

key challenges of these approaches are higher computational cost, view-point dependent learning and relatively large human image size requirement.

[0010] The embodiments described here address some of these problems of existing systems.

SUMMARY

[0011] The disclosed embodiments provide methods, devices and systems for intelligent analysis of video images to detect objects, such as human objects.

[0012] In certain embodiments, a method of detecting human objects in a video comprises determining that certain pixels of a video image are foreground pixels, the group of foreground pixels constituting a foreground blob set of one or more foreground blobs; for each of N locations within the video image, where N is an integer, comparing a predetermined shape with the foreground blob set to obtain a corresponding probability of a human at the location, thereby obtaining N probabilities corresponding to the N locations; and using the N probabilities, determining X humans are represented by the foreground blob set, where X is whole number.

[0013] A method of detecting human objects in a video, may comprise determining pixels of a video image of a real world scene are foreground pixels, the group of foreground pixels constituting a foreground blob set of one or more foreground blobs; and for each of N locations within the video image, where N is an integer, comparing a predetermined shape with the foreground blob set to determine X humans are represented by the foreground blob set, where X is whole number.

[0014] Methods may include determining a location of each of the X humans. The locations of each of the X humans may be determined as a location within a horizontal plane of the real world, such as a location on a physical ground plane of the real world.

[0015] The detection of the human objects may be used to count humans, for crowd analyses and for other event detections.

[0016] System and devices are disclosed which may be configured to perform such methods.

[0017] Computer readable media containing software that may be used to configure a computer to perform the operations described herein and comprise further embodiments of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] Example embodiments will be more clearly understood from the following detailed description taken in conjunction with the accompanying drawings. The figures represent non-limiting example embodiments as described herein.

[0019] Figure 1 illustrates an exemplary video surveillance system according to an exemplary embodiment of the invention.

[0020] Figure 2 illustrates an exemplary frame from a video stream from the video surveillance system according to an exemplary embodiment of the invention.

[0021] Figure 3A illustrates an exemplary flow diagram for target detection and counting according to an exemplary embodiment of the invention.

[0022] Figure 3B illustrates an example where several human models occupy a two-dimensional video image, each corresponding to a different location with respect to the two-dimensional video image

[0023] Figure 3C illustrates a single row of (x, y) identifying coordinates 321 each associated with a corresponding human model 320.

[0024] Figure 3D illustrates an exemplary method for calculating a human probability map.

What is claimed is:

1. A method of detecting human objects in a video, comprising:
determining pixels of a video image are foreground pixels, the group of foreground pixels constituting a foreground blob set of one or more foreground blobs;
for each of N locations within the video image, where N is an integer, comparing a predetermined shape with the foreground blob set to obtain a corresponding probability of a human at the location, thereby obtaining N probabilities corresponding to the N locations;
using the N probabilities, determining X humans are represented by the foreground blob set, where X is whole number.
2. The method of claim 1, further comprising using the probability map to determine a location of each of the X humans.
3. The method of claim 2, wherein the determined location of each of the X human is a location within an image plane corresponding to the video image.
4. The method of claim 2, wherein the determined location of each of the X human is a location with respect to a physical ground plane corresponding to the real world.
5. The method of claim 1,
wherein determining foreground pixels of the video image comprises comparison of a first frame of a video image without foreground objects with comparison of a second frame of the video image containing the foreground objects.
6. The method of claim 1, wherein the predetermined shape is the same for each of the N locations.
7. The method of claim 1, wherein the predetermined shape for at least some of the N locations has a different size.

8. The method of claim 7,
wherein a size of the predetermined shape for each of the N locations is determined in response to calibration of a video system, and
wherein the video system is used to obtain the video image.
9. The method of claim 8,
wherein the calibration of the video system comprises determining an image size of a portion of the video image corresponding to an average human size at each of the N locations, and
wherein the size of the predetermined shape for each of the N locations is determined in response to the corresponding image size.
10. The method of claim 1,
wherein, for each of the N locations, the corresponding predetermined shape comprises an estimate of a foreground image part to be occupied in the video image when a human exists at the corresponding location.
11. The method of claim 10, wherein the estimate of the foreground image part for each of the N locations is calculated based on a projection of a model of a human in the real world onto an image plane of the video image.
12. The method of claim 1,
wherein the video image comprises a plurality of image frames, each image frame comprising a two dimensional image having the N locations, each of the N locations identified by a corresponding x, y coordinate pair within the two dimensional image.
13. The method of claim 12,
wherein each of the N locations is associated with a corresponding predetermined shape with respect to an image plane corresponding to the video image.

