December 2019

Method of Capturing a Video and a Set of Selected High-Quality Images During Camera Shutter Long-Press

Wei Hong

Marius Renn

Radford Juang

Lin Chen

Method of Capturing a Video and a Set of Selected High-Quality Images During Camera Shutter Long-Press

Abstract:

This publication describes techniques to provide high-quality images automatically in response to a user long-pressing (making a press and hold gesture) a shutter button on an imaging device. The high-quality images may be selected from after the long-press of the shutter is detected by the imaging device, before the long-press is detected by the imaging device, and/or before the shutter is pressed. A machine-learning technique that ranks video images by their quality and diversity is used to select desirable high-quality images. Both high-quality images and video can be provided to a user in a digital multimedia container format (*e.g.*, MP4) that is compatible with existing media players.

Keywords:

Capture, image, high-resolution, high dynamic range (HDR), video, long-press, shutter button, camera, imaging device, quality score, diversity score, rank, RAW format, YUV format, frame selection, RAW frame selection, YUV frame selection, image deduplication filter, user interface (UI)

Background:

In some implementations, imaging devices (*e.g.*, a digital camera, a smartphone with image-capturing capabilities) include a functionality permitting a user, when in a camera mode, to long-press a shutter button to record a video that will be saved to the device in a video format (*e.g.*, MP4). After the video is recorded, the user can then export a video frame from the video in a still image format (*e.g.*, .jpg). Such photographs, taken by an imaging device, typically have a higher-resolution (*e.g.*, 16MP (4920 x 3264 pixels)) and higher dynamic range in comparison to video recordings (*e.g.*, 2MP (1920 x 1080 pixels) made by the same imaging device. An exported video frame will have the same resolution as the video from which it was exported.

In other implementations, imaging devices record video files that include both a video and a set of selected high-quality images upon detecting a long-press of a shutter button. When the user reviews the captured video file, the user can playback the video as a regular video, the user can export any of the high-quality images, and the user can export any video frame of the video in a still image format.

It is desirable to have a technological solution for imaging devices where the most desirable images (*e.g.*, high-resolution images, high dynamic range (HDR) images) from a recorded video are identified, processed, and provided to the user.

Description:

This publication describes techniques to provide high-quality images automatically in response to a user long-pressing a shutter button on an imaging device. The high-quality images are captured in a high-resolution video stream in addition to a standard video stream that is captured. Figure 1 illustrates an example imaging device that can perform the operations described herein.

