** and **643** in a single image **614** than when portions of the objects are
5 in multiple images **690'**, **690''**, **690'''**, **690''''**. Note that in some examples, accurate segmentation of such objects using image tiles is possible when these objects (e.g. **650** and **653**) appear within a single image tile **604**.

Thus, merging first information derived from a low-resolution but "big picture" image, with second information derived from higher-resolution image tiles which focus
10 on a smaller part of the captured image, can obtain the benefits of both: accurate detection of large objects that cross image tile borders, together with detection of smaller objects (e.g. captured from a higher distance) which are detectable only at a higher resolution.

Detection can be performed on images captured at the highest-possible resolution devices of the then-current technology, to detect even small objects at even large
15 distances. Advantage is taken of the leading-edge cameras. However, detection on both the big-picture view of such high-resolution captured images **400**, and the more detailed high-resolution view, can be performed, using neural networks configured for fewer pixels than the number in the captured image – to improve performance – faster detection, using relatively cheaper neural networks **327** (e.g. an off the shelf NN), and thus using
20 relatively fewer computing resources. Note that a smaller NN typically works faster than a NN for a larger number of pixels. Thus, the method these figures facilitate use of leading-edge technology cameras **210** of a higher resolution than the standard and inexpensive NN of that point in time can handle. The specific values of the capability of cameras and NNs will of course change with the technology, over time. Phrased
25 differently, the method enables the use of standard hardware and NN's, for the fast detection of objects at the full possible camera resolution. The method achieves a balance between camera and NN capabilities.

In some edge cases, there may be detections on reduced-resolution image **500**, but not on the set **600** of image tiles. Consider a theoretical case where one large object is
30 detected in image **500**, and the image exactly fits exactly two adjacent image tiles **604**, **605**. One the image tiles, there is no background behind the object, and there is no detection. In such a case, zero tile-based detected objects were detected, and the merger

of the two detections, to generate a final set of detected objects, utilizes only the detected objects **557** detected at the lower resolution. In still other cases, only empty sea and sky are seen in an image **400** captured at time *t*, and there are zero objects detected in both image **500** and in the set **600** of tiles.

5 In some examples of edge cases, if there are zero objects detected in image **500**, the objects **677** detected per **Fig. 6** are the final set of detected objects. If there are zero objects detected per **Fig. 6**, the objects **557** detected in image **500** are the final set of detected objects. A "true" merging of the two detections occurs when at least one object is detected in each of the reduced-resolution detection and the image tile-based detection
10 methods.

In many practical implementations, it is advantageous to include an additional detection step, and in some cases also a merge step, to the above detection process. This additional step is optional, and in some cases can provide additional technical advantages.

For one or more LIAS-defined areas **612** associated with the set **600** of image
15 tiles, the system **300** performs a Local Inquiry and Segmentation (LIAS) at a spatial resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network **237**. The segmentation process in some examples uses known *per se* methods. The set of local-inquiry-based detected objects is merged with the image-tiles-based set of detected objects. An updated or modified final set of detected objects **690** can thus be
20 obtained, with the assistance of this additional process.

The merging removes duplicate detections of the same object. It can yield an improved set, in that it removes the duplicate detections, and/or provides more specific segmentation of nearby, touching, and/or overlapping objects. In some cases, it is easier to accurately performing segmentation of such objects in a single image **614** than when
25 portions of the objects are in multiple images **604**, **605** and **606**.

As will be disclosed with reference to **Fig. 7**, the merging of the various detected objects can be done in various orders.

The spatial resolution used for the LIAS method is referred to herein as a LIAS resolution or local inquiry resolution. In some examples, it is the same as the image-tiles
30 spatial resolution associated with one or more of the image tiles **601**. In some examples, the LIAS method uses a full-resolution detection, performed at the resolution (pixel

density) of the captured image **400**. This method is referred to herein also as a LIAS method.

In some implementations, the defined area(s) **612** is indicative of detection of one or more tiles-based detected objects in more than one image tile. In some examples, the LIAS-defined area(s) is indicative of detection of tiles-based detected object(s) in an overlap area **627**, **667** of image tiles. Thus, object **630** was detected **690'**, **690''**, **690'''**, **690''''**. Similarly, object **645** appears in both tiles **603** and **606**. Such areas are referred to herein also as overlap detection areas. The overlap can be between horizontally adjacent and/or vertically adjacent tiles.

In other examples, the defined area(s) is indicative of detection of tiles-based detected object(s) across boundaries of tiles, and not necessarily of detection in overlap areas. Thus, for example, object **657** appears in all of tiles **604**, **605** and **606**, and its detection **677** is across several tile boundaries. The defined areas are referred to herein also as LIAS-defined areas, local search areas or LIAS areas.

The LIAS method can thus be used to clarify situations on tiles boundaries and overlaps.

