CLAIMS

1. A video conferencing system, comprising:

at least one camera, each configured to generate an overview video output stream representative of a meeting environment; and

at least one video processing unit configured to:

automatically analyze the at least one overview video output stream to identify at least a first meeting participant and a second meeting participant represented in the at least one overview video output stream;

assign a first role designation to the first meeting participant and a second role designation to the second meeting participant based on one or more characteristics associated with at least one of the first meeting participant or the second meeting participant determined based on analysis of the at least one overview video output stream;

determine, based on priorities associated with the first and second role designations, relative framing priorities associated with the first meeting participant and the second meeting participant; and

generate as an output, based on the relative framing priorities, one or more primary video streams featuring the first meeting participant and/or the second meeting participant;

wherein the at least one video processing unit is configured to periodically update the roles assigned to each of the first meeting participant and the second meeting participant based on analysis of the at least one overview video output stream.

2. The system of claim 1, wherein generating as an output, based on the relative framing priorities, one or more primary video streams featuring the first meeting participant and/or the second meeting participant comprises:

selecting, based on the relative framing priorities, a framed representation of the first meeting participant or a framed representation of the second meeting participant; and

generating, as an output, a primary video stream featuring the selected framed representation of the first meeting participant or the second meeting participant.

3. The system of claim 2, wherein the video processing unit is configured to select a framed representation of the meeting participant associated with the higher relative framing priority.

4. The system of claim 2 or claim 3, wherein the selected framed representation is of the second meeting participant, and wherein the at least one video processing unit is configured to:

determine, based on the first role designation assigned to the first meeting participant, a likelihood that the first meeting participant will begin to speak;

perform, in response to the likelihood that the first meeting participant will begin to speak meeting a likeliness condition, one or more actions to prepare the video processing unit to output a primary video stream featuring a framed representation of the first meeting participant;

determine, based on analysis of the at least one overview video output stream, whether the first meeting participant begins to speak; and

in response to a determination that the first meeting participant begins to speak, outputting for display the primary video stream including the framed representation of the first meeting participant.

5. The system of claim 4, wherein performing one or more actions to prepare the video processing unit to output a primary video stream featuring the first meeting participant comprises one or more of:

selecting the framed representation of the first meeting participant for the primary video stream; and

generating the primary video stream featuring the framed representation of the first meeting participant.

6. The system of claim 4 or claim 5, wherein the determination of the likelihood that the first meeting participant will begin to speak is further based on one or more determined

quantities characterizing conversational dynamics between the first meeting participant and one or more other meeting participants featured in the one or more video output streams.

7. The system of claim 6, wherein the one or more determined quantities include one or more determined vectors representing one or more conversational exchanges between the first meeting participant and one or more other meeting participants featured in the one or more video output streams.

8. The system of claim 7, wherein each vector represents a speaking episode during which the first meeting participant or one of the one or more other meeting participants is determined to have spoken.

9. The system of claim 8, wherein the magnitude of each vector represents a duration of the speaking episode, and the direction of each vector represents the meeting participant that speaks next.

10. The system of any one of claims 6 to 9, wherein the one or more determined quantities include at least one of:

    the identities of the one or more other meeting participants;

    one or more durations characterizing one or more episodes during which the first meeting participant or one or more other meeting participants is determined to have spoken;

    one or more frequencies characterizing the frequency of episodes during which the first meeting participant or one or more other meeting participants is determined to have spoken;

    one or more periods characterizing one or more gaps between episodes during which the first meeting participant or one or more other meeting participants is determined to have spoken.

11. The system of any one of claims 6 to 10, wherein the determination of the likelihood that the first meeting participant will begin to speak is further based on one or more determined attributes associated with the first meeting participant.

12. The system of claim 11, wherein the one or more determined attributes include at least one of: a time since the meeting participant last spoke, a frequency with which the meeting participant speaks, a turn of the head to face the direction of a speaker, a gesture, a facial expression, a change of facial expression, a nod, a gaze direction, a non-verbal sound, and a drawing of breath.

13 The system of claim 1, wherein generating as an output, based on the relative framing priorities, one or more primary video streams featuring the first meeting participant and/or the second meeting participant comprises:

generating, as an output, a first primary video stream including a framed representation of the first meeting participant during a first period of time and a second primary video stream including a framed representation of the second meeting participant during a second period of time;

wherein relative durations of the first and second periods of time are determined based on the relative framing priorities associated with the first meeting participant and the second meeting participant.

