(12) UK Patent (19) GB (11) 2555331 (13) B

(45) Date of B Publication 27.10.2021

(54) Title of the Invention: IMAGE PROCESSING METHOD

(51) INT CL: ***H04N 7/18*** (2006.01) ***G06K 9/00*** (2006.01) ***G06K 9/32*** (2006.01) ***G06T 7/00*** (2017.01)
G06T 7/20 (2017.01) ***H04N 5/14*** (2006.01) ***H04N 19/115*** (2014.01) ***H04N 19/117*** (2014.01)
H04N 19/124 (2014.01) ***H04N 19/137*** (2014.01) ***H04N 19/142*** (2014.01) ***H04N 19/167*** (2014.01)
H04N 19/503 (2014.01) ***H04N 19/87*** (2014.01)

(21) Application No: **1720694.7**

(22) Date of Filing: **12.05.2016**

Date Lodged: **12.12.2017**

(30) Priority Data:
(31) **1508074** (32) **12.05.2015** (33) **GB**

(86) International Application Data:
PCT/GB2016/051362 En 12.05.2016

(87) International Publication Data:
WO2016/181150 En 17.11.2016

(43) Date of Reproduction by UK Office **25.04.2018**

(72) Inventor(s):
Ilya Romanenko
Michael Tusch

(73) Proprietor(s):
Apical Ltd.
(Incorporated in the United Kingdom)
110 Fulbourn Road, Cambridge, Cambridgeshire, CB1 9NJ, United Kingdom

(74) Agent and/or Address for Service:
EIP
Fairfax House, 15 Fulwood Place, LONDON, WC1V 6HU, United Kingdom

(56) Documents Cited:
WO 2012/033758 A2 **WO 1999/053686 A1**
US 20120062732 A1 **US 20100119157 A1**
US 20080013787 A1 **US 20070223812 A1**

(58) Field of Search:
As for published application 2555331 A viz:
INT CL **G06K, G06T, H04N**
Other: **EPO-Internal, WPI Data**
updated as appropriate

Additional Fields
INT CL **G06T, H04N**
Other: **WPI, EPODOC**

GB 2555331 B

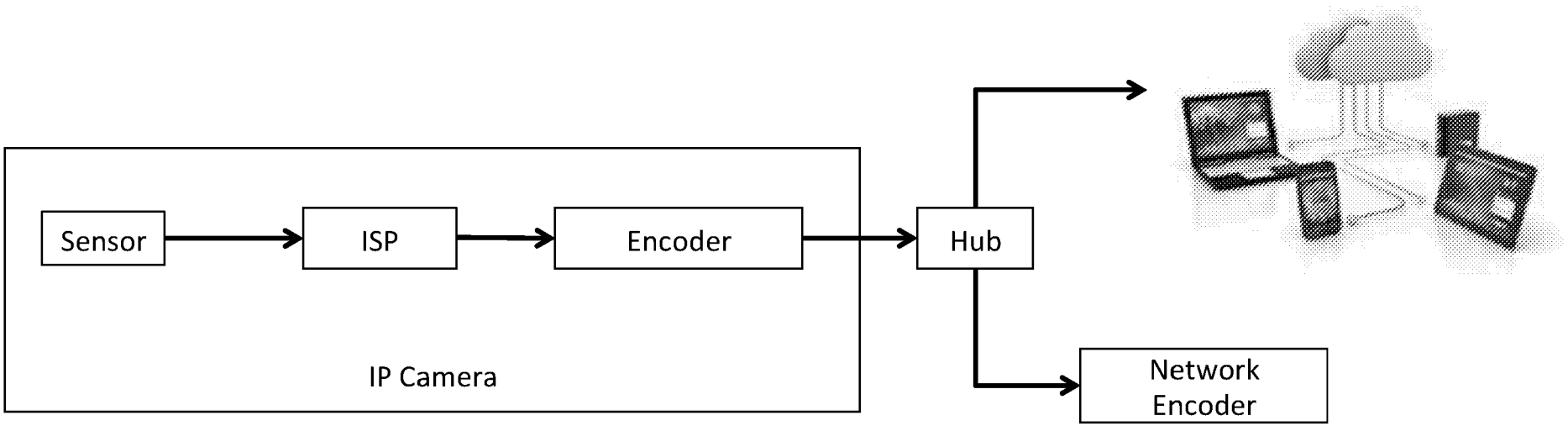

FIGURE 1 (Prior art)