In some implementations, the LIAS-defined area **612**, **618** comprises portions of at least two image tiles, e.g. portions of **604**, **605**, **607** and **608**, or portions of **604**, **605** and **606**, that is comprise a plurality of partial image tiles – as illustrated in **Fig. 6A**. In the examples of the figure, the defined areas are intersections of at least two image tiles. In some implementations, the defined area **612** is not equal or identical to any image tile **604**.

In still some other implementations (not shown), a LIAS-defined area comprises one or more entire image tiles, e.g. the entirety of tiles **601** and **602**. This can occur, for example, where the defined area comprises more pixels than any of the image tiles **601** and **602**, but the number of pixels in the defined area is small enough to require a relatively small and fast/ inexpensive neural network (compared to the total number of pixels in captured image **400**).

Note that also objects **640** and **645**, for example, appear in more than one tile, and defined areas can be defined around them as well. However, these areas are not shown in the figure, purely for ease of exposition.

In some implementations, the LIAS process comprises generating at least one additional image tile **614**, corresponding to the one or more LIAS areas **612**, and performing the local search within the additional image tile(s). This additional image tile is defined specifically for the search. This image tile is referred to herein also as a LIAS,
5 local inquiry, or local search image tile **614**. For example, the LIAS image tile **614** is defined based on captured image **400**, around those pixels which are believed to contain the object **630**, which has apparently been detected in or across more than one image tile **604**. The LIAS image tile(s) **614** are fed into a local-inquiry neural network **327**, to perform the local inquiry detection.

10 In some cases, it is easier to accurately detect an object in a single image **614** than to detect portions of the object in multiple images **604**, **605**, **607** and **608**. The image **614** can be chosen to be sizes so as to include the entire object **630**, while maintaining a high enough resolution for the needed detection requirements, while keeping the LIAS NN **327** within the desired size. Again, in some implementations the use of standard hardware
15 and networks provides fast detection at full resolution, with segmentation.

The local search can utilize the same NN **327** as is used for the image-tiles-based detection, or a different NN of same size, or even a NN of a different size.

In some cases, the object **657** is relatively large, and it crosses the boundaries of a comparatively large number of tiles **604**, **605**, **606**. The local search area **618**, which
20 includes the large object, is thus also of a comparatively large size. In some examples of such cases, the system can decide to not perform LIAS with respect to such an area **618**, and instead to rely on full-image-based detection in reduced resolution image **500**. Such a decision can be made if the defined area **618** is of a size which requires an overly large neural network, which is beyond the capabilities of the LIAS sub-system, and/or which
25 will not have the desired performance.

Since each tile is defined based on pixels of captured image **400**, once the pixels of interest in tiles **604**, **605**, **607** and **608**, corresponding to detected object **630**, are determined, they can be mapped to the corresponding pixels on image **400**, and the local inquiry image tile **614** be defined accordingly, e.g. per known *per se* methods. The same
30 type of pixel mapping can also be used for the merging of image-tiles-based detected objects and objects detected at reduced resolution, and for the merging of the image-tiles-based detected objects and the one or more local-inquiry-based of detected objects **690**.

Note that, in contrast to e.g. image-tile-detection objects **670**, image-tile-detection objects such as **680'** and **685'**, as well as detection of the objects **650** and **653**, are detected fully in only one tile, and thus no LIAS process need be performed for them.

Note that reference numbers are not shown for each image-tile-based detected
5 object, and LIAS-based detected object, in **Fig. 6**, for reasons of clarity of exposition on the figure.

In other examples, the method comprises the image-tile based detection, and the LIAS-based detection, and a merging of the two detections.

In some example implementations, the method disclosed with reference to **Figs.**
10 **4, 5, 6** and **7** is performed in real time. Thus, in some examples the image(s) **400** is captured in real time, and the system **300** is configured to perform the detection method in real time. In some examples, this can facilitate an avoidance of collision between the waterborne craft **100** and one or more of the objects **133, 140**. The real time, or on-line, nature of the method can thus provide advantages in real life situations such as collision.

Thus, in some examples, the system **300** is configured to perform the method
15 within 0.01 to 60 seconds. In another example, the system is configured to perform the method within 0.01 to 1 second. In another example, the system is configured to perform the method within 1 to 60 seconds. In another example, the system is configured to perform the method within 10 to 90 seconds. Such time frames can in some cases enable
20 a response of waterborne craft within 1 to 15 min. Based on the distance between watercraft **110** and a particular object, their speeds, and their headings, a larger or smaller response time is needed to avoid a collision, or e.g. to make a turn in the correct fashion. These numbers are all non-limiting examples.

In other implementations, the method is performed in near real time.

In some examples, the specific engineering choices when designing the method
25 (e.g. steps and their order), and of the equipment/components used, are attempting to perform an optimization of at least time needed to perform, cost of computing, and good enough performance. That is, in many cases there is a need to accurately, and correctly, detect (and possibly classify, identify, and/or determine position and size of) even small
30 and/or distant objects, in a quick manner to enable a timely response/reaction, while using limited and comparatively expensive resources. The various types of detection, and their merging, can facilitate this optimized behaviour.