14. The system of claim 13, wherein the framing priority associated with the first meeting participant is higher than the framing priority associated with the second meeting participant; and wherein the duration of the first period of time is longer than the duration of the second period of time.

15. The system of claim 13 or claim 14, wherein the first period of time and/or the second period of time is discontinuous.

16. The system of any preceding claim, wherein the one or more detected characteristics include one or more of whether a participant is speaking, a length of time a participant speaks, a percentage of a time duration during which a participant speaks, whether a participant is standing, a looking direction of a participant, changes in a looking direction of a participant, gestures performed by a participant or reactions performed by a participant.

17. The system of claim any preceding claim, wherein the at least one video processing unit is further configured to assign an observed property to each of the first meeting

participant and the second meeting participant based on analysis of the at least one overview video output stream.

18. The system of claim 17, wherein the observed property includes a demeanor; optionally wherein the demeanor includes at least one of serious, humorous, questioning, listening, or disengaged.

19. The system of claim 17 or claim 18, wherein the relative framing priorities associated with the first meeting participant and the second meeting participant are determined based on priorities associated with the assigned observed properties.

20. The system of claim 19, wherein the priority associated with at least one of the observed properties is user-assignable.

21. The system of any preceding claim, wherein the at least one video processing unit is further configured to assign an observed action to each of the first meeting participant and the second meeting participant based on analysis of the at least one overview video output stream; optionally wherein the observed action includes at least one of standing, sitting, raising a hand, interacting with a whiteboard, applauding, speaking, sleeping, writing, or laughing.

22. The system of claim 21, wherein the relative framing priorities associated with the first meeting participant and the second meeting participant are determined based on priorities associated with the assigned observed actions.

23. The system of claim 22, wherein the priority associated with at least one of the observed actions is user-assignable.

24. The system of any preceding claim, wherein the at least one video processing unit is configured to:

receive a user-assigned priority associated with the first meeting participant; and

determine the relative framing priorities based on based on the user-assigned priority associated with the first meeting participant.

25. The system of any one of claims 1 to 23, wherein the at least one video processing unit is configured to:

receive a user-assigned role designation associated with the first meeting participant and a user-assigned priority associated with the user-assigned role designation;

assign the user-assigned role designation to the first meeting participant; and

determine the framing priority for the first meeting participant based on the user-assigned priority associated with the user-assigned role designation.

26. The system of any preceding claim, wherein the at least one camera includes a plurality of cameras; optionally wherein the at least one video processing unit is further configured to automatically analyze the overview video output streams from each of the plurality of cameras, and, based on at least one identity indicator, track and correlate representations of meeting participants across the overview video output streams.

27. The system of claim 26, wherein the at least one video processing unit is included in one of the plurality of cameras.

28. The system of any preceding claim, wherein the at least one video processing unit includes one or more microprocessors onboard a camera.

29. The system of any preceding claim, wherein role designations include a presenter role designation, a contributor role designation, an audience role designation, and/or a user-assigned role designation.

30. The system of any preceding claim, wherein the at least one video processing unit is configured to implement at least one trained network configured to output role designations for identified meeting participants based on an input including one or more captured video frames from the at least one overview video output stream.

31. The system of any preceding claim, wherein the at least one video processing unit is configured to cause at least one display to show the one or more primary video streams.

32. The system of any preceding claim, further including at least one directional audio unit.

33. The system of any preceding claim, wherein the primary video stream output includes a speaker shot, a presenter shot, an overview shot, a reaction shot, an over-the-shoulder shot, or a group shot.

34. The system of any preceding claim, wherein the video processing unit is configured to transition between primary video stream outputs using hard cut transitions and/or smooth cut transitions

35. A video processing method performed by at least one video processing unit, the method comprising:

automatically analyzing at least one overview video output stream representative of a meeting environment generated by at least one camera to identify at least a first meeting participant and a second meeting participant represented in the at least one overview video output stream;

assigning a first role designation to the first meeting participant and a second role designation to the second meeting participant based on one or more characteristics associated with at least one of the first meeting participant or the second meeting participant determined based on analysis of the at least one overview video output stream;

determining, based on priorities associated with the first and second role designations, relative framing priorities associated with the first meeting participant and the second meeting participant; and

generating as an output, based on the relative framing priorities, one or more primary video streams featuring the first meeting participant and/or the second meeting participant;

wherein the method further comprises periodically updating the roles assigned to each of the first meeting participant and the second meeting participant based on analysis of the at least one overview video output stream.