14. The method of claim 1, further comprising, for each of the N locations, calculating a recall ratio of the corresponding predetermined shape and the foreground blob to determine an associated probability.

15. The method of claim 14, wherein for each of the N locations, calculating the recall ratio comprises determining a ratio of (a) an area comprising an overlap of an area occupied by the predetermined shape and the foreground blob and (b) an area of the foreground blob.

16. The method of claim 1, further comprising:
creating a probability map with the N probabilities;
determining local maximums of probabilities of the probability map.

17. The method of claim 16, further comprising:
selecting a first location of the N locations corresponding to a local maximum of the probability map;
obtaining a first predetermined shape corresponding to the first location;
calculating a first ratio of (a) an area comprising an overlap of an area occupied by the first predetermined shape and the foreground blob and (b) an area of the foreground blob; and
calculating a second ratio of (a) an area comprising an overlap of an area occupied by the first predetermined shape and the foreground blob and (b) a remaining area occupied by the first predetermined shape that is not overlapped with the foreground blob.

18. The method of claim 17, further comprising:
selecting a second location of the N locations corresponding to a local maximum of the probability map;
obtaining a second predetermined shape corresponding to the second location;
calculating a third ratio of (a) an area comprising an overlap of an area occupied by the second predetermined shape and the foreground blob and (b) an area of the foreground blob; and
calculating a fourth ratio of (a) an area comprising an overlap of an area occupied by the second predetermined shape and the foreground blob and (b) a remaining area occupied by the second predetermined shape that is not overlapped with the foreground blob.

19. The method of claim 18, wherein the first, second, third and fourth ratios are used to determine that X humans are represented by the foreground blob.

20. The method of claim 18, further comprising:
calculating a fifth ratio of (a) an overlap of an area occupied by the second predetermined shape and an area occupied by the first predetermined shape and (b) an area of the second predetermined shape.

21. The method of claim 16, further comprising calculating a precision value and a recall value for each of m locations of the N locations, m being an integer, each of the m locations corresponding to a local maximum of the probability map.

22. The method of claim 21, wherein each of the m locations are selected sequentially 1 to m, a selection of an (m-1)th location excluding selection of an mth location that falls within a first predetermined distance of the (m-1)th location.

23. The method of claim 22, wherein each of the m locations are selected sequentially 1 to m wherein the selection of a next location of the m locations comprises selecting a location that is closest to a bottom edge of the video image for those locations corresponding to a local maximum that have not been excluded.

24. A method of detecting human objects in a video, comprising:
determining pixels of a video image of a real world scene are foreground pixels, the group of foreground pixels constituting a foreground blob set of one or more foreground blobs;
for each of N locations within the video image, where N is an integer, comparing a predetermined shape with the foreground blob set to determine X humans are represented by the foreground blob set, where X is whole number and a location of each of the X humans is determined as a location within a horizontal plane of the real world.

25. The method of claim 24, further comprising detecting the existence of a crowd by reviewing at least some of the locations of the X humans.

26. The method of claim 24, further comprising determining an existence of a crowd when it is determined that Y of the X humans are located within a first area of the horizontal plane of the real world.

27. The method of claim 26, wherein the first area comprises a predetermined geometric shape having a predetermined area size within the real world.

28. The method of claim 26, wherein the first area comprises an area defined by a circle.

29. The method of claim 26, further comprising determining a crowd density within the first area.

30. The method of claim 29, further comprising comparing the crowd density to a threshold and sending at least one of a report and an alarm when the crowd density exceeds the threshold.

31. The method of claim 24, further comprising:
determining a first crowd density within a first area corresponding to a first frame of the video image;
determining a second crowd density within the first area corresponding to a second frame of the video image;
determining a crowd gathering event in response to the first crowd density and the second crowd density.

