

mmusic
Internet-Draft
Intended status: Standards Track
Expires: November 1, 2012

P. Capelastegui
Universidad Politecnica de
Madrid
April 30, 2012

3D Video in the Session Description Protocol (SDP)
draft-capelastegui-mmusic-3dv-sdp-00

Abstract

This document defines a mechanism to describe 3D video streams in the Session Description Protocol (SDP). This includes 3D video streams composed of multiple video views, or of a combination of views and depth maps. Several 3D video formats are supported, including simulcast, video-plus-depth, and frame-packing.

A new decoding dependency, "3dd", is defined, describing the association between media stream belonging to a 3D video stream. In addition, a new SDP media-level attribute, "3dvFormat", is defined, describing the format used by media streams within a 3D video stream.

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 November 1, 2012.

Copyright Notice

Copyright (c) 2012 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	3
1.1. Requirements Language	3
2. Definitions	3
3. Decoding dependency of 3D video streams	4
4. The "3dvFormat" media attribute	5
4.1. The "depth-map-simulcast" 3D format attribute	5
4.2. The "depth-map-metadata" 3D format attribute	6
4.3. The "stereo-view" 3D format attribute	7
4.4. The "frame-pack" 3D format attribute	8
5. Usage with SDP offer/answer model	9
5.1. Backward compatibility	10
6. Examples	10
6.1. Example session with single 3D video option	10
6.2. Test Scenario: Multiple 3D options	11
7. Formal Grammar	13
7.1. "3dvFormat" media attribute	13
7.2. "depth-map-simulcast" 3D format attribute	13
7.3. "depth-map-metadata" 3D format attribute	13
7.4. "stereo-view" 3D format attribute	13
7.5. "frame-pack" 3D format attribute	13
8. Security Considerations	13
9. IANA Considerations	13
10. Normative References	14
Author's Address	15

1. Introduction

3D video applications convey depth information by showing a different view for each eye of a user. In order to achieve this, 3D video streams need to include additional information compared to conventional 2D video streams, either in the form of extra views, or auxiliary maps such as depth maps, or a combination thereof. These views and maps can be transported in a variety of ways, including, among others: as separate RTP streams (simulcast), frame-packed in a single video stream [HDMIV1.4a], or using the video-plus-depth format [ISO/IEC 23002-3].

The Session Description Protocol (SDP) [RFC4566] lacks the means to describe neither of these transport techniques for 3D video. This document extends SDP to support the description of multimedia sessions using 3D video encapsulated as simulcast streams, using frame-packing techniques, or using the video-plus-depth format.

[RFC5583] defines a mechanism to signal the decoding dependency of media descriptions in SDP. This document extends that mechanism by defining a new SDP decoding dependency type, '3dd', describing the association between media streams belonging to a 3D video stream. In addition, a new SDP media-level attribute, '3dvFormat', is defined to describe the format used by media streams composing a 3D video stream. Several formats for 3D video are described in this specification, including simulcast stereo video, simulcast video and depth map, various frame-packing schemes, and streams using video-plus-depth.

1.1. Requirements Language

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 RFC 2119 [RFC2119].

2. Definitions

3D video stream: A video stream that conveys depth information by showing different perspectives of a scene to each eye of an observing user. A 3D video stream is typically composed of multiple video streams ('views'), or a combination of video streams and auxiliary data maps such as depth maps

2D video stream: A video stream that lacks 3D depth information.

View: A video stream that represents a specific point of view of the scene in a 3D video stream.

Depth Map: An auxiliary data stream that associates a Z-value to each pixel of a view within a 3D video stream. Depth maps are often encoded as grey scale video streams.

Simulcast: A method for the transmission of 3D video streams that consists on sending additional views and auxiliary data streams as a separate RTP streams.

Video-plus-depth: (also known as MPEG-C Part 3) A method for the transmission of 3D video streams consisting on encapsulating a depth map as metadata within a 2D video stream. This mechanism, standardized in [ISO/IEC 23002-3], is compatible with MPEG-2 and H.264/AVC, and allows for backwards compatibility.

Frame-packing: A method for the transmission of 3D video streams that consists on multiplexing several views and/or auxiliary data within a single video stream, using either spatial multiplexing or time multiplexing. Frame-packing is supported by standards like [HDMIV1.4a] and [ITU-T H.264].

3. Decoding dependency of 3D video streams

The "depend" SDP attribute, defined in [RFC5583] describes the decoding dependency between two or more media descriptions. This specification defines a new dependency type for the "depend" attribute:

- o 3dd: 3D video dependency - indicates that the described media stream belongs to a 3D video stream, and requires other media streams to render the 3D video. When "3dd" is used, all required media streams for the Operation Point MUST be identified by identification-tag and fmt-dependency following the "3dd" string.