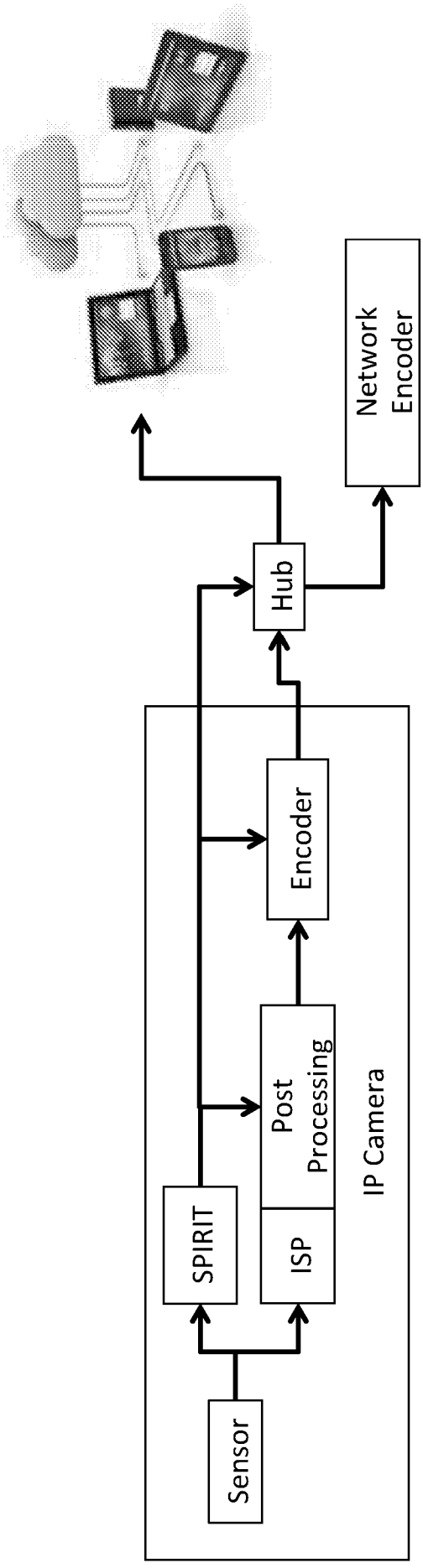
Network Encoder
Hub
Encoder
Post Processing
ISP
SPIRIT
Sensor
IP Camera

