

mmusic
Internet-Draft
Intended status: Standards Track
Expires: May 25, 2012

B. Greevenbosch
Y. Hui
Huawei
November 22, 2011

Signal 3D format
draft-ietf-mmusic-signal-3d-format-00

Abstract

This document introduces the SDP attribute "3dFormat", which provides format description of stereoscopic 3D video. In addition, the grouping mechanism for SDP is extended to cater for stereoscopic 3D video.

Note

Discussion and suggestions for improvement are requested, and should be sent to mmusic@ietf.org.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 25, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	4
2.	Requirements notation	5
3.	Definitions	6
4.	The "3dFormat" attribute	8
5.	Grouping	11
6.	Combinations of attribute values and group usage	12
7.	SDP offer/answer with 3D support	14
8.	SDP offer/answer without 3D support	16
8.1.	Frame packing	16
8.2.	2D and auxiliary as a single stream	16
8.3.	2D and auxiliary as two separate streams	16
8.4.	Simulcast of L- and R-views	17
9.	Examples	18
9.1.	One single frame compatible stream	18
9.2.	Two separate streams	18
9.3.	C-stream and depth map stream	18
9.4.	Stereoscopic 3D video with two different formats	19
10.	Formal ABNF grammar of the "3dFormat" attribute	21
11.	Security Considerations	22
12.	IANA Considerations	23
12.1.	"3dFormat" attribute	23
12.2.	"3DS" value for "group" semantics	24
13.	Acknowledgements	25
14.	Normative References	26
	Authors' Addresses	27

1. Introduction

In stereoscopic 3D multimedia applications, two views are displayed, one for the left eye and one for the right eye.

There are various ways of formatting the views of Stereoscopic 3D video. Examples of 3D formats are frame packing (see [[HDMIV1.4a](#)] and [[ISO/IEC 14496-10](#)]) and the combination of 2D video and auxiliary data such as depth maps or parallax maps (for both, see [[ISO/IEC 23002-3](#)]). Stereoscopic 3D video may be carried over a single stream or over several streams, depending on its 3D format.

In multimedia streaming applications, the Session Description Protocol (SDP) [[RFC4566](#)] can be used to provide to the receiver sufficient information about the media streams, and to enable the receiver to join and participate in the session.

This document defines an extension to SDP that provides sufficient information about the format of stereoscopic 3D video carried in the media stream(s). Before accessing the stream(s), the receiver can use the 3D format description from SDP to determine whether it has the capability to receive and render the stereoscopic 3D video content, and whether it can participate in the session.

The mentioned SDP extension is a new SDP attribute "3dFormat", which provides the format description of stereoscopic 3D video. The design of the attribute is based on the following requirements, which are listed only for informational purposes:

- o It MUST be possible to signal that the left and right views are carried in a single stream, by the use of frame packing.
- o It MUST be possible to signal that 2D video and auxiliary video (such as depth maps) are carried in a single stream.
- o It MUST be possible to signal that the left and right views are carried in two separate streams.
- o It MUST be possible to signal that 2D video and auxiliary video (such as depth maps) are carried in separate streams.

To bind multiple video streams that carry a single stereoscopic 3D video, this document also extends the SDP grouping mechanism from [[RFC5888](#)].

2. Requirements notation

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

3. Definitions

2D video

Video that does not in itself contain depth or parallax information.

auxiliary video

A sequence of depth or parallax maps, which are used to add depth to 2D video.

C-view

The centre view: a visual entity as seen from a viewpoint between the left and right eyes. The C-view can be used to calculate the L- and R-views.

C-stream

A 2D video stream consisting of a sequence of C-views.

depth map

A two dimensional map, each pixel of which defines the depth of one or more pixels in an associated 2D video frame.

depth map stream

An auxiliary stream, which contains a sequence of depth maps. The depth map stream is synchronised with the associated 2D video stream.

frame packing

A format that packs the L- and R-views into a single 2D video stream. The packing may be done spatially, where each video frame is divided into sub-frames, one containing the L-view and one containing the R-view. The packing can also be done sequentially, where alternating video frames represent L- and R-views.

legacy answerer

An answerer (in the SDP offer/answer model [[RFC3264](#)]) that does not support the "3dFormat" attribute. The legacy answerer can be the streaming server or the streaming client, but is not compliant to this document.

L-view

A visual entity that is to be projected to the left eye.

L-stream

A 2D video stream consisting of a sequence of L-views.

parallax map

A two dimensional map, each pixel of which defines the parallax of one or more pixels in an associated 2D video frame.

parallax map stream

An auxiliary stream, which contains a sequence of parallax maps. The parallax map stream is synchronised with the associated 2D video stream.

