

**Internal distribution code:**

- (A) [ - ] Publication in OJ
- (B) [ - ] To Chairmen and Members
- (C) [ - ] To Chairmen
- (D) [ X ] No distribution

**Datasheet for the decision  
of 14 October 2024**

**Case Number:** T 2075/21 - 3.5.04

**Application Number:** 13840196.3

**Publication Number:** 2941865

**IPC:** H04N19/00

**Language of the proceedings:** EN

**Title of invention:**

INDICATION OF VIEW DEPENDENCY ON REFERENCE VIEW IN MULTIVIEW  
CODING FILE FORMAT

**Applicant:**

Qualcomm Incorporated

**Headword:**

**Relevant legal provisions:**

EPC Art. 56

RPBA 2020 Art. 15a(1)

**Keyword:**

Main request and auxiliary request - inventive step (no)  
Oral proceedings - request for oral proceedings to be held by  
videoconference (refused)

**Decisions cited:**

**Catchword:**



**Beschwerdekammern**  
**Boards of Appeal**  
**Chambres de recours**

Boards of Appeal of the  
European Patent Office  
Richard-Reitzner-Allee 8  
85540 Haar  
GERMANY  
Tel. +49 (0)89 2399-0

Case Number: T 2075/21 - 3.5.04

**D E C I S I O N**  
**of Technical Board of Appeal 3.5.04**  
**of 14 October 2024**

**Appellant:** Qualcomm Incorporated  
(Applicant) 5775 Morehouse Drive  
San Diego, CA 92121-1714 (US)

**Representative:** Reddie & Grose LLP  
The White Chapel Building  
10 Whitechapel High Street  
London E1 8QS (GB)

**Decision under appeal:** **Decision of the Examining Division of the  
European Patent Office posted on 22 June 2021  
refusing European patent application  
No. 13840196.3 pursuant to Article 97(2) EPC.**

**Composition of the Board:**

**Chair** B. Willems  
**Members:** B. Le Guen  
W. Ungler

## **Summary of Facts and Submissions**

- I. The appeal is against the decision to refuse European patent application No. 13 840 196.3.
- II. The prior-art documents cited in the decision included the following:
  - D4: Teruhiko Suzuki et al., "Study Text of ISO/IEC 14496-10:2012/DAM2 MVC extension for inclusion of depth maps", Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, 2nd Meeting, Shanghai, CN, No. JCT3V-B1001, 2 November 2012, XP030130410
  - D9: ISO/IEC JCT1/SC29/WG11, "Text of ISO/IEC 14496-15:2010/DAM 2 Carriage of HEVC", 102nd MPEG Meeting, Shanghai, CN, No. n13036, 5 November 2012, XP030273045
- III. The decision under appeal was based on the grounds that the subject-matter of claim 1 of the main request and of the auxiliary request then on file did not involve an inventive step (Article 56 EPC) in view of the combined disclosures of documents D9 and D4 and the common general knowledge of the person skilled in the art.
- IV. The applicant (appellant) filed notice of appeal. With its statement of grounds of appeal, the appellant re-filed the claims of the main request and of the auxiliary request on which the decision under appeal

was based and provided arguments as to why the examining division erred in its findings.

V. The appellant was summoned to oral proceedings. In a communication under Article 15(1) RPBA, the board introduced the following documents into the appeal proceedings pursuant to Article 114(1) EPC:

D11: Telecommunication standardization sector of ITU: "Recommendation ITU-T H.264", H.264 (01/2012), 9 August 2012, retrieved from the internet: URL: <https://www.itu.int/rec/T-REC-H.264-201201-S/en>

D12: Ying Chen and Ye-Kui Wang, "AHG7: Comments on MVC+D", Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 2nd Meeting, Shanghai, CN, 13-19 October 2012, JCT3V-B0055, XP030055076

In its communication, the board also set out its preliminary opinion that the subject-matter of claim 1 of neither the main request nor the auxiliary request involved an inventive step in view of the combined disclosures of documents D9 and D4 and the common general knowledge of the person skilled in the art.