Like other dependency types, 3dd is used in combination with the "DDP" grouping semantic, which is defined in [RFC5583], and based on the SDP grouping framework [RFC5888]. Whenever a 3D video stream is composed of multiple media descriptions, these media descriptions MUST be included in the same DDP group.

The media decoding dependency terminology defined in [RFC5583] can be applied to 3D video streams as follows:

- o **Media Bitstream:** A 3D video stream is considered a Media Bitstream for the purposes of 3dd decoding dependency.
- o **Media Partition:** Each separate media description composing a 3D video stream is considered a Media Partition. Note that each Media Partition usually contains a single video view or depth map, but can also include multiple of views/maps, e.g. when using frame-packing techniques.
- o **Operation Point:** A subset of a 3D Media Bitstream that includes all Media Partitions required for reconstruction at a certain point of quality, number of views or depth maps, or other property. Note that a valid Operation Point for a 3D Media Bitstream can be a 2D video lacking any depth information.

4. The "3dvFormat" media attribute

a=3dvFormat:<fmt> <attribute>:<value>

This section defines a new media-level attribute for SDP, "3dvFormat", which can be used to describe the transport format of a media stream in a 3D video stream. This attribute can indicate that a media description corresponds to a specific view within a 3D stream, or to a depth map, or to a combination of views or depth maps encapsulated with frame-packing techniques or with the video-plus-depth mechanism.

A media description can have multiple "3dvFormat" attributes; each attribute is mapped to a media format specified for the media, indicated by <fmt>. Only one "3dvFormat" attribute is allowed per media format.

Each "3dvFormat" attribute indicates a property (known as a "3D format attribute") associated to a media format of its media description. The 3D format attribute consists on an attribute-value pair, with the form "<attribute>:<value>". This specification defines four 3D format attributes: "depth-map-simulcast", "depth-map-metadata", "stereo-view", "and frame-pack".

New 3D format attributes can be defined, but they MUST be registered with IANA, following the registry described in Section 9.

4.1. The "depth-map-simulcast" 3D format attribute

a=3dvFormat:<fmt> depth-map-simulcast:<associated_video>

The 3D format attribute "depth-map-simulcast" indicates that a media stream represents a depth map associated with a view within the same 3D video stream. A depth map described by this attribute is

transmitted as a separate transport stream from its corresponding view.

<associated-video> is the media stream identification (the "a=mid" attribute, as defined in [RFC5888]) of the video stream associated with this depth map.

A media description with the "depth-map-simulcast" 3D format attribute MUST be included in a DDP group. This group MUST include a video stream representing the view associated with the depth map. Finally, the depth map media description MUST include a "depend" attribute with the "3dd" dependency type, indicating dependency to one or more media formats within that video stream.

Example:

```
a=group:DDP 1 2
m=video 1111 RTP/AVP 99
a=rtpmap:99 H264/90000
a=mid:1
m=video 1112 RTP/AVP 99
a=rtpmap:99 H264/90000
a=3dvFormat:99 depth-map-simulcast:1
a=mid:2
a=depend:99 3dd 1:99
```

The example shows two media descriptions forming a 3D video stream, of which the first one (mid:1) represents a video view, and the second one (mid:2) the depth map for that view. The depth map cannot be used without its corresponding view, and this is reflected in the "depend" attribute.

4.2. The "depth-map-metadata" 3D format attribute

```
a=3dvFormat:<fmt> depth-map-metadata:<associated_video>
```

The 3D format attribute "depth-map-metadata" indicates that a media stream represents a depth map associated with a view within the same 3D video stream. A depth map described by this attribute is transmitted as part of the same transport stream as its corresponding view, in the form of metadata. If the view associated with this depth map is a MPEG-2 or H.264/AVC video stream, the depth map follows the format defined in MPEG-C part 3 [ISO/IEC 23002-3].

<associated-video> is the media stream identification (the "a=mid" attribute, as defined in [RFC5888]) of the video stream associated with this depth map.

A media description with the "depth-map-simulcast" 3D format attribute MUST be included in a DDP group. This group MUST include a video stream representing the view associated with the depth map. Finally, the depth map media description MUST include a "depend" attribute with the "3dd" dependency type, indicating dependency to that video stream.

It is important to note that, when a media format with a "depth-map-metadata" is used, the transport information for that media stream such as port, connection address or transport protocol MUST be ignored. In this case, the depth map is transmitted as part of the media stream of its associated view, rather than as a separate stream.