Attention is now drawn to **Fig. 7**, schematically illustrating an example generalized representation **700** of a data flow **700**, in accordance with some embodiments of the presently disclosed subject matter. The rectangles denote actions, methods, processes, or associated systems, while the parallelograms indicate input/output data items.

The example data flow starts at block **710**, with one or more images **400** captured by an imaging device(s) **210**. In the example, the high resolution of this image is 6000 x 3376. Note that all numbers in this figure are for illustrative purposes only.

Considering first the right side of the flow, in block **720** the resolution of image **400** is decreased / reduced / degraded, to yield reduced resolution image **500**. In the example, the reduces spatial resolution is 1280 x 1280.

The image **500** is sent **724** to low-resolution neural network(s) **327**, e.g. a deep neural network (DNN). The output **728** of the NN is a set #1 of detected objects **557**, detected at the reduced resolution.

Considering now the left side of the flow, in block **730** the captured image **400** is split into one or more image tiles **601**. In the example, the split is into 15 tiles of size 1280 x 1280 pixels, e.g. arranged as 5 x 3. The output **733** is the set **600** of tiles.

The image tile(s) is sent **735** to a high-resolution image tiles neural network(s) **327**, e.g. a deep neural network. The output **737** of the NN is a second set #2 of detected objects **737** (tile-based detected objects), detected at the relatively high resolution.

In block **740**, areas **612** are defined for local inquiry and segmentation. For example, LIAS image tiles **614** are defined. This is done, for example, for objects **657** which are detected in more than one image tile. An example output **745** is the set #2 of detected objects, combined with the LIAS requests.

In block **750**, the LIAS is performed on the set #2, per the LIAS requests. The LIAS image tile(s) is sent **750** to a high-resolution local inquiry and segmentation neural network(s) **327**, e.g. a deep neural network. The output **755** of the NN is a third set #3 of detected objects (local-inquiry-based detected objects), detected at this relatively high resolution.

In block **760**, the detected objects #3 and detected objects #2 are merged. A merged set #4 of objects, e.g. detected at high resolution, is the output **765**.

In block 770, the set of detected objects #4 and low-resolution detected objects #1 are merged. A merged set #5 of detected objects is the output 780.

Note that the example data flow 700 discloses three different detections and two mergers. In the non-limiting example of the figure, the order of the merging is to first
5 merge the results of the image tiles detection and of the local search, and then to merge the output of this merging with the results of the reduced-resolution detection on image 500. In another example, the order of merging is the reverse. Other orders and combinations of merging are possible. For example, in one implementation the three outputs are all merged, as three parallel inputs, at the same time.

10 Thus, depending on the particular implementation chosen, it can be that one merge process yields a "final" (or "intermediate") set of detected objects, and the second merge process modifies that set to yield an "updated" or "modified" final set of detected objects.

In some implementations, the low-resolution detection process (blocks 720, 724, 728) is performed prior to the high-resolution (i.e. image-based) detection process
15 associated with image tiles 601. In other examples, the order is reversed. In still other cases, the processes are performed at least partially in parallel, e.g. simultaneously. In some implementations, this is an engineering decision, balancing the number / amount of computer resources utilized and the time to perform the entire process.

To exemplify this wide range of options, Fig. 7 illustrates the non-limiting
20 example of LIAS detection occurring after image tiles-based detection, and the two occur in parallel with the low-resolution image detection. On the other hand, Fig. 8, disclosed below, illustrates the non-limiting example of three detections occurring all in a serial manner – low-resolution detection, followed by tile-based detection and then by local inquiry-based detection.

25 Attention is drawn to Figs. 8A-8C, schematically illustrating a generalized flow chart diagram 800 of a flow of a process or method, for object detection, in, in accordance with some embodiments of the presently disclosed subject matter. This detection process 800 is, in some examples, carried out by systems such as those disclosed with reference to Figs. 2, 3A and 3B. The flow 800 starts at 803 of Fig. 8A.

30 According to some examples, one or more images 400 are captured (block 803). In some examples, this is performed by one or more imaging devices 210, e.g. cameras. In some examples, the cameras are controlled by system 300 using captured images interface

and control module **333** of processing circuitry **320**, e.g. interfacing via I/O module **380** and communication network **247**.

According to some examples, the one or more captured images **400** are received (block **807**). In some examples, this is performed by captured images interface and control
5 module **333**, e.g. interfacing via I/O module **380** and communication network **247**.

According to some examples, inputs from the imaging device(s) **210** are synchronized with inputs from IMU **242** (or INS **240**) (block **809**). In some examples, this is performed utilizing input/output module **380**, or by some other module (e.g. a dedicated module not shown in the figure), e.g. interfacing with the INS/IMU using
10 communication network **247**. This synchronization information can facilitate association over time of e.g. IMU readings with the correct image **400** captured at or near the same time.

