**AUTOHIGHLIGHT ALGORITHM INTEGRATION**

**ON TASMANIA**

# Purpose

This document provides the overview of the Autohighlight algorithms (Engines) that needs to be integrated to Tasmania platform through Qualcomm’s video analytics manager (VAM) framework. The Auto highlight algorithm is based on the captured audio and IMU data.

# Requirements

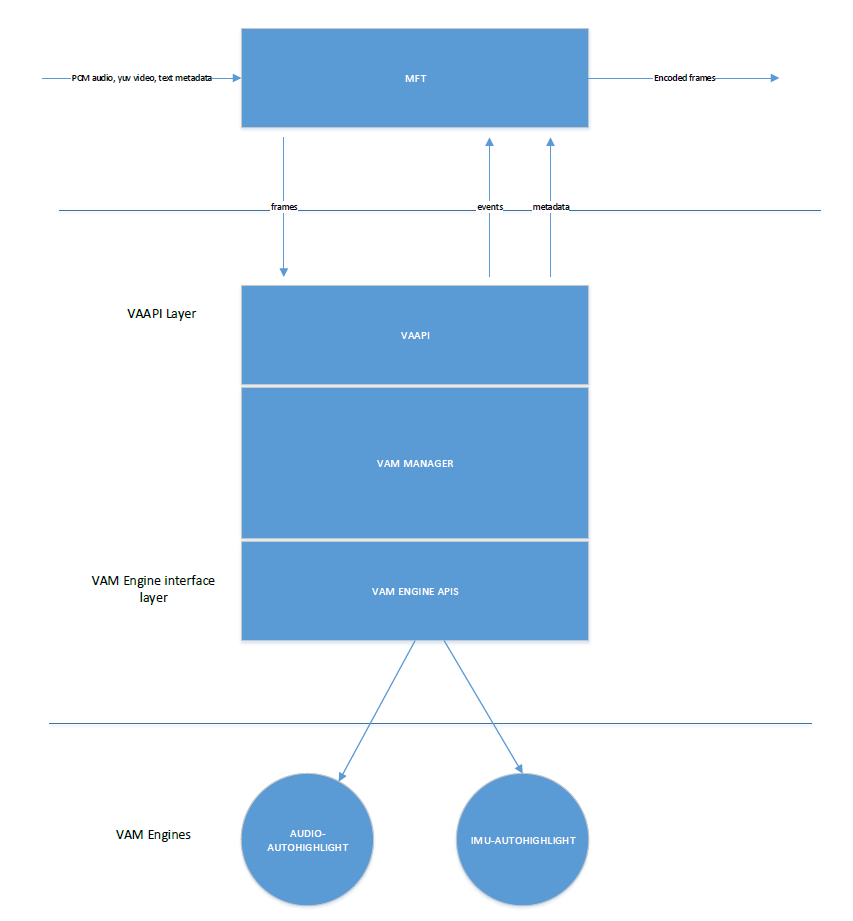
* IMU Rate: 200Hz.  
  Audio Rate: 48Khz, 2 channel, 32bit signed (this is the input) and Algorithm downsamples it to 24khz.
* IMU data is fed to the IMU algorithm every 1 sec and audio pcm data is fed to the Audio algorithm every 22msec. i.e. 1k samples every 22msec.

# Overview

Media formatter(MFT) initializes and performs start/stop of VAM. IMU and PCM audio data are passed to the VAM through MFT.

VAM internally feeds these data to VAM engines(algorithm) integrated and passes on the events(Autohighlight) generated through registered

call-backs back to MFT. MFT takes care of muxing these events as part of MP4 in GPMF format.

**Figure 1:** VAM Framework.

General events that VAM Framework generates are:

1) Audio related events: audio\_auto\_highlight\_event.

2) Sensor data related events: imu\_auto\_highlight\_event.

Types of VAM Engines are:

1) **Audio VAM Engine:** The input for this engine is pcm audio frames, this generates events.

E.g.: Audio-Auto-Highlight Engine.

2) **Sensor Data VAM Engine:** The input for this engine is metadata (GYRO, ACCL...) as text frames, this generates events processing the input frames.

