

CLAIMS

1. A method of processing video data, the method comprising: obtaining 360-degree video data, the 360-degree video data including a plurality of pictures divided into a plurality of motion-constrained tiles; generating, using the 360-degree video data, a media file including a plurality of tracks, each of the plurality of tracks containing a corresponding set of the plurality of motion-constrained tiles, wherein each set of the plurality of motion-constrained tiles corresponds to a corresponding viewport of a plurality of viewports of the 360-degree video data; generating, for the media file, an adaptation set including a set of tile representations corresponding to the plurality of viewports, the set of tile representations including a first tile representation encapsulating a first track of the plurality of tracks, the first track including a plurality of tiles at a first tile location in the plurality of pictures, wherein the plurality of tiles at the first tile location in the plurality of pictures are included in the first tile representation, and correspond to a viewport of the 360-degree video data; generating, for the adaptation set, a first tile base representation encapsulating tile representations associated with a first version of the 360-degree video data and a second tile base representation encapsulating tile representations associated with a second version of the 360-degree video data; and generating first 360-degree video specific metadata for the first tile base representation and second 360-degree video specific metadata for the second tile base representation, the first 360-degree video specific metadata including information related to the viewport of the 360-degree video data corresponding to the plurality of tiles.
2. The method of claim 1, wherein the first tile representation includes a Dynamic Adaptive Streaming over HTTP (DASH) representation.
3. The method of claim 1, wherein the media file includes a manifest file associated with an adaptive streaming format.
4. The method of claim 1, wherein the media file includes a media presentation description (MPD).
5. The method of claim 1, wherein the set of tile representations include the plurality of motion-constrained tiles, wherein the plurality of pictures divided into the plurality of motion-constrained tiles are part of at least one enhancement layer, and wherein the set of tile representations further include a plurality of pictures of a base layer.
6. The method of claim 5, wherein the plurality of pictures of the base layer do not include motion-constrained tiles.
7. The method of claim 5, wherein the plurality of pictures of the base layer include motion-constrained tiles.
8. The method of claim 1, wherein the first tile base representation encapsulates all motion-

constrained tiles of pictures in the first version of the 360-degree video data, and wherein the second tile base representation encapsulates all motion-constrained tiles of pictures in the second version of the 360-degree video data.

9. The method of claim 1, wherein the first tile base representation includes a first DASH representation, and wherein the second tile base representation includes a second DASH representation.

10. The method of claim 1, wherein the first 360-degree video specific metadata includes information associated with at least one or more of a video projection type, one or more viewport directions, or region-wise mapping of one or more pictures of the first version of the 360-degree video data.

11. The method of claim 1, wherein the first 360-degree video specific metadata is defined for the first tile representation and is applicable to a region in a picture of the 360-degree video data that is covered by a motion-constrained tile at the first tile location in the plurality of pictures of the 360-degree video data, the first 360-degree video specific metadata including information associated with at least one or more of a video projection type, one or more viewport directions, or region information of the first tile representation.

12. The method of claim 1, further comprising generating a descriptor indicating one or more most interested regions in the 360-degree video data.

13. The method of claim 12, wherein the descriptor is generated on a tile base representation level, wherein a tile base representation encapsulates all motion-constrained tiles of pictures in at least one layer.

14. An apparatus for processing video data, comprising: a memory configured to store 360-degree video data; and a processor configured to: obtain the 360-degree video data, the 360-degree video data including a plurality of pictures divided into a plurality of motion-constrained tiles; generate, using the 360-degree video data, a media file including a plurality of tracks, each of the plurality of tracks containing a corresponding set of the plurality of motion-constrained tiles, wherein each set of the plurality of motion-constrained tiles corresponds to a corresponding viewport of a plurality of viewports of the 360-degree video data; generate an adaptation set including a set of tile representations corresponding to the plurality of viewports, the set of tile representations including a first tile representation encapsulating a first track of the plurality of tracks, the first track including a plurality of tiles at a first tile location in the plurality of pictures, wherein the plurality of tiles at the first tile location in the plurality of pictures are included in the first tile representation and correspond to a viewport of the 360-degree video data; generate, for the adaptation set, a first tile base representation encapsulating tile representations associated with a first version of the 360-degree video data and a second tile base representation encapsulating tile representations associated with a second version of the 360-degree video data; and generate first 360-degree video specific metadata for the first tile base representation and second 360-degree video

specific metadata for the second tile base representation, the first 360-degree video specific metadata including information related to the viewport of the 360-degree video data corresponding to the plurality of tiles.