According to some examples, an image resolution reduction is performed (block **812**). For example, the spatial resolution of captured images **400** is reduced, and one or
15 more reduced resolution images **500** is generated – for example using methods disclosed with reference to **Figs. 5** and **7**. In some examples, this is performed by resolution reduction module **335**. In other examples, the resolution reduction is performed by an external system (not shown).

According to some examples, object detection is performed, on the reduced
20 resolution image(s) **500**, utilizing a neural network(s) **327** of relatively low resolution (block **815**). In some examples, this is performed by reduced resolution detection module **337**, for example using methods disclosed with reference to **Figs. 5** and **7**. One or more detected objects **557**, detected at reduced resolution, are obtained.

According to some examples, a set **600** of image tiles **601** is derived from at least
25 one captured image **400** (block **820**). The image(s) **400** is split into image tiles – for example using methods disclosed with reference to **Figs. 6** and **7**. Each image tile **601** of the set corresponds to a sub-section of the captured image **400**. In some examples, this is performed by tiles creation module **340**. In other examples, the creation or generation of the image tiles is performed by an external system (not shown).

30 According to some examples, a tile-based detection of objects is performed, on image tiles **601**, **602** of the set **600**, utilizing a neural network(s) **327** of relatively high resolution (block **825**). In some examples, this is performed by tiles-based detection

module **345**, for example using methods disclosed with reference to **Figs. 6** and **7**. One or more tile-based detected objects **677**, detected at high resolution, are obtained.

The flow continues **A** to block **827** on **Fig. 8B**.

According to some examples, one or more LIAS-defined areas **612** on image **400**
5 are defined (block **827**). These local inquiry areas **612** are associated with the set of image tiles. This is done, for example, using methods disclosed with reference to **Figs. 6** and **7**. The defined areas are indicative of detection of the tiles-based detected object(s) **677** in more than one image tile **605**. This step is performed, in some examples, by local search module **348**.

10 According to some examples, at least one additional image tile **614** is generated, corresponding to the one or more LIAS-defined areas **612** (block **830**). In some examples, this is performed by local search module **348**, for example using methods disclosed with reference to **Figs. 6** and **7**.

According to some examples, a Local Inquiry and Segmentation (LIAS), i.e. a local
15 search, is performed, at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network **327** of relatively high resolution (block **832**). This LIAS process involves performing a local-inquiry detection. In some examples, this is performed by local search module **348**, for example using methods disclosed with reference to **Figs. 6** and **7**. One or more local-inquiry-based detected objects **690**, detected
20 at high resolution, are obtained.

According to some examples, local-inquiry detected object(s) and tile-based
detected object(s) are merged (block **840**). The merger is of objects detected in blocks
825 and **832**. In some examples, this is performed by results merging module **350**, or by
local search module **348**, for example using methods disclosed with reference to **Figs. 6**
25 and **7**. A merged set of detected objects, detected at high resolution, is obtained.

According to some examples, the merged set of detected object(s), from block **840**,
and object(s) detected at the reduced resolution (i.e. low-resolution-based sets) are
merged (block **841**). In the example of this figure, the merger is of objects detected in
blocks **815** and **840**. In some examples, this is performed by results merging module **350**,
30 for example using methods disclosed with reference to **Figs. 6** and **7**. In some examples,
a final set of detected objects, or an updated final set, is obtained.

According to some examples, the final set of detected object(s) is merged with an additional final set(s) of detected objects, associated with a different field(s) of View (FOV) of the imaging device(s) (block **842**). In the example of this figure, the merger is of objects detected in blocks **841** for two different FOVs. In some examples, this is performed by results merging module **350**.

Thus, the low resolution **530**, tiles-based **690'**, **690''**, **690'''**, **690''''**, and LIAS-based **690** detected objects of a first FOV are merged with the corresponding objects associated with a second adjacent FOV. Consider, for example, a first image **400** captured at time *t* by a first camera **226** of **Fig. 2**, with a particular view direction and thus seeing FOV **A**, and a second image **400'** (not shown in the figures) captured at the same time *t* by a second camera **227**, seeing an adjacent FOV **B**. If a physical object **645** is detected in both images, there are technical advantages to merging the two detections, to remove the duplication.

The two FOVs **A** and **B** are different. In the particular case of this non-limiting example, the two FOVs have partial overlap. Object **645** in some examples appears in the area of FOV overlap. In others, object **645** is relatively large, and crosses the borders of the FOVs, much like object **657** of **Fig. 6** crosses the borders of three image tiles. As disclosed above with reference to **Fig. 2**, in other examples the adjacent FOVs touch at their edges/borders, or even have gaps between them, and cameras such as pan/tilt and gimbaled cameras can be utilized to facilitate non-identical FOVs.

