CLAIMS

1. A computer-implemented method of operating a video conference endpoint, the video conference endpoint including a video camera which captures images showing a field of view, wherein the method comprises:

receiving data defining of a spatial boundary within the field of view, the spatial boundary being at least in part defined by a distance from the video camera;

capturing an image of the field of view;

identifying one or more persons within the field of view of the video camera;

estimating a position for each person identified within the field of view of the video camera;

determining, from the estimated position of each person identified within the field of view of the video camera, which of the one or more identified persons are within the spatial boundary, and

generating one or more video signals for transmission to a receiver,

wherein the one or more video signals include one or more cropped regions corresponding to one or more persons determined to be within the spatial boundary, and

when one or more of the identified persons within the field of view of the video camera are determined to be not within the spatial boundary, the one or more cropped regions exclude the one or more persons determined to not be within the spatial boundary.

2. The computer-implemented method of claim 1, wherein generating the one or more video signals includes:

determining, from the estimated position(s), that at least one of the one or more persons is within the spatial boundary; and

framing the one or more persons who are determined to be within the spatial boundary to produce respective cropped regions.

3. The computer-implemented method of any preceding claim including a step of transmitting the or each video signal to the receiver.

4. The computer-implemented method of any preceding claim, further comprising a validation mode of:

labelling each person in the image within the field of view of the video camera

according to whether they are within or outside of the spatial boundary; and

presenting to a user the labelled image for validation.

5. The computer-implemented method of any preceding claim, wherein estimating the position of the or each person is performed by measuring a distance between one or more pairs of facial landmarks for the respective person.

6. The computer-implemented method of claim 5, wherein distances between a plurality of pairs of facial landmark features are measured, each distance is used to estimate a distance of the person from the video camera, and a maximum and/or minimum estimated distance among the estimated distances is used to estimate the position of the or each person.

7. The computer-implemented method of claim 5 or 6, wherein estimating the distance includes estimating an orientation of the face of the person relative to the camera, and selecting the pairs of facial landmarks used to estimate the position based on the estimated orientation.

8. The computer-implemented method of any preceding claim, wherein estimating the position of the or each person includes estimating an orientation of the camera using one or more accelerometers within the video conference endpoint.

9. The computer-implemented method of any preceding claim, wherein estimating the position of the or each person includes the use of one or more distance sensors within the video conference endpoint.

10. The computer-implemented method of any preceding claim, wherein the spatial boundary is further defined at least in part by an angular extent of the captured image.

11. The computer-implemented method of any preceding claim, wherein the method includes a user input step, in which a user provides the data defining the spatial boundary.

12. The computer-implemented method of claim 11, wherein the user provides the data via a user interface.

13. The computer-implemented method of claim 11, wherein the user provides the data by entering the video conference endpoint into a data entry mode, in which video conference

endpoint tracks the location of the user, and the user prompts the video conference endpoint to use one or more locations of the user to define the spatial boundary.

14. A video conference endpoint including a video camera configured to capture an image showing a field of view, and a processor, wherein the processor is configured to:

receive data defining a spatial boundary within the field of view, the spatial boundary being at least in part defined by a distance from the video camera;

obtain an image of the field of view from the video camera;

identify one or more persons within the field of view of the video camera;

estimate a position of the or each person within the field of view of the video camera; and

generate one or more video signals, which include one or more cropped regions corresponding to one or more persons determined to be within the spatial boundary, for transmission to a receiver.

15. The video conference endpoint of claim 14, wherein generating the one or more video signals includes:

determining, from the estimated position(s), that at least one of the one or more persons is within the spatial boundary; and

framing the one or more persons who are determined to be within the spatial boundary to produce respective cropped regions.

16. The video conference endpoint of claim 14 or 15, wherein the video conference endpoint is connected to a receiver via a network, and the processor is configured to transmit the one or more video signals to the receiver.

17. The video conference endpoint of any of claims 14 - 16, wherein the processor is configured to perform a validation mode of:

labelling each person in the image within the field of view of the camera according to whether they are within or outside of the spatial boundary; and

presenting to a user the labelled image for validation.

18. The video conference endpoint any of claims 14 – 17, wherein the processor is configured to estimate the position of the or each person by measuring a distance between one or more pairs of facial landmark features of the respective person.

19. The video conference endpoint of any of claims 18, wherein the processor is configured to measure a plurality of distances between a plurality of pairs of facial landmark features, to estimate a distance of the person from the video camera using each of the measured distances, and use a maximum and/or minimum estimated distance among the estimated distances to estimate the position of the or each person.