15. The apparatus of claim 14, wherein the first tile representation includes a Dynamic Adaptive Streaming over HTTP (DASH) representation.

16. The apparatus of claim 14, wherein the media file includes a manifest file associated with an adaptive streaming format.

17. The apparatus of claim 14, wherein the media file includes a media presentation description (MPD).

18. The apparatus of claim 14, wherein the set of tile representations include the plurality of motion-constrained tiles, wherein the plurality of pictures divided into the plurality of motion-constrained tiles are part of at least one enhancement layer, and wherein the set of tile representations further include a plurality of pictures of a base layer.

19. The apparatus of claim 18, wherein the plurality of pictures of the base layer do not include motion-constrained tiles.

20. The apparatus of claim 18, wherein the plurality of pictures of the base layer include motion-constrained tiles.

21. The apparatus of claim 14, wherein the first tile base representation encapsulates all motion-constrained tiles of pictures in the first version of the 360-degree video data, and wherein the second tile base representation encapsulates all motion-constrained tiles of pictures in the second version of the 360-degree video data.

22. The apparatus of claim 14, wherein the first tile base representation includes a first DASH representation, and wherein the second tile base representation includes a second DASH representation.

23. The apparatus of claim 14, wherein the first 360-degree video specific metadata includes information associated with at least one or more of a video projection type, one or more viewport directions, or region-wise mapping of one or more pictures of the first version of the 360-degree video data.

24. The apparatus of claim 14, wherein the first 360-degree video specific metadata is defined for the first tile representation and is applicable to a region in a picture of the 360-degree video data that is covered by a motion-constrained tile at the first tile location in the plurality of pictures of the 360-degree video data, the first 360-degree video specific metadata including information associated with at least one or more of a video projection type, one or more viewport directions, or region information of the first tile representation.

25. The apparatus of claim 14, wherein the processor is further configured to generate a descriptor indicating one or more most interested regions in the 360-degree video data.

26. The apparatus of claim 25, wherein the descriptor is generated on a tile base representation level, wherein a tile base representation encapsulates all motion-constrained tiles of pictures in at least one layer.

27. The apparatus of claim 14, wherein the apparatus comprises a mobile device with a camera for capturing pictures.

28. A method of processing one or more media files, comprising: obtaining a media file including a plurality of tracks, each of the plurality of tracks containing a corresponding set of a plurality of motion-constrained tiles of a plurality of pictures representing 360-degree video data, wherein each set of the plurality of motion-constrained tiles corresponds to a corresponding viewport of a plurality of viewports of the 360-degree video data, wherein the media file further includes an adaptation set including a set of tile representations corresponding to the plurality of viewports, the set of tile representations including a first tile representation encapsulating a first track of the plurality of tracks, the first track including a plurality of tiles at a first tile location in the plurality of pictures, wherein the plurality of tiles at the first tile location in the plurality of pictures are included in the first tile representation, and correspond to a viewport of the 360-degree video data; obtaining, from the adaptation set, a first tile base representation encapsulating tile representations associated with a first version of the 360-degree video data and a second tile base representation encapsulating tile representations associated with a second version of the 360-degree video data; obtaining first 360-degree video specific metadata for the first tile base representation and second 360-degree video specific metadata for the second tile base representation, the first 360-degree video specific metadata including information related to the viewport of the 360-degree video data corresponding to the plurality of tiles; and processing the media file based on at least one of the first 360-degree video specific metadata or the second 360-degree video specific metadata.

29. The method of claim 28, wherein the first tile representation includes a Dynamic Adaptive Streaming over HTTP (DASH) representation.

30. The method of claim 28, wherein the media file includes a manifest file associated with an adaptive streaming format.

31. The method of claim 28, wherein the media file includes a media presentation description (MPD).

32. The method of claim 28, wherein the set of tile representations include the plurality of motion-constrained tiles, wherein the plurality of pictures divided into the plurality of motion-constrained tiles are part of at least one enhancement layer, and wherein the set of tile

representations further include a plurality of pictures of a base layer.

33. The method of claim 32, wherein the plurality of pictures of the base layer do not include motion-constrained tiles.

34. The method of claim 32, wherein the plurality of pictures of the base layer include motion-constrained tiles.

35. The method of claim 28, wherein the first tile base representation encapsulates all motion-constrained tiles of pictures in the first version of the 360-degree video data, and wherein the second tile base representation encapsulates all motion-constrained tiles of pictures in the second version of the 360-degree video data.