FIGURE 2

FIGURE 3

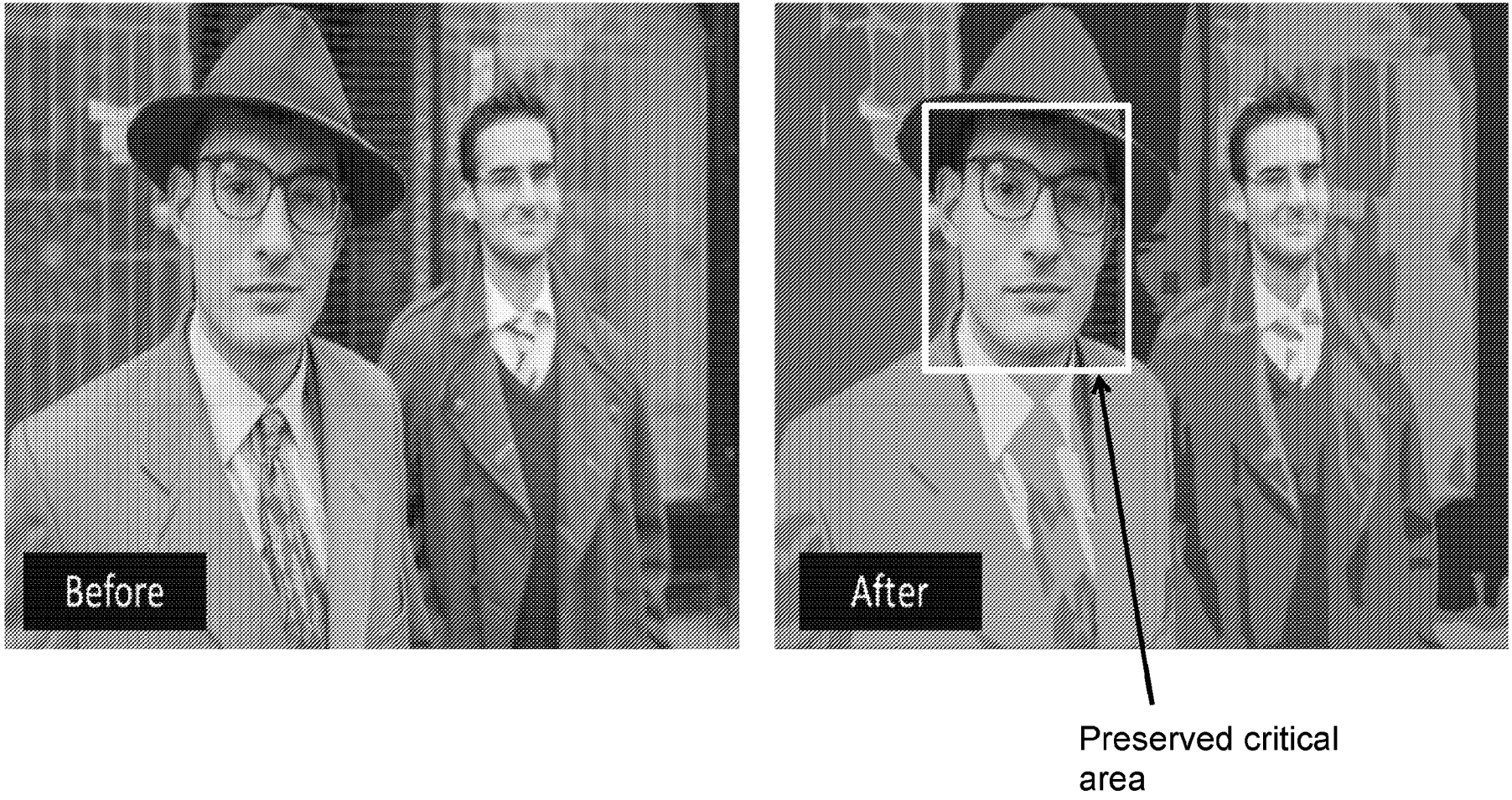

FIGURE 4

IMAGE PROCESSING METHOD

BACKGROUND OF THE INVENTION

1. Field of the Invention

The field of the invention relates to method of image processing and to related systems, devices and computer program products.

2. Technical Background

The introduction and proliferation of low cost, high resolution video cameras are major factors in the rapid growth of the home security market. Typical installations now combine multiple Internet Protocol (IP) cameras in and around the home with both video recording solutions and with internet based "home monitoring" services.

However, despite modern techniques of video compression, even homes with multi-gigabit internet connections can easily exhaust their total bandwidth with even a modest number of cameras, perhaps even just one. In the light of growing resolution of ip cameras, i.e. switching to 4k standard, the issue of bandwidth does not seem to be resolved in a nearest future only by improvements in video encoding techniques.

There is no doubt that such systems can reassure people that their home is safe whilst they are away, and web based monitoring services can text or email when "alarms" are triggered, such as when motion is detected. However, the high incidence of false alarms, and the wasteful use of network bandwidth to, say, connect your camera video phone to

your phone app and transmit still and video images, severely limits the useful application of these systems.

3. Discussion of Related Art

Figure 1 shows a diagram of typical current systems, in which an IP camera comprises an image sensor, an Image Signal Processor (ISP) and an encoder. Such systems necessarily stream compressed video, primarily around the home network to a storage device, essentially a video recorder. Additionally, some systems may "broadcast" video either directly to a device connected directly via the Internet, or through a web based service – perhaps providing some formatting operations such as scaling the video to a resolution more suited to display on a mobile device.

Such systems may use methods to limit the "on" time of the video, such as using PIR motion detectors to trigger recording. However, such systems are both prone to error, necessarily needing the decision to be "over prudent", and the result remains a hard decision; to send video or not.

Additionally, there is a need for a method to reduce the bandwidth requirement through the transmission and at the storage end while preserving critical information in the image or video captured, while at the same time recording certain regions in the image at best possible quality.

SUMMARY OF THE INVENTION

The invention is an image processing method comprising the steps of acquiring a video, using an object detection engine to detect a person in the video, the object detection engine being integrated with an image signal processing pipeline, transmitting the video over a network, determining that the detected person has moved less than a pre-set distance; and responsive to the determining, pausing transmission of the video.