R-view

A visual entity that is to be projected to the right eye.

R-stream

A 2D video stream consisting of a sequence of R-views.

stereoscopic 3D video

The L- and R-streams, ready to be projected to the viewer's left and right eyes.

sub-frame

A part of a video frame.

4. The "3dFormat" attribute

The media-level SDP attribute "3dFormat" signals the format of stereoscopic 3D video. The attribute transfers this information through two parameters: one indicating the format type of the stereoscopic 3D video carried in the media stream(s), and the other indicating the type of the video component, which is a constituent element of the stereoscopic 3D video. The video component type depends on the format type of the stereoscopic 3D video. The syntax of the attribute is defined as follows:

`a=3dFormat:<Format Type> <Component Type>`

The <Format Type> can have the following values (as indicated between the quotes):

"FP" Frame Packing

The L- and R-views are packed into a single stream. The packing may use a side-by-side, top-and-bottom, interleaved, checkerboard or frame sequential format.

"SC" Simulcast

The L- and R-streams are transmitted separately.

"2DA" 2D + auxiliary

2D video and auxiliary data (such as depth maps or parallax maps) are transmitted. These can be transmitted in a single stream, as well as in two separate streams.

The <Component Type> can have the following values (as indicated between the quotes):

"C" Centre view

The associated stream is a C-stream.

"CD" centre view and depth map

The associated stream contains both the C-view and depth map sequences.

"ChB" Checkerboard

The video frame consists of alternating pixels from the corresponding L- and R-views, as illustrated by Figure 1.

"CP" Centre view and parallax map

The associated stream contains both the C-view and parallax map sequences.

"D" Depth map

The associated stream is a sequence of depth maps.

"L" Left view

The associated stream is the L-stream.

"LD" Left view and depth map

The associated stream contains both the L-view and depth map sequences.

"LIL" Line Interleaved

Each video frame consists of alternating scan lines from the L- and R-views.

"LP" Left view and parallax map

The associated stream contains both the L-view and parallax map sequences.

"P" Parallax map

The associated stream is a sequence of parallax maps.

"R" Right view

The associated stream is the R-stream.

"SbS" Side by Side

Each video frame is divided in two equally sized sub-frames, spatially positioned side by side of each other. One sub-frame contains the L-view, whereas the other contains the R-view.

"Seq" Frame Sequential

The single video stream consists of alternating frames from the L- and R-streams. Additional signalling, e.g. AVC SEI messages [ISO/IEC 14496-10], is needed to signal which frames contain L- and which contain R-views.

"TaB" Top and Bottom

Each video frame is divided in two equally sized sub-frames, spatially positioned above each other. One sub-frame contains the L-view, whereas the other contains the R-view.


```
+--+--+--+--+
|L|R|L|R|L|R|
+--+--+--+--+
|R|L|R|L|R|L|
+--+--+--+--+
|L|R|L|R|L|R|
+--+--+--+--+
```

The checkerboard pattern. The transmitted video frame is composed of pixels from the L- and R-views. Samples from the L-view are indicated with "L", whereas samples from the R-view are indicated with "R".

Figure 1

5. Grouping

When multiple streams carry a single stereoscopic 3D video, (e.g. C-stream and parallax map, or separately transmitted L- and R-streams), the grouping mechanism from [[RFC5888](#)] MUST be used.

However, to cater for the special requirements of 3D signalling, the semantics are expanded:

```
group-attribute      = "a=group:" semantics *(SP identification-tag)
semantics             = "LS" / "FID" / "3DS" / semantics-extension
semantics-extension  = token
```

The grouping is needed when multiple streams carry a single stereoscopic 3D video. This is the case when the <format type> is "SC", or the <format type> is "2DA" and the 2D video and auxiliary data are transmitted as multiple streams. A group with the "3Ds" semantics is called a "3DS group".

A 3DS group MUST NOT contain data that is (potentially) inconsistent with other data in the 3DS group:

- o A 3DS group MUST NOT contain both a parallax map stream and a depth map stream.
- o A 3DS group MUST NOT contain more than one parallax map stream.
- o A 3DS group MUST NOT contain more than one depth map stream.
- o A 3DS group MUST contain at least one 2D video stream.
- o If a 3DS group contains an L- and an R-stream, it MUST NOT contain a depth map or a parallax map.
- o If a 3DS group contains only one 2D video stream, it MUST also contain a parallax map stream or a depth map stream.
- o If a 3DS group contains a parallax map stream or a depth map stream, it MUST also contain a 2D video stream.