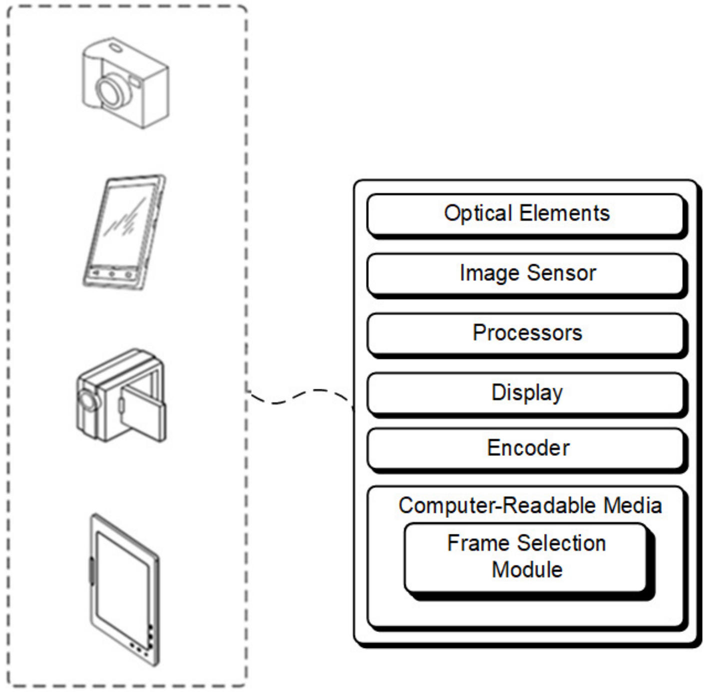

Figure 1

An imaging device can capture multiple consecutive images, such as a video of a scene. In the example of Figure 1, the imaging device is illustrated as being a camera, a smartphone, a video recorder, or a tablet computer. The imaging device includes an image sensor for detecting information used to make an image. The imaging device may include one or more optical elements (*e.g.*, a lens, a mechanical shutter, an electrical shutter, an aperture). The imaging device further includes at least one processor (*e.g.*, an image processor for processing images). The image processor (*e.g.*, an Image Signal Processor (ISP)) is utilized to improve the quality of images generated by the imaging device through image-processing settings, as well as to convert image files into different formats. The imaging device may also include a display for displaying a user interface (UI). The UI is configured to receive input from a user of the imaging device and may include one or more of a touchscreen, a button, a dial, or a keypad.

The imaging device also includes executable instructions of a frame selection module. The frame selection module may be implemented on the computer-readable media (CRM) of the imaging device. The CRM may include any suitable memory or storage device such as random-access memory (RAM), read-only (ROM), or flash memory. The frame selection module represents functionality that uses a machine-learning algorithm to analyze incoming lower-resolution video frames for quality scores (*e.g.*, face quality aesthetic quality, blurriness) and diversity scores (*e.g.*, differences to nearby frames), and ranks the video frames by a linear combination of the scores. The frame selection module selects the highest-ranked video frames, maintains the highest-ranked video frames in RAW format in a RAW frame ring buffer, converts the highest-ranked video frame in the RAW frame ring buffer to a high-quality YUV image file at a fixed time interval, and maintains the high-quality YUV graphics files in a YUV frame ring buffer. The frame selection module filters the high-quality YUV image files for similar images and provides the remaining high-quality YUV image files to an encoder. The encoder outputs a digital multimedia container format (*e.g.*, MP4) file that includes high-quality images. When a user reviews the MP4 file in the UI, the user can playback the video and export high-quality images. A link to view high-quality images may be displayed to a user, and the high-quality images may be available for selection while the user is viewing the corresponding video in the UI.

Upon the initialization of the image capture functionalities of the imaging device (*e.g.*, a long-press, or press and hold of a camera shutter button), the imaging device can capture a video and a set of high-quality images (high-resolution, HDR) as illustrated in Figure 2.

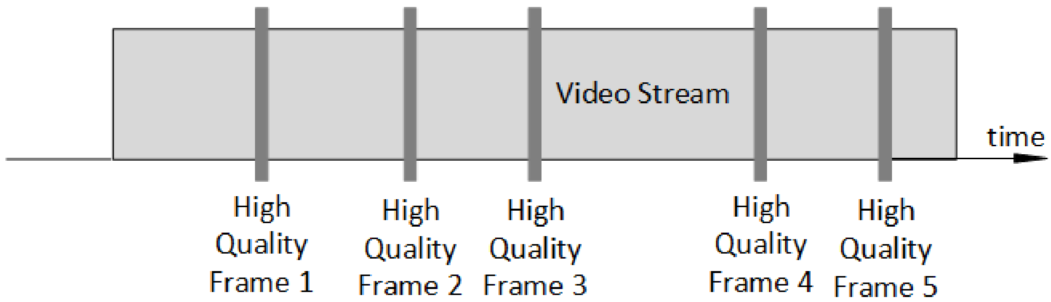

Figure 2

However, due to the temporal nature of a long-press input, it can take 500 milliseconds to detect that detected user input is a long-press. To accommodate the time after a shutter press but prior to long-press detection, the imaging device can record video in response to a shutter press. In an implementation, the imaging device can transmit video frames to the frame selection module for analysis and store a burst of video (*e.g.*, approximately forty-five video frames (about 1.5 seconds of video)) in a video frame ring buffer. The imaging device can capture high-quality images prior to long-press detection, as illustrated in Figure 3.