In some examples, the merging is done separately for each type of detection, and then the results are merged. Thus, the process can comprise at least one of the following:

- (a) the low-resolution detection process comprises merging the detected object(s), which were detected at the reduced resolution in a first image captured at a first angular FOV of the imaging device(s), with additional detected object(s), which were detected at the reduced resolution in a second image captured in a second, non-identical, angular FOV of the of the imaging device(s).
- (b) the high-resolution, tiles-based, detection process comprises merging the tiles-based detected object(s) with additional tiles-based detected objects. The tiles-based detected object(s) is associated with the image tiles of the set, captured in the first FOV. The additional tiles-based detected objects are associated with additional image tiles captured in the second FOV.

(c) the high-resolution, local inquiry-based, detection process comprises merging the local inquiry-based detected object(s) with additional local inquiry-based detected objects. The local inquiry-based detected object(s) is associated with the image tiles, captured in the first FOV. The additional local inquiry-based detected objects are associated with the additional image tiles captured in the second FOV.

In this example, the three outputs of these three merges are then merged to obtain a final set of detected objects that considers the multiple FOVs / multiple cameras.

In another non-limited example, merging is done separately for each FOV. Thus, for image **400** associated with the first FOV, merging is done (in whichever order) of the reduced-resolution-based, tiles-based, and LIAS-based detected objects. The same is done again, for image **400'** associated with the second FOV. Then the two outputs are merged. Thus, the order and combinations of mergers can have various implementations.

According to some examples, spatial position of detected objects of the final set are classified (block **847**). In some examples, this is performed by classification and identification module **358**, or by another module not shown in **Fig. 3B**, e.g. using known *per se* methods. In some examples, the position of a detected objects is determined based at least a relative position of the detected object, relative to the captured image **400**, and based at least on an orientation of the waterborne platform **110** associated with same captured image, i.e. based on the orientation of the platform at the time *t* of capture of image **400**.

The flow continues **B** to block **850** on **Fig. 8C**.

According to some examples, detected objects of the final set are classified (block **850**). In some examples, this is performed by classification and identification module **358**, e.g. using known *per se* methods. For example, a comparatively general class, e.g. aircraft carrier, tanker, fishing boat, pleasure yacht, buoy etc., is determined. This step is optional, and it is performed in only some implementations.

According to some examples, detected objects of the final set are identified (block **854**). In some examples, this is performed by classification and identification module **358**, e.g. using known *per se* methods. For example, a more specific class is identified, e.g. "XXXX class aircraft carrier", or a specific vessel is identified, e.g. "USS YYY". This step is optional, and it is performed in only some implementations.

Note that in some implementations or applications, e.g. for the basic case of collision avoidance, there is no need to classify nor identify. In some other examples, only some of the objects are classified and/or identified. Similarly, in some other examples, classification and/or identification is performed in steps before generation of the final set of detected objects – e.g. on the objects detected at reduced resolution, and/or on those detected using the image tiles etc.

According to some examples, information indicative of the final set of detected objects is output (block **857**). In some examples, this is performed by input/output module **380**, sending the output to a display or to an external system (not shown) via communications network **247**. In some examples, the outputted information comprises position information (e.g. longitude and latitude coordinates) associated with one or more detected objects of the final set.

Other information that can be sent, in some implementations, includes, among others, the bearing of an object **120** relative to watercraft **110**, the captured image of the object, an estimated bounding box (e.g. size in pixels), an estimation of the object's size (e.g. height), an estimate of its distance from watercraft **110**.

The output can be e.g. in the form of a list of objects and parameters/information, and/or a display.

According to some examples, the method is repeated at least one additional time (block **870**). All, or some, of blocks **805** to **857** are repeated. In some implementations, this is repeated in a continual manner, throughout the period of motion of craft **110**, to ensure e.g. safe navigation that is e.g. collision-free. In some examples, the method is also repeatedly performed while the ship is at anchor, to provide a continual monitoring of the nearby aquatic environment. In some examples, this is performed utilizing repetition and tracking module **355**.

According to some examples, tracking of the detected objects is performed (block **875**). This tracking can be facilitated by the repetition. In some examples, this is performed utilizing repetition and tracking module **355**. A continuous identity of a particular detected object is identified, and tracked, over time (e.g. in consecutive frames), in the Field of View of a single imaging device, as well as across multiple FOVs of multiple devices, and its trajectory can be evaluated. This can be performed for multiple

detected objects. This step can be performed continually, as the detection process is repeated over time per step 870.

In some implementations, based on the tracking, the system 300 can provide at least certain example technical advantages. For example, the speed and heading of objects 127
5 can be estimated. Also, if relatively major changes are detected between several images 400 captured over time, an alert can be sent to e.g. an external system, and/or to a visual (and/or audio) display, indicating that there may be an issue. Examples of such situations are that a large ship 133 disappears from image to image, or the object 127 is suddenly heading towards the ship 110, or it has sudden acceleration (e.g. it was traveled slowly,
10 and appeared as a fishing boat, but it now is moving at speedboat speeds). More steps concerning determining changes in detected-object position are disclosed in the following blocks.