32. The method of claim 24, further comprising:
determining a first crowd density within a first area corresponding to a first frame of the video image;

determining a second crowd density within the first area corresponding to a second frame of the video image;

determining a crowd dispersing event in response to the first crowd density and the second crowd density.

key challenges of these approaches are higher computational cost, view-point dependent learning and relatively large human image size requirement.

[0010] The embodiments described here address some of these problems of existing systems.

SUMMARY

[0011] The disclosed embodiments provide methods, devices and systems for intelligent analysis of video images to detect objects, such as human objects.

[0011A] According to various embodiments, a method of detecting human objects in a video may be provided. The method of detecting human objects in a video may include: determining pixels of a video image are foreground pixels, a group of the foreground pixels constituting a foreground blob set of one or more foreground blobs; for each of N predetermined shapes at corresponding ones of N predetermined locations within the video image, where N is a positive integer greater than one, calculating a first value by comparing an overlap of the corresponding predetermined shape with the foreground blob set, the first value being used to obtain a corresponding probability of a human at the corresponding predetermined location, thereby obtaining N probabilities corresponding to the N locations; using the N probabilities, determining X humans are represented by the foreground blob set, where X is a whole number; and providing at least one of a report, an alarm, and an event detection using the determination of the representation of X humans, wherein a size of the corresponding predetermined shape for each of the N locations is determined in response to calibration of the video system, and wherein the comparing of the corresponding predetermined shape with the foreground blob set for each of the N predetermined shapes comprises analyzing an amount of an overlapping area of the corresponding predetermined shape with the foreground blob set.

[0011B] According to various embodiments, a method of detecting human objects in a video. The method of detecting human objects in a video may include: determining pixels of a video image of a real world scene are foreground pixels, a group of the foreground pixels constituting a foreground blob set of one or more foreground blobs; for each of N predetermined shapes at corresponding ones of N predetermined locations within the video image, where N is a positive integer greater than one, calculating a first value by comparing an overlap of the corresponding predetermined shape with the foreground blob set, the first value being used to determine X humans are represented by the foreground blob set, where X is whole number and a location of each of the X humans is determined as a location within a horizontal plane of the real world; and providing at least one of a report, an alarm, and an event detection when the crowd density exceeds a threshold value, using the determination of the representation of X humans, wherein a size of the corresponding predetermined shape for each of the N locations is determined in response to calibration of the video system, wherein the video system is used to obtain the video image, wherein the comparing of the corresponding predetermined shape with the foreground blob for each of the N predetermined shapes comprises analyzing an amount of an overlapping area of the corresponding predetermined shape with the foreground blob set.

[0011C] According to various embodiments, a video surveillance system may be provided. The video surveillance system may include: a video source configured to provide a video image of a real world scene; and a computer, configured to detect foreground pixels of the video image, a group of the foreground pixels constituting a foreground blob set of one or more foreground blobs, wherein for each of N predetermined shapes at corresponding ones of N predetermined locations within the video image, where N is a positive integer greater than one, configured to calculate a first value by comparing an overlap of the corresponding predetermined shape with the foreground blob set, configured to determine X humans are

represented by the foreground blob set using the first value, where X is whole number, configured to compare the corresponding predetermined shape with the foreground blob for each of the N predetermined shapes by analyzing an amount of an overlapping area of the corresponding predetermined shape with the foreground blob set and configured to provide at least one of a report, an alarm, and an event detection when the crowd density exceeds a threshold value, using the determined representation of X humans.

[0012] In an example, a method of detecting human objects in a video comprises determining that certain pixels of a video image are foreground pixels, the group of foreground pixels constituting a foreground blob set of one or more foreground blobs; for each of N locations within the video image, where N is an integer, comparing a predetermined shape with the foreground blob set to obtain a corresponding probability of a human at the location, thereby obtaining N probabilities corresponding to the N locations; and using the N probabilities, determining X humans are represented by the foreground blob set, where X is whole number.

[0013] In an example, a method of detecting human objects in a video, may comprise determining pixels of a video image of a real world scene are foreground pixels, the group of foreground pixels constituting a foreground blob set of one or more foreground blobs; and for each of N locations within the video image, where N is an integer, comparing a predetermined shape with the foreground blob set to determine X humans are represented by the foreground blob set, where X is whole number.