VI. In a letter dated 22 July 2024, the appellant requested that the oral proceedings be held by videoconference.

VII. In a communication dated 5 September 2024, the board informed the appellant that in view of the expected complexity of the discussions it did not consider it appropriate to hold the oral proceedings by videoconference.

VIII. By letter dated 12 September 2024, the appellant confirmed that it would attend the in-person oral proceedings scheduled for 14 October 2024.

IX. The oral proceedings before the board were held on the scheduled date.

The appellant's final requests were that the decision under appeal be set aside and that a European patent be granted on the basis of the claims of the main request filed with the statement of grounds of appeal or, alternatively, on the basis of the claims of the auxiliary request filed with the statement of grounds of appeal.

At the end of the oral proceedings, the chair announced the board's decision.

X. Claim 1 of the **main request** reads as follows:

*"A method of processing video data, the method comprising:*

*parsing (220) a track of video data, wherein the track includes one or more views;*

*parsing a view identifier box from at least one of a sample entry and a multi-view group entry, wherein the at least one of the sample entry and the multi-view group entry are associated with the track, wherein parsing the view identifier box comprises, for each reference view of each of the one or more views in the track, parsing (222) a two-bit syntax element `dependent_component_idc` that indicates whether the texture view and the depth view of the reference view are required for decoding at least one of the one or more views in the track with a value, wherein, when the value of `dependent_component_idc` is equal to 0, only*

*the texture view of the reference view is required for decoding at least one of the one or more views in the track, when the value of dependent\_component\_idc is equal to 1, only the depth view of the reference view is required for decoding at least one of the one or more views in the track, and when the value of dependent\_component\_idc is equal to 2, both the texture view and the depth view of the reference view is required for decoding at least one of the one or more views in the track; and*

*decoding the at least one of the one or more views in the track of video data based on the indicated texture view or depth view of the reference view."*

- XI. Claim 1 of the **auxiliary request** reads as follows (with features that have been added as compared to claim 1 of the main request being underlined; features that have been deleted have been crossed out):

*"A method of processing video data, the method comprising:*

*parsing (220) a track of video data, wherein the track includes one or more views;*

*parsing a view identifier box from at least one of a sample entry and a multi-view group entry, wherein the at least one of the sample entry and the multi-view group entry are associated with the track, wherein parsing the view identifier box comprises, for each reference view of each of the one or more views in the track, parsing a syntax element ref view id that specifies a view identifier for the reference view, and parsing (222) a two-bit syntax element dependent\_component\_idc that indicates whether the texture view and the depth view of the reference view are required for decoding at least one of the one or more views in the track with a value, wherein, when the*

*value of dependent\_component\_idc is equal to 0, only the texture view of the reference view is required for decoding at least one of the one or more views in the track, when the value of dependent\_component\_idc is equal to 1, only the depth view of the reference view is required for decoding at least one of the one or more views in the track, and when the value of dependent\_component\_idc is equal to 2, both the texture view and the depth view of the reference view is required for decoding at least one of the one or more views in the track; and*

*decoding the at least one of the one or more views in the track of video data based on the indicated texture view or depth view of the reference view."*

XII. The appellant's arguments, where relevant to the present decision, can be summarised as follows.

- (a) The distinguishing features made it possible to identify only those view components of a reference view that were required to decode a given view. This facilitated extracting only the required components from the file, which in turn allowed a reduction of the buffer requirements and the computational load of a decoder.
- (b) No combination of the prior-art documents on file led to the claimed subject-matter.
- (c) Although documents D4 and D12 disclosed that the texture-view and depth-view components of the same view could have different view dependencies in the MVC+D standard, none of the cited prior-art documents taught that buffer space could be saved by extracting only the view component(s) of a reference view that was (were) required for decoding a given view.