Optional features include any one or more of the following:

- the object detection engine is integrated with the image signal processing pipeline in an embedded image processor.
- the object detection engine receives raw digital video from an image sensor and analyses the raw digital video on a frame by frame basis.
- the object detection engine outputs location data or co-ordinates of one or more persons, their poses, body parts such torso, upper bodies, faces, hands, legs at each video frame.
- extracting a thumbnail image, such as a person's face, and sending a text alert with the thumbnail to an end user, wherein the end user can choose to access a live video feed.
- detecting an "allowed person" based on pre-defined rules, and not sending an alert to the end user.

Another aspect is an embedded image processor configured to perform a method in accordance with the method defined above. The object detection engine and image signal processing pipeline may be integrated with or form part of a GPU (graphics processing unit) or a SoC (system on chip).

Note that the term 'video' used in this specification includes still images.

BRIEF DESCRIPTION OF THE FIGURES

Aspects of the invention will now be described, by way of example(s), with reference to the following Figures, in which:

Figure 1 shows a typical current system (prior art).

Figure 2 shows the system architecture of the present invention.

Figure 3 shows an example of a system for people detection in which a security camera is placed outside an entry door.

Figure 4 shows another example of a system for people detection.

DETAILED DESCRIPTION

The present invention addresses the limitations of current systems through a novel use of people or object detection results, avoiding false alarms and providing a means to efficiently manage data bandwidth. The invention provides a system enabling an efficient use of communication bandwidth and storage.

A video or still camera is provided including an object detection engine for detecting a predefined class of object in an image and for manipulating the image to preserve detail and information necessary to the given application of the camera, whilst reducing the overall bandwidth to transmit, or capacity to store, that image or video.

Integrating an object detection engine with the image signal processing (ISP) pipeline of the camera, that is the processing that is done to convert the output of a CMOS image sensor into a recognizable picture, offers many advantages. The output of an object detection engine is used to control the parameters of the encoder.

ISPs typically already include scalers to crop and resize images, filters to blur images, and noise reduction technologies that can be applied more or less significantly to areas of the image. However, it is also possible to use the results of the people detection engine, such as "annotation data" to control the parameters of the encoder.

Figure 2 shows a diagram of the system's architecture. An IP camera comprises an image sensor for capturing an image data, an object detection engine (SPIRIT) for detecting an object within the captured image data, an image signal processor including a post processing module for processing the captured image by taking into account the object detection engine output, and an encoder for compressing the processed image by taking into account the detection engine output. The system adopts advanced detection techniques not only as a trigger, but also as a means to modify and manage the video transition bandwidth (not excluding still images) in a graceful, predictable manner, without the loss of critical information. The proposed system may for example employ a people detection engine, in this case referred to as the "SPIRIT" engine, to extract the coordinates of people and their faces within each video frame. Since the detection is

done within the camera, the video can be manipulated prior to transmission. In fact the video may be modified, such as scaled or cropped, even before compression.

One implementation of the invention is a video or still camera including a means to detect a predefined class of object in an image and manipulate that image to preserve detail and information necessary for a specific application, whilst reducing the overall bandwidth needed to transmit image or video, or the capacity to store that image or video.

Optional implementations may include the following features, alone or in combination:

- Image or video data sent to an encoder may be modified according to one or more of the following:
    - determine that whilst people have been detected, but have not moved more than a pre-set distance, then pause the output, thereby exploiting the temporal compression of the encoder; (This results in on the fly frame rate control);
    - zoom to the region of interest, for instance a person;
    - segmentation of scene to separate people from background;
    - blur / low-pass filter / pastelize… remove detail of background;
    - preserve detail of faces (including resolution, texture, colour);
    - enhance detail – application of localized gamma correction, contrast enhancement, sharpening… of critical areas like faces.
- Modifying the behavior of the encoder may include for example:
    - compression ratio;
    - quantization;
    - output resolution and/or frame rate.
- The system may be augmented by a means to monitor the usage of bandwidth, and further to adopt more or less aggressive compression techniques as described above.
- The system may include a means for the end user (consumer) of the video to request greater/lesser compression, remotely.
- The system may include a means for the end user (consumer) of the video to alter the compression strategy.