An example advantage of the tracking is as follows: waterborne objects 657 such as ships move relatively slowly. Therefore, high-resolution cameras 210 can be utilized
15 to repeatedly capture images 400, at a relatively slow rate (e.g. 2 seconds), which will not overload the pipeline of the image processing in system 300. This can facilitate relatively easy tracking of objects, at high accuracy, using hi-resolution cameras. In some examples, such a slow rate of image capture is less relevant for images captured by e.g. an airborne craft such as an aircraft.

20 The tracking can provide continual situation awareness as the waterborne craft 110 travels. In addition, monitoring of images over time can in some cases facilitate detection of false alarms. A non-limited example of false alarms is an object apparently seen in time but not at time $t+1$, e.g. what was seen at time t was in fact a reflection of the sun.

The flow ends. As indicated above, the sequence of some or all of these steps can
25 in some cases continue, in a loop.

Thus, per the flow chart, the presently disclosed subject matter includes a computerized method of object detection, the method performed by a processing circuitry of an object detection system, the method comprising:

A) receive one or more images captured by at least one imaging device mounted on a
30 waterborne platform,
the one or more images comprising one or more objects, which are located within a body of water;

B) perform the following processes:

i) perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:

(1) receive a set of image tiles, derived from at least one image of the one or more images,

where each image tile of the set corresponds to a sub-section of the at least one image,

where a spatial resolution of each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and

(2) perform a tile-based detection of objects, on image tiles of the set, utilizing at least one image tiles neural network, thereby obtaining one or more tile-based detected objects;

ii) for one or more defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network, thereby obtaining one or more local-inquiry-based detected objects,

where the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile,

where each LIAS-defined area comprises portions of at least two image tiles; and

iii) merge the one or more local-inquiry-based detected objects with at least one of the image-tiles-based set of detected objects and the one or more detected objects detected at reduced resolution, thereby obtaining a final set of detected objects.

Similarly, per the flow chart, the presently disclosed subject matter includes a computerized method of object detection performed by a processing circuitry of an object detection system, the method comprising:

A) receive one or more images captured by at least one imaging device mounted on a waterborne platform,

the one or more images comprising one or more objects, which are located within a body of water;

B) perform the following processes:

i) perform a low-resolution detection process, comprising:

1. receive one or more reduced resolution images, generated by performing an image resolution reduction, the image resolution reduction comprising reducing the spatial resolution of the one or more images; and

2. perform object detection, on the one or more reduced resolution images, utilizing a neural network, thereby obtaining one or more detected objects detected at reduced resolution;

ii) perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:

(1) receive a set of image tiles, derived from at least one image of the one or more images,

wherein each image tile of the set corresponds to a sub-section of the at least one image,

wherein a spatial resolution of each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and

(2) perform a tile-based detection of objects, on image tiles of the set, utilizing at least one image tiles neural network, thereby obtaining one or more tile-based detected objects;

iii) for one or more defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network, thereby obtaining one or more local-inquiry-based detected objects,

where the one or more defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile,

where each LIAS-defined area comprises portions of at least two image tiles;

and

iv) merge the one or more detected objects detected at reduced resolution, the image-tiles-based set of detected objects, and the one or more local-inquiry-based detected objects,

thereby obtaining a final set of detected objects.

Thus, the method can combine:

(A) low-resolution detection and image-tiles-based detection (i.e. the high resolution detection process),

- (B) image-tiles-based detection and local inquiry-based detection, or
- (C) all three of image-tiles-based detection, local inquiry-based detection, and low-resolution detection.

Any of these methods can yield a final set of detected objects.

5 In some embodiments, one or more steps of the flowcharts exemplified herein may be performed automatically. The flow and functions illustrated in the flowchart figures may for example be implemented in system **300** and in processing circuitry **320**, and they may make use of components described with regards to **Figs. 2, 3A and 3B**. It is also noted that whilst the flowchart is described with reference to system elements that realize
10 steps, such as for example system **300** and processing circuitry **320**, this is by no means binding, and the operations can be carried out by elements other than those described herein.

It is noted that the teachings of the presently disclosed subject matter are not bound by the flowcharts illustrated in the various figures.

15 For example, some of the operations or steps can be integrated into a consolidated operation, or can be broken down into several operations, and/or other operations may be added. As a non-limiting example, in some cases blocks **840, 841, 842**, can be combined.

In embodiments of the presently disclosed subject matter, fewer, more and/or different stages than those shown in the figures can be executed. As one non-limiting
20 example, certain implementations may not include one or more of block **875**.

Similarly, in some implementations, the operations can occur out of the illustrated order. One or more stages illustrated in the figures can be executed in a different order and/or one or more groups of stages may be executed simultaneously. As one example, blocks **812, 815** can be performed before or after blocks **820-840**.