- (d) In any case, none of the cited prior-art documents taught that the buffer savings would be significant enough to justify spending additional bits in the file format to indicate the required view component(s).
- (e) There were other ways of signalling the required view component(s) of a reference view than by adding a two-bit flag. For example, the skilled person could do so by replacing `ref_view_id[i][j]` with syntax elements directly identifying the respective depth/texture component of the reference view.

### **Reasons for the Decision**

1. The appeal is admissible.

#### *Summary of the invention*

2. As summarised in section 4.1 of the statement of grounds of appeal, the invention relates to a file format for storing multiview video data files, and to storing information about the coded multiview bitstream as file format metadata. This information is used by a media player to understand the nature and structure of a bitstream in a video file, in order to identify an operating point within the capabilities of its decoder and then extract the necessary parts of the bitstream to pass to the decoder for decoding.
3. A multiview bitstream contains image sequences captured from multiple camera angles, and for each of these "texture-view" components it may also contain an associated "depth-view" component which contains information about the depth of objects in the scene. The invention concerns the inclusion of metadata that

indicates the dependency between view components so that a media player parsing the metadata is made aware of these dependencies, and can select an appropriate operating point accordingly.

4. More specifically, the invention concerns the inclusion, in the view identifier box of a track of multiview video data, of a two-bit syntax element which indicates whether the texture-view component, the depth-view component, or both view components of an identified reference-view component are required for decoding one or more views included in the track.

*Main request - inventive step*

5. An invention is to be considered to involve an inventive step if it is not obvious to a person skilled in the art in view of the state of the art (Article 56 EPC).
6. Document D9 discloses a method of processing video data comprising the following steps:
  - parsing a track of video data, wherein the track includes one or more views (see D9, clause 7.1: *"This clause specifies the storage format of MVC data [...] Extractor: a structure to enable efficient extraction of NAL units from other tracks [...] The MVC file format allows storage of one or more views into a track [...]"*; and clause 7.6.1: *"A multi-view video stream is represented by one or more video tracks in a file. Each track represents one or more views of the stream"*);
  - parsing a view identifier box from at least one of a sample entry and a multi-view group entry (see D9, clause 7.6.3.1.3: *"View Identifier Box"*; and

clause 7.6.3.1.3.1: "*Container: Sample Entry [...] or MultiviewGroupEntry*"), wherein the at least one of the sample entry and the multi-view group entry are associated with the track (see clause 7.6.3.1.3.1: "*When included in a sample entry, this box indicates the views included in the track. When included in a Multiview Group entry, this box indicates the views included in the respective tier*"; and clause 3.1.21: "*An MVC tier represents a particular set of temporal subsets of a particular set of views*"), wherein parsing the view identifier box comprises, for each reference view of each of the one or more views in the track, parsing "*the number of views which are referenced by the view with view\_id[i]*" and "*the view identifier of the j-th view that may be directly or indirectly referenced by the view with view\_id[i]*" (see D9, clause 7.6.3.1.3.3).

7. It is undisputed that the subject-matter of claim 1 of the main request differs from the method disclosed in document D9 by the following features:
  - parsing the view identifier box comprises, for each reference view of each of the one or more views in the track, parsing a two-bit syntax element `dependent_component_idc` that indicates whether the texture view and the depth view of the reference view are required for decoding at least one of the one or more views in the track with a value, wherein, when the value of `dependent_component_idc` is equal to 0, only the texture view of the reference view is required for decoding at least one of the one or more views in the track, when the value of `dependent_component_idc` is equal to 1, only the depth view of the reference view is

required for decoding at least one of the one or more views in the track, and when the value of `dependent_component_idc` is equal to 2, both the texture view and the depth view of the reference view is required for decoding at least one of the one or more views in the track.

- the method further comprises a step of decoding the at least one of the one or more views in the track of video data based on the indicated texture view or depth view of the reference view.