E.g.: IMU-Auto-Highlight Engine.

# OVERVIEW OF VAM

# The application layer intakes encoded video from the camera and configuration settings from the

# VASim\_rule\_config.json file and translates this information into C structure. This data is then

# passed to VAM through VAAPI. From there, it can be passed to VA engines (for example, object

# tracking and facial recognition) using the VAM engine API. Customers can also develop custom

# VA engines using the following VA utility SDKs:

#  FastCV – Currently available

#  SVCE – Currently unavailable, will be added in a later version of VAM

#  VASS-ext – Currently unavailable, will be added in a later version of VAM

# VAM produces encoded metadata frames and event data in the .json format. Using the VASim

# tool, customers can view a rendering of the output data on a PC.

# PASSSING DATA TO VAM USING VAAPI

Following is the highlight data structure passed to VAM,

Highlight Data Structure:

*typedef struct high\_light*

*{*

*uint32\_t time; // highlight time*

*uint32\_t in\_time; //event start time*

*uint32\_t out\_time; //even end time*

*int32\_t longitude; //event location (1E7 format)*

*int32\_t lattitude; //event location (1E7 format)*

*float altitude; //event height (meters)*

*uint32\_t event\_type; //4CCs: AIMU, AACD, FUSE, JUMP, //MANL, etc.*

*float confidence; //0 to 100.0%*

*float score;*

*} high\_light;*

QTI provides the following pseudocode template for passing data to VAM using VAAPI:

*vaapi\_init();*

*vaapi\_start();*

*vaapi\_enroll\_obj();*

*// call back function registrations*

*vaapi\_register\_audio\_event\_cb();*

*vaapi\_register\_imu\_event\_cb();*

*vaapi\_register\_metadata\_cb();*

*vaapi\_register\_frame\_processed\_cb ();*

*//for(video frame input)*

*vaapi\_process();*

*vaapi\_stop();*

*vaapi\_deinit();*

VAM Framework creates one engine instance for each of the algorithms that are integrated and each of them runs on their own thread. The frame data which VAM framework receives will be passed onto the engines and engines process the frame and generates the relevant events and metadata. These events then passed on to the upper layer through registered call-backs.

**vaapi\_init:**

This function begins VAM initialization. The dyn\_lib\_path structure contains the engine plug-in folder.

int32\_t vaapi\_init(const vaapi\_source\_info \*info, const char \*dyn\_lib\_path);

The vaapi\_source\_info data structure is as follows:

*struct vaapi\_source\_info*

*{*

*char data\_folder[VAAPI\_PATH\_LEN]; //Path length is 512.*

*vaapi\_img\_format img\_format; //It contains one of //the following formats:*

*/\* enum vaapi\_img\_format*

*{*

*vaapi\_format\_invalid = 0,*

*vaapi\_format\_yv12,*

*vaapi\_format\_nv12,*

*vaapi\_format\_nv21,*

*vaapi\_format\_YUVJ420P,*

*vaapi\_format\_YUVJ422P,*

*vaapi\_format\_YUVJ444P,*

*vaapi\_format\_GRAY8,*

*vaapi\_format\_RGB24*

*}; \*/*

*uint8\_t frame\_l\_enable;*

*uint32\_t frame\_l\_width[3];*

*uint32\_t frame\_l\_pitch[3];*

*uint32\_t frame\_l\_height[3];*

*uint32\_t frame\_l\_scanline[3];*

*uint8\_t frame\_s\_enable;*

*uint32\_t frame\_s\_width[3];*

*uint32\_t frame\_s\_pitch[3];*

*uint32\_t frame\_s\_height[3];*

*uint32\_t frame\_s\_scanline[3];*

*char is\_test\_mode;*

*};*

**vaapi\_start:**

This function starts the VAM framework, allocates the thread for each of the engine objects, each engine objects run on their own thread.

int32\_t vaapi\_start();

**vaapi\_enroll\_obj:**

This function enrols a single object for VAM. It should be called once per item.

int32\_t vaapi\_enroll\_obj( vaapi\_event\_type type, vaapi\_enrollment\_info \*enroll\_info);