- The system may be augmented by a means to define different regions of interest in a scene, such that greater bandwidth may be dedicated to video relating to that region (note: for instance, higher bandwidth at front door, lower for further away).
- The system may comprise different strategies for compression based on the scene, including zooming in to a person's face when they are the only person in the scene, and applying greater background blurring when there are multiple people and the field of view is preserved.
- The system may extract thumbnail images, say of faces, and broadcast/record these in favour of video. This includes sending a text alert – so user can see a thumbnail, if they recognize the person then do nothing, but if not they can click on it to go to the live video feed.
- The system may determine if someone is approaching or leaving, and may include rules to ignore people leaving.
- The system may recognize people, including "rules" to define them as an "allowed" person – so no need to send anything.
- The system may send annotation data (or subset thereof) to a central hub to allow the aggregation of detections.
- A multi-camera system may be augmented with a web-based service that can take the compressed/cropped/thumbnails from several cameras and construct (on the fly) a single multi-camera view to transmit to the user (saving bandwidth / data usage) on the receiving device.
- A multi-camera system /hub may include rules to be defined to trigger recording or streaming of video based on events detected on multiple cameras (like – camera 1 saw someone in the yard, camera 2 saw them approach the back door…)
- The system may include an application/program/browser, characterized in that the received video is interspersed with / includes (or through any other channel) data to allow an "always on" view, which may be live video, periodic low frame rate vide, stills or textual/graphic information.
- The "compressed video" may be replaced altogether by a symbolic/graphical representation of the scene, including sending this as data to be rendered by a "player" (so this is not video!).

- The symbolic/graphical representation of the scene may also add real faces to avatars.
- The system may be performed in real-time.
- Modifying or altering the video is performed in 3D.

Use case examples:

Figure 3 shows an example of a system for people detection in which a security camera is placed outside an entry door. The system prevents false alarms due to passing traffic, such as a car present in background. Image A shows the original video captured by the camera. In order to save bandwidth, the image may be cropped. Cropping the image also further reduces the likelihood of any motion in the scene such as passers by being included in the scene and hence may also reduce the number of false alarms. Other example of cropped background information also includes plants blowing in the wind or flapping leaves in the background. The system therefore enables to take advantage of the motion estimation based compression in the encoder. Image C shows a processed image to be recorded or broadcasted, in which the person's head has been segmented from the background scene, allowing detail to be preserved in critical areas, such as the person's face. The rest of the image (background) has been blurred, allowing Image C to be more efficiently compressed.

Figure 4 shows another example of a system for people detection and demonstrates how the proposed techniques, whilst dramatically altering the image to vastly increase the achievable compression, still retains the "useful" information required by the observer. **Figure 4** shows a "before" image as captured by the image sensor and an "after" image that has been processed through the system. The "after" image shows the critical area where details have been preserved while the rest of the image has been blurred. In this case, the system has recognised a person of interest and the critical area corresponds to the area of interest corresponding the recognised person's face. The video may also be altered such that a crop is used and focuses on the recognised person's face. In another example, another crop may be focused on many person of interest.

Note

It is to be understood that the above-referenced arrangements are only illustrative of the application for the principles of the present invention. Numerous modifications and alternative arrangements can be devised without departing from the spirit and scope of the present invention. While the present invention has been shown in the drawings and fully described above with particularity and detail in connection with what is presently deemed to be the most practical and preferred example(s) of the invention, it will be apparent to those of ordinary skill in the art that numerous modifications can be made without departing from the principles and concepts of the invention as set forth herein.

CLAIMS

1. An image processing method comprising the steps of acquiring a video and then altering the video using an object detection engine integrated with an image signal processing pipeline prior to transmission of the video, in order to modify the bandwidth needed to transmit the modified video.

2. The method of Claim 1 in which the object detection engine is integrated with the image signal processing pipeline in an embedded image processor.

3. The method of any preceding Claim, in which the object detection engine receives raw digital video from an image sensor and analyses the raw digital video on a frame by frame basis.

4. The method of any preceding Claim, in which the object detection engine is used to detect a predefined class of object, such as a person.

5. The method of Claim 4, in which the object detection engine outputs location data or co-ordinates of one or more persons, their poses and body parts, such as torso, hands, legs and faces at each video frame.

6. The method of Claim 4-5, in which the video transmission is paused when it is determined that detected person(s) have not moved more than a pre-set distance.