8. As put forward by the appellant (see section XII.(a) above), the distinguishing features facilitate extracting, from a file encapsulating an MVC+D stream, only the component(s) of a reference view that is (are) required to decode a given view.
9. At the oral proceedings before the board, it was agreed that the objective technical problem could be formulated as improving the ISO/IEC 14496-15 storage standard (i.e. document D9) to provide a more efficient storage of an MVC+D stream.
10. Document D4 is a draft of an amendment to the ISO/IEC 14496-10 video coding standard (document D11) including Annex I ("*Multiview and depth video coding*"). Clauses I.7.3.2.1.4 and I.7.3.2.1.5 on pages 31 and 32 of document D4 define a "*Sequence parameter set MVC extension syntax*" and a "*Sequence parameter set 3DVC extension syntax*". The sequence parameter set MVC extension syntax and semantics are specified in clause H.7.3.2.1.4 of Annex H of document D11, Annex H being the extension dedicated to "*Multiview video coding*" (without depth). Pursuant to clause H.7.4.2.1.4 of document D11, the sequence parameter set MVC

extension specifies *"inter-view dependency relationships for the coded video sequence"*, the syntax elements `num_anchor_refs_10`, `anchor_ref_10`, `num_anchor_refs_11` and `anchor_ref_11` specifying the number and identities of the reference views that can be referred to predict a given view (which is a texture view component in the context of Annex H). The use of a 3DVC sequence parameter set in addition to the use of the MVC sequence parameter set in Annex I of document D4 makes the MVC+D standard backward-compatible with the MVC decoder for the decoding of texture data. The sequence parameter set 3DVC extension semantics are specified in clause I.7.4.2.1.4 of document D4. According to that clause, *"num\_anchor\_refs\_10[ i ], anchor\_ref\_10[ i ][ j ], num\_anchor\_refs\_11[ i ], anchor\_ref\_11[ i ][ j ], num\_non\_anchor\_refs\_10[ i ], non\_anchor\_ref\_10[ i ][ j ], num\_non\_anchor\_refs\_11[ i ] and non\_anchor\_ref\_11[ i ][ j ] [parsed from seq\_parameter\_set\_3dvc\_extension() specified in clause I.7.3.2.1.5] **apply to depth view components [only]**"* (emphasis added by the board; see also document D4, page 41, last paragraph: *"More specifically, the view dependency information of the 3DVC sequence parameter set extension applies only to the depth views"*). This means that Annex I was conceived to allow depth-view components to have different view dependencies than texture-view components, otherwise there would have been no need to specify additional anchor and non-anchor references specifically for depth-view components. This is confirmed by the statement made in document D12 that *"In MVC+D, the texture view and depth view with the same view id may have different view dependencies, which are specified in two sequence parameter set MVC extension syntax tables"*. This fact was acknowledged by the appellant at the oral proceedings before the board.

11. It is correct that no combination of the prior-art documents on file leads on its own to the claimed subject-matter (see the appellant's argument in section XII.(b) above). However, the skilled person's common general knowledge and ability to perform routine work and experimentation should also be taken into account when assessing whether the claimed invention involves an inventive step (with regard to the definition of the skilled person, see Case Law of the Boards of Appeal of the European Patent Office, 10th edition, 2022, "Case Law", I.D.8.1.1).
  
12. In cases in which depth-view components have different view dependencies than texture-view components, the skilled person would have observed that more view components than strictly necessary were extracted and buffered when the extraction was carried out on the basis of the reference-view identifiers specified in clause 7.6.3.1.3 of document D9 only. In observing this, the skilled person would have realised that buffer space could be saved by extracting only the view component(s) of a reference view that was (were) required for decoding a given view (see the appellant's argument in section XII.(c) above). This can only be achieved by providing an indication of the required reference-view component(s) in the file format.
  
13. Since there are two possible view components (texture and depth), only three cases can occur: only the texture-view component of a reference view is required for decoding a given view, only the depth-view component is required or both the texture-view and depth-view components are required. Signalling one of three possible options using two bits is part of the common general knowledge of the skilled person.