The vaapi\_enrollment\_info data structure is as follows:

*struct vaapi\_enrollment\_info*

*{*

*char id[VAAPI\_UUID\_LEN]; //FR: personID*

*char display\_name[VAAPI\_NAME\_LEN];*

*vaapi\_object\_type type; //Type of the object*

*/\* enum vaapi\_object\_type*

*{*

*vaapi\_object\_type\_unknown = 0, //default type, //should alwyas be replaced*

*vaapi\_object\_type\_people = 1,*

*vaapi\_object\_type\_vehicle = 2,*

*vaapi\_object\_type\_face = 3,*

*}; \*/*

*char img\_id[VAAPI\_UUID\_LEN]; //FR: image id*

*vaapi\_img\_format img\_format;*

*uint8\_t \*img\_data[3];*

*uint32\_t img\_width[3];*

*uint32\_t img\_pitch[3];*

*uint32\_t img\_height[3];*

*};*

**vaapi\_register\_audio\_event\_cb:**

This function registers the VAM event call back function when an audio event is detected.

int32\_t vaapi\_register\_audio\_event\_cb(vaapi\_audio\_event\_cb\_func func, void \*usrData);

**vaapi\_register\_imu\_event\_cb:**

This function registers the VAM event call back function when an imu event is detected.

int32\_t vaapi\_register\_imu\_event\_cb(vaapi\_imu\_event\_cb\_func func, void \*usrData);

**vaapi\_register\_frame\_processed\_cb:**

This function registers the VAM frame process finished callback function. Callback happens once per frame.

int32\_t vaapi\_register\_frame\_processed\_cb(vaapi\_frame\_processed\_cb\_func func, void\* usrData);

**vaapi\_process:**

This function feeds one frame of data to VAM framework, which is passed onto the VAM Engine and it takes care of processing it and generating corresponding events and metadata.

int32\_t vaapi\_process(struct vaapi\_frame\_info \*frame\_info);

The vaapi\_frame\_info data structure is as follows:

*struct vaapi\_frame\_info*

*{*

*uint64\_t time\_stamp;*

*uint8\_t \*frame\_l\_data[3];*

*uint8\_t \*frame\_s\_data[3];*

*uint32\_t obj\_count;*

*vaapi\_object \*objects; //Object data structure is as //follows:*

*/\* struct vaapi\_object*

*{*

*uint32\_t id;*

*char display\_name[VAAPI\_NAME\_LEN];*

*vaapi\_position pos; //Position is as follows:*

*/\* struct vaapi\_position*

*{*

*uint32\_t x;*

*uint32\_t y;*

*uint32\_t width; // 0 <= (X + Width) <= 10000*

*uint32\_t height; // 0 <= (Y + Height) <= 10000*

*}; \*/*

*vaapi\_object\_type type;*

*uint8\_t confidence;*

*int64\_t reserve[VAAPI\_RESERVE\_ITEM];*

*char reserve\_str[VAAPI\_RESERVE\_ITEM]....[VAAPI\_NAME\_LEN];*

*}; \*/*

*};*

**vaapi\_stop:**

This function stops the VAM framework, cleans up the thread context.

