DJI Media File Metadata White Paper

Part 3: Metadata in Video

Version 1.1

2020-05-07

The third-party agreements, methods and specifications described or used in this document are owned by the original owner; in the process of using this document, the developer needs to carry out the DJI and the use requirements specified by the above-mentioned owners at the same time; The problems caused by the use of the above content without authorization are not related to DJI, and DJI does not bear any legal responsibility related to them. This document are copyrighted by DJI with all rights reserved. No part of this document shall be reproduced in any form without prior written consent of or authorization from DJI.

Disclaimer

This document does not contain any forecast information, also with DJI's technical, product and Market. In the process of using this document, please consciously abide by the relevant requirements of local laws and regulations. Any legal problems caused by violation of local laws are not related to DJI.

DJI will not acquire the proprietary data of developers or users in any form or any form. Developers are requested to use encryption and other methods to protect various data and personal information during the process of acquiring and using media files metadata.

DJI is a trademark of SZ DJI TECHNOLOGY CO., LTD. (abbreviated as “DJI”) and its affiliated companies. Names of products, brands, etc., appearing in this document are trademarks or registered trademarks of their respective owner companies.

Contents

Version History

Version	Date	Description
1.1	2020-05-07	First release version based on version 1.0 of DJI Media File Metadata Whitepaper Part 1.

1. Introduction

1.1 Organization of Document Set

The metadata whitepaper has three parts:

- *Part 1, Overview*, introduce the basic concept and organization of metadata in media files product by DJI. It's an overview of what kinds of metadata we provide and the basic representation model for metadata in media files.
- *Part2, Metadata in still-image*, describe metadata specifications and contents in the still-image. This part mainly contains metadata in Exif and XMP standards, both metadata follow industrial standards and extend by DJI to give more rich information.
- *Part3, Metadata in Video*, describe metadata in video contains both metadata for whole video (global metadata) and metadata carried on every video frame (timed metadata). DJI introduce our own data model and specification in this part for video timed metadata.

1.2 About This Document

This document, *DJI Media File Metadata Whitepaper Part 3: Metadata in Video*, describe the global and timed metadata in DJI video files. We introduce DJI new standard to store video timed metadata which can extend very rich metadata for various applications. From this document you can find that the metadata organization in video files and how to access those metadata to your own application.

2. Global metadata in MP4/MOV

MP4 and MOV formats allow storage user-defined metadata in the atom structure. We exploited this feature to add several global metadata atoms in the `udta` atom, which is sub-atom under the top-level `moov` atom.

NOTICE: This is not universal to all existing DJI products. Please check the udta atom formats if your software depends on these metadata. But we will keep the compatibility from metadata was added to a product.

2.1.1 Camera Serial Number Atom

Camera Serial Number Atom (defined by the `©csn` atom type) stores product serial number as a NULL terminated string.

2.1.2 Unique ID Atom

Unique ID Atom (defined by the `©uid` atom type) stores a 32-bit unique ID of this video file.

2.1.3 Model Name Atom

Model Name Atom (defined by the `©mdl` atom type) stores product model name as a NULL terminated string.

2.1.4 Resolution Atom

Resolution Atom (defined by the `©res` atom type) stores various video media information contains the following data elements.

Resolution Width

A 32-bit integer that specifies the width of resolution in pixel.

Resolution Height

A 32-bit integer that specifies the height of resolution in pixel.

Frame Rate

A 32-bit integer that specifies the frame rate of video. This element is deprecated since it can only express an integer frame rate. You can get frame rate by using tools like FFmpeg or calculate it by checking total frames and total durations in the `stts` atom.

TV System

A 2-byte string that specifies the TV system of video, "N\0" for NTSC system, "P\0" for PAL system.

Video Type

An 8-bit integer that specifies the type of video. Following types is available:

- 0 = Normal video
- 2 = Slow motion video
- 3 = Timelapse video
- 4 = HDR video
- 5 = Hyperlapse video
- 6 = Loop video

Speed Ratio

An 8-bit integer that specifies the playing speed ratio of video. For example, this value means how many times slow down on a slow motion video or how many times speed up on a hyperlapse video.

3. DJI Video Timed Metadata Format

Recording metadata for every video frame provides more powerful ability to post processing and analyzing. For example, recording IMU data for every frame make offline EIS possible. GPS and speed data can be used to generate track map or speedometer by post processing.

There are some solutions for video timed metadata storage used by difference manufacturers. Based on requirements on performance, extensibility, and developer-friendly, we released our new DJI Video Timed Metadata Format (DVTM for short). We believe that the DVTM format can energize developers with metadata.

3.1 Metadata format based on Google protocol buffers

Protocol Buffers is a language and platform independent data serialization scheme published by Google. High efficiency and flexible of Protocol buffers are suitable for metadata storage. As a developer, you can use API of various languages provide by protocol buffers to access our metadata including Java, Python, Objective-C, C++, Dart, Go, Ruby, C#.

It's easy to adapt your exist platform or favorite language. The compatible and extensible design of Protocol Buffers helps us to add more and more metadata in the future, unleash our productivity to developers.

Video timed metadata of all our products is described in `dvtm_library.proto`, which act a library of all type of metadata defined by DJI. Each product or product line adapts different subset named `dvtm_product.proto` to specify what kind of metadata is supported with it.

Protocol buffers provide a compiler called `protoc` which can translate those `.proto` files into different language files, so that you can use corresponding language APIs to decode. Protocol buffers has two main version, 2 and 3. We choose protocol buffers 3 for reasons such as more languages support.

3.2 Integrate with MP4/MOV file format

Like video and audio, metadata in MP4 (MPEG-4 part 12: ISO base media file format, ISO/IEC 14496-12) or MOV (QuickTime File Format) file is also organized as a track, called `meta` track, which is a collection of frames under a specific rate. You can read metadata frames by parsing `moov/trak/mdia/minf/stbl` boxes just like read video frames. In most cases, `meta` track has same frame rate with video track, but not always.

Every frame in `meta` track stores a message described by `DvtmProductName`. Every sub-message in it is optional and repeatable due to the property of the protocol buffers 3, so that a fixed frame rate of `meta` track can support both higher or lower frequency metadata. For

example, we can store a 60 Hz sensor data into the `meta` track of 30 FPS, because this sensor data can be repeated stored for 2 times in a frame. And a 15 Hz sensor data can also be embedded in 30 FPS `meta` track, because data is interleaved stored on metadata frame.

4. References

1. https://developer.apple.com/library/archive/documentation/QuickTime/QTFF.
2. https://developers.google.com/protocol-buffers.