[0014] In an example, methods may include determining a location of each of the X humans. The locations of each of the X humans may be determined as a location within a horizontal plane of the real world, such as a location on a physical ground plane of the real world.

[0015] The detection of the human objects may be used to count humans, for crowd analyses and for other event detections.

[0016] System and devices are disclosed which may be configured to perform such methods.

[0017] Computer readable media containing software that may be used to configure a computer to perform the operations described herein and comprise further examples.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] Example embodiments will be more clearly understood from the following detailed description taken in conjunction with the accompanying drawings. The figures represent non-limiting example embodiments as described herein.

[0019] Figure 1 illustrates an exemplary video surveillance system according to an exemplary embodiment of the invention.

[0020] Figure 2 illustrates an exemplary frame from a video stream from the video surveillance system according to an exemplary embodiment of the invention.

[0021] Figure 3A illustrates an exemplary flow diagram for target detection and counting according to an exemplary embodiment of the invention.

[0022] Figure 3B illustrates an example where several human models occupy a two-dimensional video image, each corresponding to a different location with respect to the two-dimensional video image

[0023] Figure 3C illustrates a single row of (x, y) identifying coordinates 321 each associated with a corresponding human model 320.

[0024] Figure 3D illustrates an exemplary method for calculating a human probability map.

What is claimed is:

1. A method of detecting human objects in a video, comprising:
determining pixels of a video image are foreground pixels, a group of the foreground pixels constituting a foreground blob set of one or more foreground blobs;
for each of N predetermined shapes at corresponding ones of N predetermined locations within the video image, where N is a positive integer greater than one, calculating a first value by comparing an overlap of the corresponding predetermined shape with the foreground blob set, the first value being used to obtain a corresponding probability of a human at the corresponding predetermined location, thereby obtaining N probabilities corresponding to the N locations;
using the N probabilities, determining X humans are represented by the foreground blob set, where X is a whole number; and
providing at least one of a report, an alarm, and an event detection using the determination of the representation of X humans,
wherein a size of the corresponding predetermined shape for each of the N locations is determined in response to calibration of the video system, and
wherein the comparing of the corresponding predetermined shape with the foreground blob set for each of the N predetermined shapes comprises analyzing an amount of an overlapping area of the corresponding predetermined shape with the foreground blob set.
2. The method of claim 1, further comprising using the N probabilities to determine a location of each of the X humans.
3. The method of claim 2, wherein the determined location of each of the X humans is a location within an image plane corresponding to the video image.
4. The method of claim 2, wherein the determined location of each of the X humans is a location with respect to a physical ground plane corresponding to the real world.
5. The method of claim 1,

wherein determining foreground pixels of the video image comprises comparison of a first frame of a video image without foreground objects with comparison of a second frame of the video image containing the foreground objects.

6. The method of claim 1, wherein the predetermined shape is the same for each of the N locations.

7. The method of claim 1, wherein the predetermined shape for at least some of the N locations has a different size.

8. The method of claim 1,
wherein the calibration of the video system comprises determining an image size of a portion of the video image corresponding to an average human size at each of the N locations, and

wherein the size of the predetermined shape for each of the N locations is determined in response to the corresponding image size.

9. The method of claim 1,
further comprising, prior to the determination of the pixels of the video image that are foreground pixels, for each of the N locations, determining the corresponding predetermined shape by estimating a foreground image part to be occupied in the video image when a human exists at the corresponding location.

10. The method of claim 9, wherein estimating the foreground image part for each of the N locations is based on a projection of a model of a human in the real world onto an image plane of the video image.

11. The method of claim 1,
wherein the video image comprises a plurality of image frames, each image frame comprising a two dimensional image having the N locations, each of the N locations identified by a corresponding x, y coordinate pair within the two dimensional image.

12. The method of claim 11,

wherein each of the N locations is associated with a corresponding one of the N predetermined shapes with respect to an image plane corresponding to the video image.

13. The method of claim 1, further comprising, for each of the N locations, calculating a recall ratio of the corresponding predetermined shape and the foreground blob to determine an associated probability.

14. The method of claim 13, wherein for each of the N locations, calculating the recall ratio comprises determining a ratio of (a) an area comprising an overlap of an area occupied by the predetermined shape and the foreground blob and (b) an area of the foreground blob.