Example:

```
a=group:DDP 1 2
m=video 1111 RTP/AVP 99
a=rtpmap:99 H264/90000
a=mid:1
m=video 1112 RTP/AVP 99 100
a=rtpmap:99 H264/90000
a=3dvFormat:99 depth-map-simulcast:1
a=rtpmap:100 H264/90000
a=3dvFormat:100 depth-map-metadata:1
a=mid:2
a=depend:99 3dd 1:99; 100 3dd 1:99
```

The example shows two media descriptions forming a 3D video stream, of which the first one (mid:1) represents a video view, and the second one (mid:2) the depth map for that view. Two possible configurations for the depth map are offered, one using simulcast (payload type 99), and the other transmitting the depth map as metadata (payload type 100). If the depth map stream is configured as metadata, the port specified in that media description (1112) will be ignored, since the depth map will be transmitted within the video view stream. On the other hand, if the simulcast option is used, the depth map will be transmitted as a separate stream using the specified port and transport, as usual.

4.3. The "stereo-view" 3D format attribute

```
a=3dvFormat:<fmt> stereo-view:<view-type>
```

The 3D format attribute "stereo-view" indicates whether a video stream is associated with the left-eye view or the right-eye view of a stereo 3D video stream.

<view-type> indicates which view is associated with the media stream. It can have the value "left", for the left-eye view, or "right", for the right-eye view.

A media description with the "stereo-view" 3D format attribute MUST be included in a DDP group. This group MUST also include another video stream containing the "stereo-view" 3D format attribute with the other stereo view as value. The media description for either of the two stereo views MUST include a "depend" attribute with the "3dd" dependency type, indicating dependency to the stream corresponding to the other view.

Example:

```
a=group: DDP 1 2
m=video 1111 RTP/AVP 99
a=rtpmap:99 H264/90000
a=3dvFormat:99 stereo-view:left
a=mid:1
m=video 1112 RTP/AVP 99
a=rtpmap:99 H264/90000
a=3dvFormat:99 stereo-view:right
a=mid:2
a=depend:99 3dd 1:99
```

The example shows two media descriptions forming a stereo 3D video stream, of which the first one (mid:1) represents the left view, and the second one (mid:2) the right view. This Media Bitstream can be configured as a 3D video stream composed of two stereo views, or as a 2D video stream including just the left eye view.

4.4. The "frame-pack" 3D format attribute

```
a=3dvFormat:<fmt> frame-pack:<fp-format>
```

The 3d attribute indicates that frame-packing mechanisms are used in a media stream, for the specified media format.

<fp-format> signals which frame-packing mode is applied. It has three possible values: "side-by-side", "top-bottom", and "frame-seq".

Of these frame-pack modes, the first two are based on spatial multiplexing, or dividing each video frame in the stream into two sub-frames, and assigning one view to each sub-frame. In "side-by-side" mode, the left sub-frame corresponds to the left eye view, and the right sub-frame to the right eye view. In "top-bottom" mode, the top sub-frame corresponds to the left eye view, and the lower sub-frame to the right eye view.

On the "frame-seq" (frame sequential) frame-packing mode, time multiplexing is used, so that half the video frames in a stream correspond to the left eye view, and the other half to the right eye view, in alternating order. In order to identify which frame corresponds to each view, additional signalling is required; in H.264/AVC video streams, this is achieved through supplemental enhancement information (SEI) metadata [ITU-T H.264].

5. Usage with SDP offer/answer model

When the extensions defined in this specification are used in the SDP offer/answer model [RFC3264], the following rules apply.

The offerer MAY include more than one "3dvFormat" attribute per media description, and the values of these "3dvFormat" can be different or duplicated. However, each media format MUST NOT have more than one "3dvFormat" attribute.

If the offerer includes a 3D video stream composed of more than one media description, all media descriptions in the stream MUST be included in a DDP group. If the 3D video stream includes streams with 3D format attributes whose description specifies any stream requirements or mandatory dependencies, those requirements or dependencies MUST be respected. Each 3D video stream in the offer SHOULD have at least one Operation Point consisting on a single 2D video stream, as well as any number of Operation Points with 3D video.

An answer MUST NOT include any "3dvFormat" attribute that is not present in the offer.

When a media format in an offered media description has a "3dvFormat" attribute, if the answer contains that media format it MUST also include the "3dvFormat" attribute, with the same parameters as the offer.