6. Combinations of attribute values and group usage

The following table summarises the possible combinations of attribute values and grouping:

	FP	SC	2DA
C			D/P, 3DS
CD			T
ChB	T		
CP			T
D			C/L, 3DS
L		R, 3DS	D/P, 3DS
LD			T
LIL	T		
LP			T
P			C/L, 3DS
R		L, 3DS	
SbS	T		
Seq	T		
TaB	T		

The table is to be read as follows:

- o The columns indicate <Format Type> values, whereas the rows indicate <Component Type> values.
- o For one particular column, we denote the <Format Type> value by "FT" and the <Component Type> value by "CT".
- o When an entry in the table is empty, it means that the corresponding combination of FT and CT is not allowed.

- o When an entry in the table contains a single <Component Type> value CTsec, it means that another stream with the <Component Type> value CTsec and the same <Format Type> value FT is needed.
- o When multiple <Component Type> values are listed, separated by a "/" symbol, only one secondary stream is needed, which must have one of the listed <Component Type> values, and the same <Format Type> value FT.
- o When an entry contains "3DS", it means that a 3DS group is needed.
- o When an entry in the table contains the letter "T" (true), it means that the corresponding combination FT and CT is allowed, that there is no required secondary stream, and that a 3DS group is not needed.

7. SDP offer/answer with 3D support

This section describes how the SDP offer/answer model (see [[RFC3264](#)]) can be used to negotiate the 3D format. It is assumed that both offerer and answerer are compliant to this document. The case where the answerer is a legacy answerer is described in [Section 8](#).

An example where the SDP offer/answer model can be used to negotiate the 3D format, is the case where the offerer offers two representations of the same stereoscopic 3D video: one frame packed and one as L/R simulcast. In this case, the answerer can select the format of its preference, according to its capabilities or as a trade-off between bandwidth and video quality.

There may also be cases where the answerer prefers to receive a 2D version, even when it supports stereoscopic 3D video and the "3dFormat" attribute. For example, this might happen when the user prefers to watch without glasses this time.

The following statements apply for the answerer:

- o The answerer **MUST NOT** omit the "3dFormat" attribute for the accepted streams. The answerer **MAY** omit the "3dFormat" attribute for the rejected streams.
- o The answerer **MUST NOT** change the value of the "3dFormat" attribute. This means, that the answerer can only choose between the 3D formats advertised in the offer.
- o In case the offer contains simulcast of the L- and R-view, the answerer **MAY** choose just one view. In this case, it **MUST** select only that view. This means that the port number of the other view **MUST** be set to zero in the answer.
- o In case the offer contains a 2D stream and an auxiliary stream as separate streams, the answerer **MAY** choose only the 2D stream. In this case, it **MUST** select the 2D stream, and **MUST NOT** select the auxiliary stream. This means that the port number of the auxiliary stream **MUST** be set to zero in the answer.
- o In case the offer contains a 2D stream and an auxiliary stream as a single stream, the answerer **MAY** choose to reject the stream by setting the port number in the answer to zero.
- o In case of frame packing, if the answerer prefers not to have frame packing, it **MUST** reject the stream by setting the port number in the answer to zero.

- o If the answerer selects multiple 3D formats, it MUST be prepared to send/receive (depending on whether it is a streaming server or client or both) associated streams simultaneously.

The following statements apply for the offerer:

- o The offerer MUST check if the "3dFormat" attribute is included in the answer. If it is not, it SHOULD handle the answer as described in [Section 8](#).
- o The offerer SHOULD list the 3D formats in order of preference.
- o When multiple 3D formats are selected, the offerer MAY initiate all associated streams. Alternatively, it MAY update its offer with a reduced number of 3D formats.
- o If all 3D formats have been rejected, the offerer MAY issue a new offer with 2D video instead.
- o If only an auxiliary stream is selected in the answer, the offerer SHOULD update its offer with only the associated 2D video stream. Alternatively, it MAY update its offer advertising another 3D format.

8. SDP offer/answer without 3D support

Since a legacy answerer does not support the "3dFormat" attribute, it might reject the offer. In this case the offerer MAY send a new offer with only a 2D video stream.

On the other hand, it is also possible that the legacy answerer accepts the offer but omits the "3dFormat" attribute in the answer. In this case the offerer is able to deduct that the answerer is a legacy answerer without 3D support. In the following subsections, we describe what the offerer still can do to provide a good user experience with a legacy answerer, for each of the 3D format styles. We assume that the offer was accepted, but a legacy answerer was detected.