36. The method of claim 28, wherein the first tile base representation includes a first DASH representation, and wherein the second tile base representation includes a second DASH representation.

37. The method of claim 28, wherein the first 360-degree video specific metadata includes information associated with at least one or more of a video projection type, one or more viewport directions, or region-wise mapping of one or more pictures of the first version of the 360-degree video data.

38. The method of claim 28, wherein the first 360-degree video specific metadata is defined for the first tile representation and is applicable to a region in a picture of the 360-degree video data that is covered by a motion-constrained tile at the first tile location in the plurality of pictures of the 360-degree video data, the first 360-degree video specific metadata including information associated with at least one or more of a video projection type, one or more viewport directions, or region information of the first tile representation.

39. The method of claim 28, wherein the media file includes a descriptor indicating one or more most interested regions in the 360-degree video data.

40. The method of claim 39, wherein the descriptor is generated on a tile base representation level, wherein a tile base representation encapsulates all motion-constrained tiles of pictures in at least one layer.

41. An apparatus for processing one or more media files, comprising: a memory configured to store the one or more media files comprising 360-degree video data; and a processor configured to: obtain a media file including a plurality of tracks, each of the plurality of tracks containing a corresponding set of a plurality of motion-constrained tiles of a plurality of pictures representing 360-degree video data, wherein each set of the plurality of motion-constrained tiles corresponds to a corresponding viewport of a plurality of viewports of the 360-degree video data, wherein the media file further includes an adaptation set including a set of tile representations corresponding to the plurality of viewports, the set of tile

representations including a first tile representation encapsulating a first track of the plurality of tracks, the first track including a plurality of tiles at a first tile location in the plurality of pictures, wherein the plurality of tiles at the first tile location in the plurality of pictures are included in the first tile representation, and correspond to a viewport of the 360-degree video data; obtain, from the adaptation set, a first tile base representation encapsulating tile representations associated with a first version of the 360-degree video data and a second tile base representation encapsulating tile representations associated with a second version of the 360-degree video data; obtain first 360-degree video specific metadata for the first tile base representation and second 360-degree video specific metadata for the second tile base representation, the first 360-degree video specific metadata including information related to the viewport of the 360-degree video data corresponding to the first set plurality of tiles; and process the media file based at least one of the first 360-degree video specific metadata or the second 360-degree video specific metadata.

42. The apparatus of claim 41, wherein the first tile representation includes a Dynamic Adaptive Streaming over HTTP (DASH) representation.

43. The apparatus of claim 41, wherein the media file includes a manifest file associated with an adaptive streaming format.

44. The apparatus of claim 41, wherein the media file includes a media presentation description (MPD).

45. The apparatus of claim 41, wherein the set of tile representations include the plurality of motion-constrained tiles, wherein the plurality of pictures divided into the plurality of motion-constrained tiles are part of at least one enhancement layer, and wherein the set of tile representations further include a plurality of pictures of a base layer.

46. The apparatus of claim 45, wherein the plurality of pictures of the base layer do not include motion-constrained tiles.

47. The apparatus of claim 45, wherein the plurality of pictures of the base layer include motion-constrained tiles.

48. The apparatus of claim 41, wherein the first tile base representation encapsulates all motion-constrained tiles of pictures in the first version of the 360-degree video data, and wherein the second tile base representation encapsulates all motion-constrained tiles of pictures in the second version of the 360-degree video data.

49. The apparatus of claim 41, wherein first tile base representation includes a first DASH representation, and wherein the second tile base representation includes a second DASH representation.

50. The apparatus of claim 41, wherein the first 360-degree video specific metadata includes

information associated with at least one or more of a video projection type, one or more viewport directions, or region-wise mapping of one or more pictures of the first version of the 360-degree video data.

51. The apparatus of claim 41, wherein the first 360-degree video specific metadata is defined for the first tile representation and is applicable to a region in a picture of the 360-degree video data that is covered by a motion-constrained tile at the first tile location in the plurality of pictures of the 360-degree video data, the first 360-degree video specific metadata including information associated with at least one or more of a video projection type, one or more viewport directions, or region information of the first tile representation.

52. The apparatus of claim 41, wherein the media file includes a descriptor indicating one or more most interested regions in the 360-degree video data.

53. The apparatus of claim 52, wherein the descriptor is generated on a tile base representation level, wherein a tile base representation encapsulates all motion-constrained tiles of pictures in at least one layer.

54. The apparatus of claim 41, further comprising: a display for displaying the video data.

55. The apparatus of claim 41, wherein the apparatus comprises a mobile device with a camera for capturing pictures.