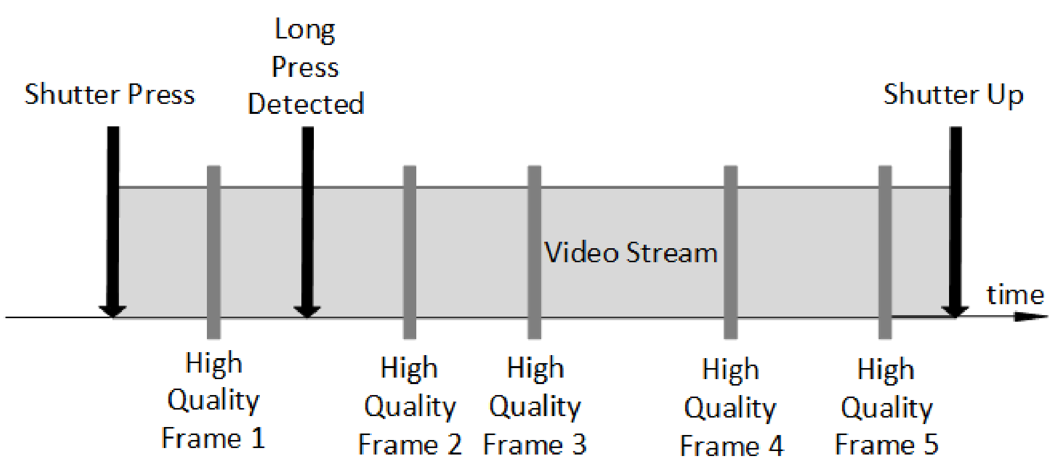

Figure 3

Alternatively, the video recording can begin prior to a shutter press using a trimming algorithm, as illustrated in Figure 4. The trimming algorithm determines the start time of the video, considering criteria. Example criteria include recorded frames before the shutter press that have a camera orientation and brightness similar to a frame when the shutter was pressed. For example, if a user was attempting to take a video of a bird and only pressed the shutter when the bird started to fly, this feature would allow time prior to the shutter press (*e.g.*, a half-second, a second) to be included on the video. Additionally, high-quality frames can be selected from the start time of the video that is prior to the shutter press.

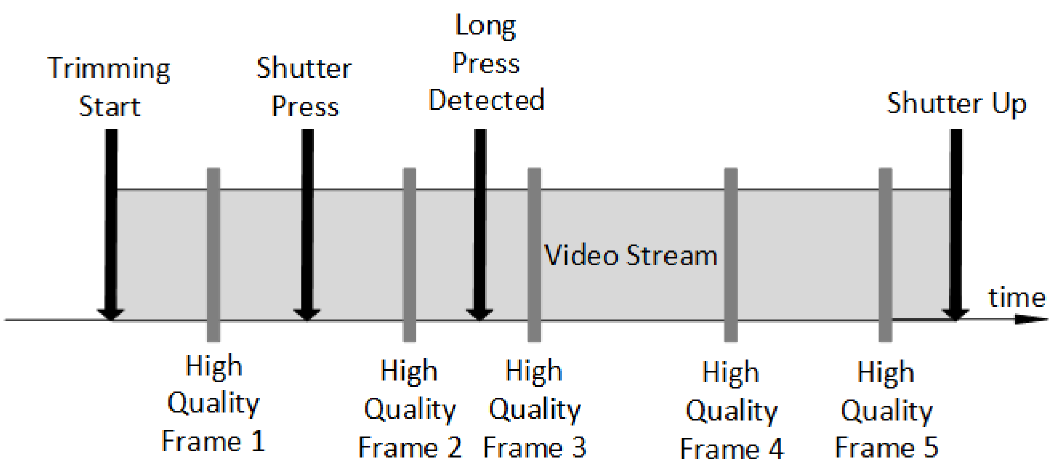

Figure 4

Figure 5, below, illustrates the diagram of capturing a video and a set of selected high-quality images.