8.1. Frame packing

In case the original offer contains frame packing, and the answer does not contain the "3dFormat" attribute, the offerer SHOULD treat that media stream as a 2D stream.

Note: in some cases, the answerer may be a legacy device that is capable of rendering a frame packed 3D stream, but does not understand the "3dFormat" attribute. For example, the user may be able to switch manually to 3D. Therefore, the server MAY stream the frame packed video as it is.

8.2. 2D and auxiliary as a single stream

If the original offer contains a 2D video and an auxiliary video in a single stream, and the answer does not contain the "3dFormat" attribute, the offerer SHOULD treat that media stream as a 2D stream.

8.3. 2D and auxiliary as two separate streams

When the offerer sends an offer to a legacy answerer, and the offer contains a 2D video and an auxiliary video in two separate streams, there are the following possibilities:

- o If the answerer selects only the 2D video stream then 2D video streaming can be done as agreed.
- o If the answerer selects only the auxiliary video, the offerer MAY treat that stream as a 2D video stream. If it does not, the offerer SHOULD update its offer without the auxiliary video.
- o If the answerer selects both video streams, but omits the "3dFormat" attribute, the offerer MAY update its offer without the

auxiliary video.

In case the offerer updates its offer by setting the port for auxiliary video to zero, it MUST NOT include the "3dFormat" attribute or use "3DS" grouping for the 2D stream.

8.4. Simulcast of L- and R-views

When the offerer sends an offer to simulcast the L- and R-view to the legacy answerer, we have the following possibilities:

- o If the answerer selects only one video stream, the offerer MAY stream the 2D video as agreed.
- o If the answerer selects both video streams, but omits the "3dFormat" attribute, the offerer MAY update its offer with only the L- or the R-stream.

In case the offerer updates its offer with only the L- or R-stream by setting one of the ports to zero, it MUST NOT include the "3dFormat" attribute or use "3DS" grouping for the offered stream.

9. Examples

9.1. One single frame compatible stream

The following is an example of an SDP description of a session which contains a single stream, in which the L- and R-streams are packed, in side by side fashion.

```
v=0
o=Alice 2890844526 2890842807 IN IP4 131.163.72.4
s=The technology of 3D-TV
c=IN IP4 131.164.74.2
t=0 0
m=video 49170 RTP/AVP 99
a=rtpmap:99 H264/90000
a=3dFormat:FP SbS
m=audio 52890 RTP/AVP 10
a=rtpmap:10 L16/16000/2
```

9.2. Two separate streams

The following is an example of an SDP description of a session with an audio stream, an L-stream and an R-stream.

```
v=0
o=Alice 2890844526 2890842807 IN IP4 131.163.72.4
s=The technology of 3D-TV
c=IN IP4 131.164.74.2
t=0 0
a=group:3DS 1 2
m=video 49170 RTP/AVP 99
a=rtpmap:99 H264/90000
a=3dFormat:SC L
a=mid:1
m=video 49172 RTP/AVP 101
a=rtpmap:101 H264/90000
a=3dFormat:SC R
a=mid:2
m=audio 52890 RTP/AVP 10
a=rtpmap:10 L16/16000/2
```

9.3. C-stream and depth map stream

The following is an example of an SDP description of a session with an audio stream, a C-stream and a depth map stream.


```
v=0
o=Alice 2890844526 2890842807 IN IP4 131.163.72.4
s=The technology of 3D-TV
c=IN IP4 131.164.74.2
t=0 0
a=group:3DS 1 2
m=video 49170 RTP/AVP 99
a=rtpmap:99 H264/90000
a=3dFormat:2DA C
a=mid:1
m=video 49172 RTP/AVP 101
a=rtpmap:101 H264/90000
a=3dFormat:2DA D
a=mid:2
m=audio 52890 RTP/AVP 10
a=rtpmap:10 L16/16000/2
```

9.4. Stereoscopic 3D video with two different formats

In the following example, there are two different formats for stereoscopic 3D video. One consists of stream 1 (C-stream) and stream 2 (parallax map stream), whereas the other consists of stream 3 (L-stream) and stream 4 (R-stream). There also is an audio stream, which can be used with both formats.


```
v=0
o=Alice 2890844526 2890842807 IN IP4 131.163.72.4
s=The technology of 3D-TV
c=IN IP4 131.164.74.2
t=0 0
a=group:3DS 1 2
a=group:3DS 3 4
m=video 49170 RTP/AVP 99
a=rtpmap:99 H264/90000
a=3dFormat:2DA C
a=mid:1
m=video 49172 RTP/AVP 101
a=rtpmap:101 H264/90000
a=3dFormat:2DA P
a=mid:2
m=video 49174 RTP/AVP 103
a=rtpmap:103 H264/90000
a=3dFormat:SC L
a=mid:3
m=video 49176 RTP/AVP 105
a=rtpmap:105 H264/90000
a=3dFormat:SC R
a=mid:4
m=audio 52890 RTP/AVP 10
a=rtpmap:10 L16/16000/2
```