To simplify the processing of 3D video configurations, when the answer includes a "3dvFormat" attribute in a media description, the same RTP payload type number used in the offer should also be used in the answer, and the answer MUST NOT include more than one media format for that media description.

If the answerer understands the DDP semantics, it is necessary to take the "depend" attribute into consideration in the Offer/Answer procedure, as indicated in [RFC5583]

5.1. Backward compatibility

Depending on implementation, a node that does not understand DDP grouping or "3d" attributes SHOULD respond to an offer using this grouping or attributes either with a refusal to the request, or with an answer that ignores the grouping or 3D video format attributes.

In case of a refused request, if the offerer has identified that the refusal of the request is caused by the use of 3D video, and it still wishes to initiate a session, it SHOULD generate a new offer without any 3D video streams.

If the request is accepted but the answer is ignoring the grouping attribute, the "depend" attribute, or a "3dvFormat", it should be assumed that the answerer is unable to send or receive 3D video streams. If the offerer still wishes to initiate a session, it SHOULD generate a new offer without any 3D video streams. Alternatively, if the answer does not include more than a single video stream, the offerer MAY initiate the session without generating a new offer, and send and receive that stream as a 2D video stream.

6. Examples

The following examples show SDP Offer/Answer exchanges for sessions with 3D video streams. Only the media descriptions and grouping attributes of the SDP are shown. For each example, two possible answers are considered: one in which the answering device is compatible with this specification, and one with a legacy answering device.

6.1. Example session with single 3D video option

The example shows a session where the 3D video stream is transmitted over a single media stream, so no grouping or decoding dependencies are needed for the SDP. The calling user agent makes a SDP offer with 2 options for configuring the 3D video stream:

- o 2D video stream
- o Single frame-packed video stream, with 2 views multiplexed side-by-side

Offer SDP:

```
m=video 1111 RTP/AVP 99 100
a=rtpmap:99 H264/90000
```

```
a=rtpmap:100 H264/90000
a=3dvFormat:100 frame-pack:side-by-side
```

Answer SDP:

```
m=video 2222 RTP/AVP 100
a=rtpmap:100 H264/90000
a=3dvFormat:100 frame-pack:side-by-side
```

The initial offer includes a media description with two media formats, with one corresponding to a 2D video stream (payload type 99) and the other to a frame-packed 3D video stream (payload type 100). Of these, the answering device chooses the frame-packed media format.

Alternate Answer SDP (legacy device):

```
m=video 2222 RTP/AVP 100
a=rtpmap:100 H264/90000
```

If this SDP offer is received by a legacy device and the session is not rejected, the answer will ignore any 3D video format attributes. In this case, the offerer can initiate the session treating the selected media format as a 2D video stream.

6.2. Test Scenario: Multiple 3D options

The example shows a session where the 3D video stream is transmitted over up to two media streams, and several options for the format of the 3D video stream are offered:

- o 2D video stream
- o Single frame-packed video stream, with 2 views multiplexed side-by-side
- o Single video stream including a depth map as metadata
- o 2 Simulcast streams, with video and depth map
- o 2 Simulcast streams, with 2 stereo views.

Offer SDP:

```
a=group:DDP 1 2
m=video 1111 RTP/AVP 99 100
a=rtpmap:99 H264/90000
a=3dvFormat:99 stereo-view:left
a=rtpmap:100 H264/90000
a=3dvFormat:100 frame-pack:side-by-side
a=mid:1
```

```
m=video 1112 RTP/AVP 99 100 101
a=rtpmap:99 H264/90000
a=3dvFormat:99 depth-map-metadata:1
a=rtpmap:100 H264/90000
a=3dvFormat:100 depth-map-simulcast:1
a=rtpmap:101 H264/90000
a=3dvFormat:101 stereo-view:right
a=mid:2
a=depend:99 3dd 1:99; 100 3dd 1:99; 101 3dd 1:99
```

Answer SDP:

```
a=group:DDP 1 2
m=video 2222 RTP/AVP 99
a=rtpmap:99 H264/90000
a=3d:99 stereo-view:left
a=mid:1
m=video 2223 RTP/AVP 102
a=rtpmap:101 H264/90000
a=3d:101 stereo-view:right
a=mid:2
a=depend:101 3dd 1:99
```

The initial offer includes two media descriptions, the first of which (mid 1) can be transmitted independently, either as a 2D video stream (payload type 99) or as a frame-packed 3D stream (payload type 100). The second media description (mid 2), on the other hand, depends on the first one for all its media formats, and can be configured as a depth map transmitted as metadata (payload type 99), as a simulcast depth map stream (payload type 100), or as a right-eye stereo view (payload-type 101). The answering device chooses the configuration with 2 simulcast stereo views.