7. The method of Claim 4-5, in which detailed object information is used to precisely model an object's avatar to allow the object to be separated from the background more precisely.

8. The method of any preceding Claim, in which altering the video involves one or more of the following:

- zooming or cropping into a region of interest or an object of interest,
- segmenting the video to separate certain object(s) from the background,
- blurring to remove background details,
- using a low pass-filter to remove certain areas, such as background details,

- pastelizing to remove certain areas, such as background details,
- preserving details such as resolution, texture or colour of a region of interest or an object of interest,
- enhancing details, such as applying localized gamma correction, enhancing contrast or sharpening on a region of interest or an object of interest.

9. The method of any preceding Claim, in which altering the video is performed prior to transmitting the video to an encoder that is internal to the camera that recorded the video.

10. The method of Claim 9, in which the encoder compresses the altered video by taking into account the output(s) of the object detection engine.

11. The method of Claim 10, in which the encoder compresses the altered video on a frame by frame basis by modifying one or more of the following: compression ratio, quantization, output resolution or frame rate.

12. The method of Claim 9-11 in which bandwidth usage is monitored and compression techniques of the encoder are adapted according to the bandwidth usage.

13. The method of Claim 9 in which an end user is able to modify compression techniques of the encoder remotely.

14. The method of any preceding Claim, in which one or more regions of interest in a scene are pre-defined, and bandwidth usage is also pre-defined specifically for each region of interest.

15. The method of any preceding Claim, in which a composition or compression target is pre-defined based on a specific scene.

16. The method of Claim 15, in which the video composition or compression target involves zooming to a person's face when they are the only person in the scene and

applying background blurring when there are multiple people and the field of view is preserved.

17. The method of any preceding Claim including the step of extracting thumbnail images, such as of a person's face, and broadcasting or recording the thumbnail images instead of video.

18. The method of any preceding Claim including the step of extracting a thumbnail image, such as a person's face, and sending a text alert with the thumbnail to an end user, wherein the end user can choose to access a live video feed.

19. The method of any preceding Claim including the step of detecting if a person is approaching or leaving, and in which any detected person who is leaving is ignored from a video composition or compression target.

20. The method of Claim 18 including the step of detecting an "allowed person" and not sending an alert to the end user.

21. The method of any preceding Claim including the step of sending annotation data (or subset thereof) to a central hub for aggregating the annotation data from the detections.

22. The method of any preceding Claim in which the video transmitted is interspersed with or includes data to allow an "always on" view, wherein data is one or more of the following: live video, periodic low frame rate video, stills or textual/graphic information.

23. The method of any preceding Claim in which a symbolic/graphical representation of a scene is transmitted instead of video.

24. The method of Claim 23 in which the symbolic/graphical representation of a scene includes avatars.

25. The method of any preceding Claim in which the method is applied to a multiple camera system, and in which the altered videos from the multiple camera system are used to reconstruct a single multiple camera view that is transmitted to an end-user.

26. The method of Claim 25 including the step of defining a set of rules to trigger recording or streaming of video based on events detected on the multiple camera system.

27. A computer vision system comprising

(e) an image sensor to capture a raw digital video in a visual, or infra-red part of the spectrum,;

(f) an object detection engine for detecting an object within the captured video and for extracting information on the captured video, the object detection engine integrated with an image signal processing pipeline;

(g) an image signal processor to alter the captured video by taking into account the information extracted by the object detection engine;

(h) an encoder for compressing the altered video by taking into account the information extracted by the object detection engine.

28. The computer vision system of Claim 27, in which the computer vision system processes video using the method of any preceding Claim 1-26.

29. A computer vision system of Claim 27-28, where one or more computer vision system, are embedded in one of the following:

- IP Camera or webcam;
- Smartphone;
- Surveillance computer vision system;
- Autonomous vehicle computer vision system;
- Robotics computer vision system.

30. An embedded image processor configured for acquiring a video and then altering the video, using an object detection engine integrated with an image signal processing pipeline, prior to transmission of the video, in order to modify the bandwidth needed to transmit the modified video or the storage needed to store the video.

31. The embedded image processor of Claim 30, operable to process video using the method of any preceding Claim 1-26.

32. The embedded image processor of Claim 31 in which the object detection engine and image signal processing pipeline are integrated with or form part of a GPU (graphics processing unit) or a SoC (system on chip).