36. A video conferencing system, comprising:

at least one camera, each configured to generate an overview video output stream representative of a meeting environment; and

at least one video processing unit configured to:

automatically analyze the at least one overview video output stream to identify at least a first meeting participant represented in the at least one overview video output stream;

assign a first role designation to the first meeting participant based on one or more characteristics associated with the first meeting participant determined based on analysis of the at least one overview video output stream;

determine, based on the first role designation assigned to the first meeting participant, a likelihood that the first meeting participant will begin to speak;

perform, in response to the likelihood that the first meeting participant will begin to speak meeting a likeliness condition, one or more actions to prepare the video processing unit to output a primary video stream featuring a framed representation of the first meeting participant for display;

determine, based on analysis of the at least one overview video output stream, whether the first meeting participant begins to speak; and

in response to a determination that the first meeting participant begins to speak, outputting for display the primary video stream featuring the framed representation of the first meeting participant;

wherein the at least one video processing unit is configured to periodically update the first role assigned to the first meeting participant based on analysis of the at least one overview video output stream.

37. A video processing method performed by at least one video processing unit, the method comprising:

automatically analyzing at least one overview video output stream representative of a meeting environment generated by at least one camera to identify at least a first meeting participant represented in the at least one overview video output stream;

assigning a first role designation to the first meeting participant based on one or more characteristics associated with the first meeting participant determined based on analysis of the at least one overview video output stream;

determining, based on the first role designation assigned to the first meeting participant, a likelihood that the first meeting participant will begin to speak;

performing, in response to the likelihood that the first meeting participant will begin to speak meeting a likeliness condition, one or more actions to prepare the video processing unit to output a primary video stream featuring a framed representation of the first meeting participant for display;

determining, based on analysis of the at least one overview video output stream, whether the first meeting participant begins to speak; and

in response to a determination that the first meeting participant begins to speak, outputting for display the primary video stream featuring the framed representation of the first meeting participant;

wherein the method further comprises periodically updating the role assigned to the first meeting participant based on analysis of the at least one overview video output stream.

38. A video conferencing system, comprising:

at least one camera, each configured to generate an overview video output stream representative of a meeting environment; and

at least one video processing unit configured to:

automatically analyze the at least one overview video output stream to identify a plurality of meeting participants featured in the at least one overview video output stream;

identify a plurality of non-speaking meeting participants amongst the identified meeting participants;

assign a respective role designation to each of the identified non-speaking meeting participants based on one or more characteristics associated with each of the non-speaking meeting participants determined based on analysis of the at least one overview video output stream; and

generate, as an output for display, a primary video stream featuring one or more of the non-speaking meeting participants, wherein selection of the one or more non-speaking meeting participants for inclusion in the primary video stream is based on the roles assigned to each of the non-speaking meeting participants;

wherein the at least one video processing unit is configured to periodically update the roles assigned to each of the non-speaking meeting participants based on analysis of the at least one overview video output stream.

39. The system of claim 38, wherein the at least one video processing unit is configured to identify a speaking meeting participant amongst the identified meeting participants; and wherein the primary video stream features the speaking meeting participant.

40. A video processing method performed by at least one video processing unit, the method comprising:

automatically analyzing at least one overview video output stream representative of a meeting environment generated by at least one camera to identify a plurality of meeting participants featured in the at least one overview video output stream;

identifying a plurality of non-speaking meeting participants amongst the identified meeting participants;

assigning a respective role designation to each of the identified non-speaking meeting participants based on one or more characteristics associated with each of the non-speaking meeting participants determined based on analysis of the at least one overview video output stream; and

generating, as an output for display, a primary video stream featuring one or more of the non-speaking meeting participants, wherein selection of the one or more non-speaking meeting participants for inclusion in the primary video stream is based on the roles assigned to each of the non-speaking meeting participants

wherein the method further comprises periodically updating the roles assigned to each of the non-speaking meeting participants based on analysis of the at least one overview video output stream.

41. At least one video processing unit configured to perform the method of any one of claims 35, 37 and 40.

42. One or more computer-readable storage media storing instructions that, when executed by at least one video processing unit, cause the at least one video processing unit to perform the method of any one of claims 35, 37 and 40.

43. The system of any one of claims 1 to 34, 36, 38 or 39, wherein the camera system is a multi-camera video conferencing camera system, and wherein the meeting environment is a meeting room, an office space, a classroom and/or a lecture hall.