Mapping On2 VP8 in Ogg
Version 1.0.0

Sebastian Dröge, Collabora Multimedia

May 20, 2010

Contents

1 Scope

This specification defines how On2 VP8 video streams are mapped to the Ogg container format.

2 Normative References

RFC 3533, *The Ogg Encapsulation Format Version 0*

Ogg Vorbis I format specification, *Comment field and header specification*

On2 VP8, *Data Format and Decoding Specification*

3 Logical Stream Identification

Ogg logical streams start with a BOS page, which must contain information to uniquely identify the codec type of the logical stream and should contain additional information about the encoded media. It must additionally contain all necessary timing information to convert the granule position (see section 5.3) into a timestamp.

For the On2 VP8 Ogg mapping the BOS page must have the following content:

Name	Size (bits)	Description
HDRID	8	Identifier for Ogg VP8 mapping header packets (`0x4F`)
ID	32	Identifier for the Ogg VP8 mapping (`0x56503830` or "VP80")
HDRTYP	8	Ogg VP8 stream info header type (`0x01`)
VMAJ	8	Mapping major version, used to version incompatible On2 VP8 mappings. This specification defines major version 1.
VMIN	8	Mapping minor version, used to version compatible On2 VP8 mappings. This specification defines minor version 0.
FW	16	Stored frame width.
FH	16	Stored frame height.
PARN	24	Pixel aspect ratio numerator.
PARD	24	Pixel aspect ratio denominator.
FPSN	32	Frame rate numerator.
FPSD	32	Frame rate denominator.

The BOS page might be followed by a single additional header page containing VORBISCOMMENT metadata for that logical stream. The structure of VORBISCOMMENT metadata is defined in the Ogg Vorbis I format specification. The identification of the VP8 VORBISCOMMENT must be the *HDRID* and *ID* as used for the BOS page content, followed by the *HDRTYP* `0x02` and followed by a space, i.e. `0x4F 0x56 0x50 0x38 0x30 0x01 0x20`.

Any future headers will also start with the *HDRID* and *ID*.

Note: All integers are unsigned integers stored in big-endian format, i.e. most significant byte first.

Note: The total size of the version 1.0 BOS page content is 26 bytes but newer minor versions of this specification might append new fields to this structure.

Note: The BOS page content and the VORBISCOMMENT packet can be placed at other places in the stream. The first byte can't be created by the VP8 encoder at this point and most likely will never be created. If this happens in the future version 2.0 of this specification has to be released.

4 Logical Stream Termination

An On2 VP8 logical stream inside the Ogg container terminates with an EOS page, which might contain zero, one or multiple On2 VP8 packets (see section 5.1).

5 Storage

5.1 Packets

The main unit in Ogg streams are packets, which can be of arbitrary size. Every Ogg packet of an On2 VP8 logical stream must contain exactly one VP8 encoded frame, no matter if the frame is a visible or invisible VP8 frame. A packet must not contain any additional data, multiple VP8 frames or parts of one or more VP8 frames.

The first packet after the BOS page must contain a single VP8 keyframe.

Packets must be placed in the logical stream in coding order, which is equal to presentation order for VP8.

5.2 Pages

Ogg Pages combine one or more packets of a single logical stream with header information, containing timing information, a checksum and other metadata.

On2 VP8 keyframe packets should begin a new Ogg page to allow simpler random access and might span multiple pages if necessary but nonetheless should be the only packet in a page. Putting keyframes together with other frames into a single Ogg pages is allowed but might work suboptimal with some software.

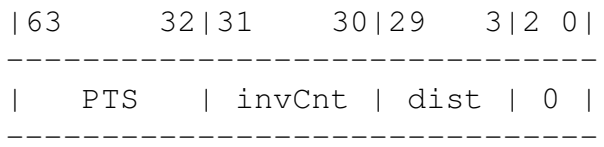
5.3 Granule Position

The granule position is a synthetic value, that combines timing information, distance to the last keyframe and other information. Every Ogg page has a single granule position value, which corresponds to the last complete packet in this Ogg page.

For On2 VP8 logical streams the granule position contains the following fields:

- *PTS*: **End** presentation timestamp (as frame count) of the last complete packet in the Ogg page. If the last complete packet contains an invisible frame this is the timestamp of the end of the next visible frame. This value increases by one per visible frame.
- *invCnt*: Number of invisible VP8 frames since the last visible frame. A visible frame gets the special value -1 (in two's complement).
- *dist*: Distance in packets (including invisible frames) to the last keyframe.

These values are to be placed as follows into the granule position field:



Note: The meaning of the 3 least significant bits is undefined and might be defined in later versions of this specification.

Note: The invCount for visible frames has the special value -1 instead of 0 to create a monotonically increasing granulepos value.