25 In the claims that follow, alphanumeric characters and Roman numerals, used to designate claim elements such as components and steps, are provided for convenience only, and do not imply any particular order of performing the steps.

It should be noted that the word “comprising” as used throughout the appended claims, is to be interpreted to mean “including but not limited to”.

30 While there has been shown and disclosed examples in accordance with the presently disclosed subject matter, it will be appreciated that many changes may be made therein without departing from the spirit of the presently disclosed subject matter.

It is to be understood that the presently disclosed subject matter is not limited in its application to the details set forth in the description contained herein or illustrated in the drawings. The presently disclosed subject matter is capable of other embodiments and of being practiced and carried out in various ways. Hence, it is to be understood that the phraseology and terminology employed herein are for the purpose of description and should not be regarded as limiting. As such, those skilled in the art will appreciate that the conception upon which this disclosure is based may readily be utilized as a basis for designing other structures, methods, and systems for carrying out the several purposes of the present presently disclosed subject matter.

It will also be understood that the system according to the presently disclosed subject matter may be, at least partly, a suitably programmed computer. Likewise, the presently disclosed subject matter contemplates a computer program product being readable by a machine or computer, for executing the method of the presently disclosed subject matter, or any part thereof. The presently disclosed subject matter further contemplates a non-transitory machine-readable or computer-readable memory tangibly embodying a program of instructions executable by the machine or computer for executing the method of the presently disclosed subject matter or any part thereof. The presently disclosed subject matter further contemplates a non-transitory computer readable storage medium having a computer readable program code embodied therein, configured to be executed so as to perform the method of the presently disclosed subject matter.

Those skilled in the art will readily appreciate that various modifications and changes can be applied to the embodiments of the invention as hereinbefore described without departing from its scope, defined in and by the appended claims.

CLAIMS:

1. A computerized method of object detection, the method performed by a processing circuitry of an object detection system, the method comprising:
 - A) receive one or more images captured by at least one imaging device mounted on a waterborne platform,
the one or more images comprising one or more objects, which are located within a body of water;
 - B) perform the following processes:
 - i. perform a low-resolution detection process, comprising:
 1. receive one or more reduced resolution images, generated by performing an image resolution reduction, the image resolution reduction comprising reducing the spatial resolution of the one or more images; and
 2. perform object detection, on the one or more reduced resolution images, utilizing a neural network, thereby obtaining one or more detected objects detected at reduced resolution;
 - ii. perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:
 1. receive a set of image tiles, derived from at least one image of the one or more images,
wherein each image tile of the set corresponds to a sub-section of the at least one image,
wherein a spatial resolution of the each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and
 2. perform a tile-based detection of objects, on image tiles of the set, utilizing at least one image tiles neural network, thereby obtaining one or more tile-based detected objects; and
 - iii. merge the one or more detected objects detected at reduced resolution, and the one or more tiles-based detected objects,
thereby obtaining a final set of detected objects.
2. The computerized method of claim 1, the method further comprising:

C) for one or more LIAS-defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network, thereby obtaining one or more local-inquiry-based detected objects,

wherein the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile,

wherein each LIAS-defined area comprises portions of at least two image tiles;

and

D) merge the one or more local-inquiry-based detected objects with at least one of the image-tiles-based set of detected objects and the one or more detected objects detected at reduced resolution,

thereby facilitating obtaining an updated final set of detected objects.

3. The computerized method of any one of claims 1 to 2, the method further comprising:
E) performing the image resolution reduction.

4. A computerized method of object detection, the method performed by a processing circuitry of an object detection system, the method comprising:

A) receive one or more images captured by at least one imaging device mounted on a waterborne platform,

the one or more images comprising one or more objects, which are located within a body of water;

B) perform the following processes:

i) perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:

(1) receive a set of image tiles, derived from at least one image of the one or more images,

wherein each image tile of the set corresponds to a sub-section of the at least one image,

wherein a spatial resolution of the each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and

(2) perform a tile-based detection of objects, on image tiles of the set, utilizing at least one image tiles neural network, thereby obtaining one or more tile-based detected objects;