Therefore, the board finds that adding a two-bit syntax element in the view identifier box would have been an obvious way of signalling the view components of a reference view required to decode a given view. The association of the three cases mentioned above to the values 0, 1 and 2, as specified in claim 1 of the main request, would have been an obvious possibility among several.

14. It is correct that the cited prior-art documents do not teach that the buffer savings would be significant enough to justify spending additional bits in the file format to indicate the required view component(s) (see the appellant's argument in section XII.(d) above). However, on account of the fact that two bits only represent one third of a typical 8-bit component (luminance, chrominance) of a pixel of a view component, the skilled person would have expected that the buffer savings achieved by not extracting the non-required view component(s) of a reference view compensate for the signalling of an additional two-bit syntax element.
  
15. The board cannot exclude that there would have been other (possibly obvious) ways of signalling the required view component(s) of a reference view in the file format (see the appellant's argument in section XII.(d) above). However, the fact that there might be alternative options has no bearing on the obviousness of one specific option (see Case Law, I.D. 9.21 b)). In any case, the board is not persuaded that the option put forward by the appellant (namely replacing `ref_view_id[i][j]` with syntax elements directly identifying the respective depth/texture component of the reference view; see the appellant's argument in section XII.(d) above) would have been

considered by the skilled person. Document D4 does not define a syntax element in the NAL unit header directly identifying a depth-view component or a texture-view component on its own. According to clauses I.7.3.1.1 and I.7.4.1.1 of document D4, the NAL unit header extension syntax and semantics are inherited from clause H.7.3.1.1 of document D11. The syntax element `view_id` specified in clause H.7.3.1.1 of D11 does not allow a NAL unit corresponding to a texture-view component to be distinguished from a NAL unit corresponding to a depth-view component on its own (see clause I.7.3.2.1.5 of D4). Therefore, at least one syntax element would have been required in the view identifier box to indicate which NAL unit types to extract from among those having the desired `ref_view_id`.

16. In view of the above, the board agrees with the examining division that the subject-matter of claim 1 of the main request does not involve an inventive step in view of the combined disclosures of documents D9 and D4 and the common general knowledge of the person skilled in the art.

*Auxiliary request - inventive step*

17. In comparison to claim 1 of the main request, claim 1 of the auxiliary request further specifies "*parsing a syntax element `ref_view_id` that specifies a view identifier for the reference view*".
18. A syntax element (`ref_view_id`) specifying a view identifier for each of the reference views is already disclosed in clause 7.6.3.1.3 of document D9. Adding a syntax element identifying which of the view components (texture or depth) is required for inter-view

prediction would have been obvious for the reasons provided in sections 10. to 15. above.

19. Therefore, the board also agrees with the examining division that the subject-matter of claim 1 of the auxiliary request does not involve an inventive step either, in view of the combined disclosures of documents D9 and D4 and the common general knowledge of the person skilled in the art.

*Conclusion*

20. Since neither of the requests on file is allowable, the appeal must be dismissed.

*Request to hold the oral proceedings by videoconference*

21. The assessment of inventive step in the case at hand is based on an analysis of the ISO/IEC 14496-10:2012/DAM2 standard draft (document D4), the ISO/IEC 14496-15:2010/DAM 2 standard draft (document D9), the ITU-T H.264 (01/2012) standard (document D11) and their inter-dependencies. Given the length and complexity of these disclosures, the board expected that long, strenuous discussions regarding complex technical issues would take place at the oral proceedings. For that reason, the board did not consider it appropriate to hold the oral proceedings by videoconference and refused the appellant's request to that effect (Article 15a(1) RPBA).

**Order**

**For these reasons it is decided that:**

The appeal is dismissed.

The Registrar:

The Chair:



K. Boelicke

B. Willems

Decision electronically authenticated