Alternate Answer SDP (legacy device)

```
m=video 2222 RTP/AVP 99
a=rtpmap:99 H264/90000
m=video 0 RTP/AVP 102
a=rtpmap:101 H264/90000
```

If this SDP offer is received by a legacy device and the session is not rejected, the answer will ignore any 3D video format attributes, as well as the grouping and dependency attributes. In the example above, the answering device has selected a media format for the first video stream, and disabled the second video stream. In this case, the offerer can initiate the session treating the selected media format as a 2D video stream. If the second video stream had not been disabled, the offerer should send a new offer with a single video

stream.

7. Formal Grammar

The 3d attributes defined in this document use the following Augmented Backus-Naur Form (ABNF) [RFC5234] grammar.

7.1. "3dvFormat media attribute

```
3dvformat-attribute = "a=3dvFormat:" fmt SP 3dvf-type  
; fmt is described in [RFC4566]  
; fmt is media format (usually RTP payload type)
```

```
3dvf-type= depth-scast / depth-meta / st-view / f-pack
```

7.2. "depth-map-simulcast" 3D format attribute

```
depth-scast = "depth-map-simulcast:" identification-tag  
; identification-tag is defined in [RFC5888]
```

7.3. "depth-map-metadata" 3D format attribute

```
depth-meta = "depth-map-metadata:" identification-tag  
; identification-tag is defined in [RFC5888]
```

7.4. "stereo-view" 3D format attribute

```
st-view = "stereo-view:" view-type  
view-type= "left" / "right"
```

7.5. "frame-pack" 3D format attribute

```
f-pack = "frame-pack:" fp-format  
fp-format= "side-by-side" / "top-bottom" / "frame-seq"
```

8. Security Considerations

No security issues have been identified for this specification.

9. IANA Considerations

The following contact information shall be used for all registrations included here:

Contact: Pedro Capelastegui
 email: Capelastegui@dit.upm.es
 tel: +34 915 49 57 00 ext. 3024

This document defines the following new semantics for the "depend" SDP attribute. The semantics are registered by IANA under "depend" SDP Attribute Values under "Session Description Protocol (SDP) Parameters":

Token	Semantics	Reference
3dd	3D video dependency	[THIS DOC]

This document defines a new media-level SDP attribute, "3dvFormat". The attribute is registered by IANA under "Session Description Protocol (SDP) Parameters" under "att-field (media level only)".

Attribute name: 3dvFormat
 Long form: 3D video format
 Type of name: att-field
 Type of attribute: media level only
 Subject to charset: no
 Purpose: [THIS DOCUMENT]
 Reference: [THIS DOCUMENT]
 Values: see this document and registrations below

Parameters of the "3dvFormat" SDP attribute MUST be registered under IANA following the "Specification Required" policy [RFC5226]. This document creates a new IANA registry called [REF-1] within the "Session Description Protocol (SDP) Parameters" registry, for that purpose.

The initial entries in the registry are shown below.

Token	Description	Reference
depth-map-simulcast	depth map as separate stream	[THIS DOC]
depth-map-metadata	depth map as metadata	[THIS DOC]
stereo-view	left or right stereo view	[THIS DOC]
frame-pack	frame-packed video stream	[THIS DOC]

10. Normative References

[HDMIv1.4a]
 HDMI, "HDMI Specification Version 1.4a", March 2010.

- [ISO/IEC 23002-3]
ISO/IEC JTC1/SC29/WG11, "ISO/IEC FDIS 23002-3
Representation of Auxiliary Video and Supplemental
Information", Doc. N8768, January 2007.
- [ITU-T H.264]
HDMI, "Advanced video coding for generic audiovisual
services", ITU-T Recommendation H.264 and ISO/
IEC 14496-10, 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.
- [RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax
Specifications: ABNF", STD 68, RFC 5234, January 2008.
- [RFC5583] Schierl, T. and S. Wenger, "Signaling Media Decoding
Dependency in the Session Description Protocol (SDP)",
RFC 5583, July 2009.
- [RFC5888] Camarillo, G. and H. Schulzrinne, "The Session Description
Protocol (SDP) Grouping Framework", RFC 5888, June 2010.

Author's Address

Pedro Capelastegui
Universidad Politecnica de Madrid
ETSI Telecomunicacion
Avenida Complutense, 30
Despacho B.203
Madrid 28040
Spain

Phone: +34 915 49 57 00 ext. 3024
Email: capelastegui@dit.upm.es