15. The method of claim 1, further comprising:
creating a probability map with the N probabilities; and
determining local maximums of probabilities of the probability map.

16. The method of claim 15, further comprising:
selecting a first location of the N locations corresponding to a local maximum of the probability map;
obtaining a first predetermined shape corresponding to the first location; and
analyzing an amount of an overlap of an area occupied by the first predetermined shape and the foreground blob.

17. The method of claim 16, further comprising:
calculating a first ratio of (a) an area comprising an overlap of an area occupied by the first predetermined shape and the foreground blob and (b) an area of the foreground blob,

wherein the first ratio is used to determine that X humans are represented by the foreground blob.

18. The method of claim 15, further comprising calculating a precision value and a recall value for each of m locations of the N locations, m being an integer, each of the m locations corresponding to a local maximum of the probability map.

19. The method of claim 1, further comprising:
selecting a subset of the N predetermined shapes based on the N probabilities;
and
analyzing an overlap of an area occupied by selected subset of N predetermined shapes and an area occupied by foreground blob.
20. The method of claim 1, further comprising, sequentially selecting 1 to m ones of the N locations, a selection of an (m-1)th location excluding selection of a subsequent one of the N locations that falls within a first predetermined distance of the (m-1)th location.
21. The method of claim 1, further comprising sequentially selecting 1 to m ones of the N locations, wherein the selection of a next location of the N locations comprises selecting a location based on its proximity to a bottom edge of the video image.
22. A method of detecting human objects in a video, comprising:
determining pixels of a video image of a real world scene are foreground pixels, a group of the foreground pixels constituting a foreground blob set of one or more foreground blobs;
for each of N predetermined shapes at corresponding ones of N predetermined locations within the video image, where N is a positive integer greater than one, calculating a first value by comparing an overlap of the corresponding predetermined shape with the foreground blob set, the first value being used to determine X humans are represented by the foreground blob set, where X is whole number and a location of each of the X humans is determined as a location within a horizontal plane of the real world; and
providing at least one of a report, an alarm, and an event detection when the crowd density exceeds a threshold value, using the determination of the representation of X humans,
wherein a size of the corresponding predetermined shape for each of the N locations is determined in response to calibration of the video system,
wherein the video system is used to obtain the video image,
wherein the comparing of the corresponding predetermined shape with the foreground blob for each of the N predetermined shapes comprises analyzing an amount of an overlapping area of the corresponding predetermined shape with the foreground blob set.

23. The method of claim 22, further comprising detecting the existence of a crowd by reviewing at least some of the locations of the X humans.
24. The method of claim 22, further comprising determining an existence of a crowd when it is determined that Y of the X humans are located within an area of the horizontal plane of the real world.
25. The method of claim 24, wherein the area of the horizontal plane of the real world comprises a predetermined geometric shape having a predetermined area size within the real world.
26. The method of claim 24, wherein the area of the horizontal plane of the real world comprises an area defined by a circle.
27. The method of claim 22, further comprising:
determining a first crowd density within a first area corresponding to a first frame of the video image;
determining a second crowd density within the first area corresponding to a second frame of the video image;
determining a crowd gathering event in response to the first crowd density and the second crowd density.
28. The method of claim 22, further comprising:
determining a first crowd density within a first area corresponding to a first frame of the video image;
determining a second crowd density within the first area corresponding to a second frame of the video image;
determining a crowd dispersing event in response to the first crowd density and the second crowd density.
29. A video surveillance system, comprising:
a video source configured to provide a video image of a real world scene; and
a computer, configured to detect foreground pixels of the video image, a group of the foreground pixels constituting a foreground blob set of one or more foreground blobs,

wherein for each of N predetermined shapes at corresponding ones of N predetermined locations within the video image, where N is a positive integer greater than one, configured to calculate a first value by comparing an overlap of the corresponding predetermined shape with the foreground blob set, configured to determine X humans are represented by the foreground blob set using the first value, where X is whole number, configured to compare the corresponding predetermined shape with the foreground blob for each of the N predetermined shapes by analyzing an amount of an overlapping area of the corresponding predetermined shape with the foreground blob set and configured to provide at least one of a report, an alarm, and an event detection when the crowd density exceeds a threshold value, using the determined representation of X humans.