10. Formal ABNF grammar of the "3dFormat" attribute

This section contains the formal ABNF grammar of the "3dFormat" attribute.

```
3dFormat-attribute      = "a=3dFormat:" formatType componentType
formatType              = "FP"/"SC"/"2DA"/formatType-extension
formatType-extension    = token
componentType           = "C"/"CD"/"ChB"/"CP"/"D"/"L"/"LD"/
                        "LIL"/"LP"/"P"/"R"/"SbS"/"Seq"/"TaB"/
                        componentType-extension
componentType-extension = token
```


11. Security Considerations

The authors foresee no security issues in addition to those already listed in [[RFC4566](#)].

12. IANA Considerations

12.1. "3dFormat" attribute

Following the guidelines in [[RFC4566](#)], the SDP attribute has to be registered at IANA:

- o Contact name/email: authors of this RFC
- o Attribute name: 3dFormat
- o Long-form attribute name: Attribute for signalling the format of a stereoscopic 3D video carried in the media stream(s).
- o Type of attribute: media level
- o Subject to charset: no

The "3dFormat" SDP media-level attribute is used to signal the format of stereoscopic 3D video, carried in one or more media stream(s).

The attribute has the following syntax:

a=3dFormat:<Format Type> <Component Type>

The <Format Type> indicates the format type of the stereoscopic 3D video carried in the media stream(s). It indicates whether the stereoscopic 3D video is frame packed, simulcast or consists of a 2D video stream and an auxiliary stream. The <Format Type> can have the following values (as indicated between the quotes):

"FP"	frame packed
"SC"	simulcast
"2DA"	2D + auxiliary

The <Component Type> indicates the type of the video component, which is a constituent element of the stereoscopic 3D video. It can have the following values:

"C"	centre view
"CD"	centre view and depth map
"ChB"	checkerboard
"CP"	centre view and parallax map
"D"	depth map
"L"	left view
"LD"	left view and depth map
"LIL"	line interleaved
"LP"	left view and parallax map
"P"	parallax map
"R"	right view
"SbS"	side by side
"Seq"	frame sequential
"TaB"	top and bottom

12.2. "3DS" value for "group" semantics

Following the standards action policy from [[RFC5226](#)], the following semantics have to be registered with IANA in the "Semantics for the "group" SDP Attribute" registry under "SDP Parameters":

+-----+	+-----+	+-----+
Semantics	Token	Reference
+-----+	+-----+	+-----+
3D synchronised	3DS	this RFC
+-----+	+-----+	+-----+

13. Acknowledgements

The authors would like to thank Stephen Botzko, Imed Bouazizi, Pedro Capelastegui, Roni Even, Miguel Garcia, Ted Hardie, Jonathan Lennox, Yue Peiyu and Tian Linyi for their review comments.

14. Normative References

- [HDMIv1.4a]
HDMI, "HDMI Specification Version 1.4a", March 2010.
- [ISO/IEC 23002-3]
MPEG, "MPEG video technologies part 3: Representation of auxiliary video and supplemental information", ISO/IEC FDIS 23002-3:2007(E), December 2002.
- [ISO/IEC 14496-10]
MPEG, "H.264/MPEG-4 Part 10: Advanced video coding for generic audiovisual services", ISO/IEC FDIS 14496-10:2010, March 2010.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC3264] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", [RFC 3264](#), June 2002.
- [RFC4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", [RFC 4566](#), July 2006.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", [BCP 26](#), [RFC 5226](#), May 2008.
- [RFC5888] Camarillo, G. and H. Schulzrinne, "The Session Description Protocol (SDP) Grouping Framework", [RFC 5888](#), June 2010.

Authors' Addresses

Bert Greevenbosch
Huawei Technologies Co., Ltd.
Huawei Industrial Base
Bantian, Longgang District
Shenzhen 518129
P.R. China

Phone: +86-755-28978088
Email: bert.greevenbosch@huawei.com

Hui Yu
Huawei Technologies Co., Ltd.
Huawei Nanjing R&D Center
101 Software Avenue
Yuhuatai District
Nanjing 210012
P.R. China

Phone: +86-25-56620323
Email: huiyu@huawei.com