- ii) for one or more LIAS-defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network, thereby obtaining one or more local-inquiry-based detected objects, wherein the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile, wherein each defined area comprises portions of at least two image tiles; and
 - iii) merge the one or more local-inquiry-based detected objects with the image-tiles-based set of detected objects, thereby obtaining a final set of detected objects.
5. The computerized method of any one of claims 2 to 4, wherein the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in at least one of an overlap area and a touching area of image tiles.
 6. The computerized method of any one of claims 1 to 5, wherein the one or more images comprise one or more high resolution images, wherein the at least one imaging device comprises at least one high resolution imaging device.
 7. The computerized method of any one of claims 1 to 6, the method further comprising:
 - F) deriving the set of image tiles.
 8. The computerized method of any one of claims 1 to 7, wherein the at least one image captured in real time, wherein the system is configured to perform the method in one of: real time; near real time; and online,
 - thereby facilitating an avoidance of collision between the waterborne platform and an object of the one or more objects.
 9. The computerized method of claim 8, wherein the system is configured to perform the method within 0.01 to 1 second.
 10. The computerized method of claim 8, wherein the system is configured to perform the method within 1 to 60 seconds.
 11. The computerized method of any one of claims 1 to 10, wherein the method is repeated at least one additional time.
 12. The computerized method of claim 11, wherein the system is configured to synchronize inputs from the at least one imaging device and at least one of an Inertial Measurement Unit (IMU) and an Inertial Navigation System (INS).

13. The computerized method of any one of claims 11 to 12, the method further comprising:
 - G) performing tracking of the one or more objects, based on the repetition.
14. The computerized method of any one of claims 1 to 13, wherein the waterborne platform is one of: a ship and a boat.
15. The computerized method of any one of claims 1 to 14, wherein the final set of detected objects comprises at least one of: another waterborne platform, a buoy, an obstacle, and a lighthouse.
16. The computerized method of any one of claims 1 to 15, wherein the one or more captured images having a resolution of at least 640 pixels in the at least one dimension.
17. The computerized method of any one of claims 1 to 16, wherein the one or more captured images having a resolution of at least 1080 pixels in the at least one dimension.
18. The computerized method of any one of claims 1 to 17, wherein at least one of the following is true:
 - [a] at least some adjacent image tiles of the set of image tiles have at least partial horizontal overlap and/or at least partial vertical overlap; and
 - [b] at least some adjacent image tiles of the set of image tiles touch each other.
19. The computerized method of any one of claims 1 to 18, the method further comprising:
 - H) classify the at least one detected object of the final set.
20. The computerized method of any one of claims 1 to 19, the method further comprising:
 - I) identify the at least one detected object of the final set.
21. The computerized method of any one of claims 1 to 20, wherein the at least one imaging device is comprised in one or more payloads, wherein each payload comprising one or more imaging devices, wherein the one or more payloads are configured to cover an angle up to 360 degrees around the waterborne platform.
22. The computerized method of any one of claims 1 to 21, wherein the at least one imaging device is a camera.
23. The computerized method of any one of claims 21 to 22, the method further comprising:

J) merge the final set of detected objects with an additional final set of detected objects associated with image capture in a different field of view (FOV) of at least one additional imaging device.

24. The computerized method of any one of claims 1 to 24, the method further comprising:

K) outputting information indicative of the final set of detected objects.

25. The computerized method of claim 25, wherein the outputted information comprises position information associated with at least one detected object of the final set.

26. A computerized method of object detection, the method performed by a processing circuitry of an object detection system, the method comprising:

A) receive one or more images captured by at least one imaging device mounted on a waterborne platform,

the one or more images comprising one or more objects, which are located within a body of water;

B) perform the following processes:

i. perform a low-resolution detection process, comprising:

(1) receive one or more reduced resolution images, generated by performing an image resolution reduction, the image resolution reduction comprising reducing the spatial resolution of the one or more images; and

(2) perform object detection, on the one or more reduced resolution images, utilizing a neural network, thereby obtaining one or more detected objects detected at reduced resolution;

ii. perform a high-resolution detection process based on a splitting of the one or more images to image tiles, comprising:

(1) receive a set of image tiles, derived from at least one image of the one or more images,

wherein each image tile of the set corresponds to a sub-section of the at least one image,

wherein a spatial resolution of each image tile is higher than the reduced spatial resolution of the one or more reduced resolution images; and

- 56 -

- (2) perform a tile-based detection of objects, on image tiles of the set, utilizing at least one image tiles neural network, thereby obtaining one or more tile-based detected objects;
 - iii. for one or more LIAS-defined areas associated with the set of image tiles, perform a Local Inquiry and Segmentation (LIAS) at a resolution higher than the reduced resolution, utilizing at least one local-inquiry neural network, thereby obtaining one or more local-inquiry-based detected objects,
 - where the one or more LIAS-defined areas are indicative of detection of the one or more tiles-based detected objects in more than one image tile,
 - where each LIAS-defined area comprises portions of at least two image tiles;
 - and
 - iv. merge the one or more detected objects detected at reduced resolution, the image-tiles-based set of detected objects, and the one or more local-inquiry-based detected objects,
 - thereby obtaining a final set of detected objects.
27. A computerized object detection system, comprising a processing circuitry, configured to perform the method of any one of claims 1 to 26.
28. The computerized object detection system of claim 27, further comprising the at least one imaging device.
29. A non-transitory computer readable storage medium tangibly embodying a program of instructions that, when executed by a processing circuitry of an object detection system, cause the processing circuitry to perform the method of any one of claims 1 to 26.