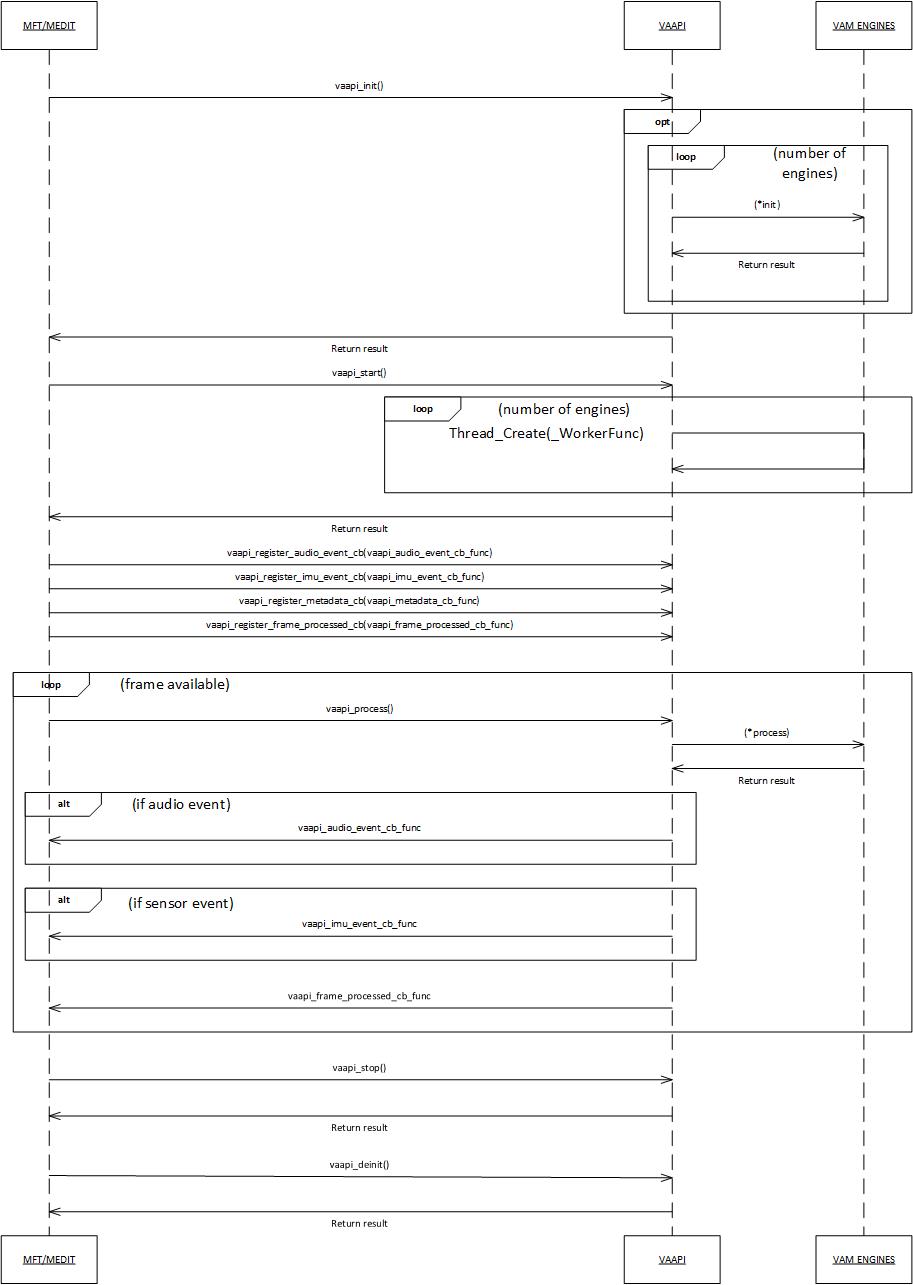
int32\_t vaapi\_stop();

**vaapi\_deinit:**

This function deinitializes the VAM Engines/Algorithms.

int32\_t vaapi\_deinit();

# Sequence Diagram



# Assumptions

Following assumptions are made based on the existing CV framework and proposed VAM framework:

* We use mft\_wrapper.cpp and mft\_wrapper\_video.cpp for binding the VAM calls from MFT module (mft\_main.c, mft\_video.c).
* Fused autohighlight data structure given back to the media-formatter through registered call-backs.
* Autohighlights are saved as part of MP4 track in GPMF format.

# References

|  |  |
| --- | --- |
| **TITLE** | **AUTHOR** |
| *CV\_Document\_v0.3.pdf* | *GoPro.* |
| *video\_analytics\_manager\_developer\_guide.pdf* | *Qualcomm Technologies, Inc.* |

# Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Details** |
| V0.1 | 11-Aug-17 | Shrihari A | Initial Draft |
| V0.2 | 16-Aug-17 | Shrihari A | Internal review comments incorporated |
| V0.3 | 19-Aug-17 | Shrihari A | GoPro review comments incorporated |