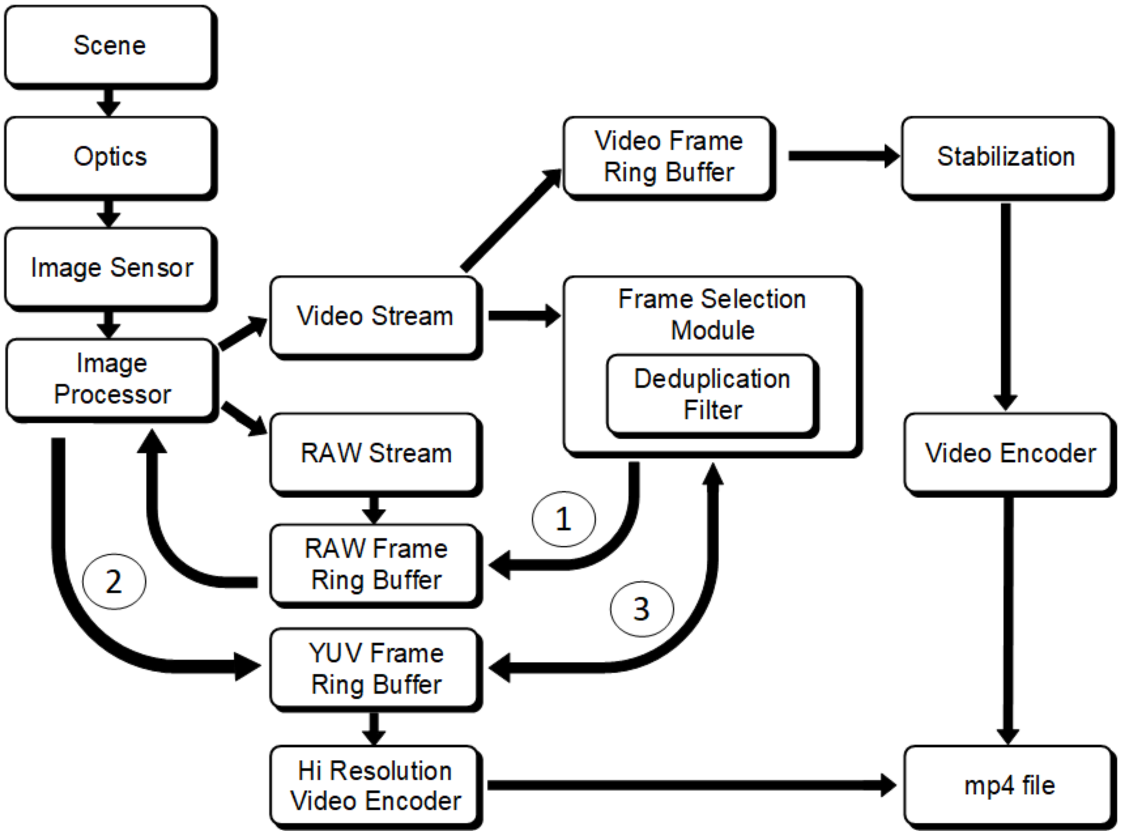

Figure 5

As illustrated in Figure 5, the imaging device will initially capture a scene under optic settings, and an image sensor will send the RAW frames to the image processor. The image processor will output a lower-resolution video stream and a high-resolution RAW video stream. The lower-resolution video stream and the high-resolution RAW video stream can be at the same rate (*e.g.*, thirty frames per second) and paired by a timestamp. The frame selection module can use machine-learning to analyze a lower-resolution video frame in real-time by evaluating the frame for quality (*e.g.*, face quality, aesthetic quality, blurriness) and for diversity based on the differences to nearby frames (*e.g.*, time difference, facial expression, camera orientation). All of the incoming frames are ranked by the frame selection module by a linear combination of each respective frame's quality and diversity scores. Using the timestamp, the frame selection module locates the high-resolution RAW frames that correspond to the highest-ranked video frames and stores them in a RAW frame ring buffer, as illustrated by the number 1. The RAW frame ring buffer may be small (*e.g.*, have a capacity of five frames) due to memory limitations on the imaging device.

The highest-ranked image in the RAW frame ring buffer is converted by the image processor from RAW to a higher quality (*e.g.*, higher-resolution, HDR) YUV image format and then is stored in a YUV frame ring buffer, as illustrated by number 2. Processing a high-quality image is power-consuming, so the RAW images may be processed slowly (*e.g.*, at a fixed time interval of one per second). The YUV frame ring buffer may be small (*e.g.*, have a capacity of five frames). If the total number of YUV images is more than the capacity of the YUV frame ring buffer, the frame selection module can remove the lowest-ranked YUV image from the YUV frame ring buffer, as illustrated by number 3.

Once a user releases the shutter button, the YUV images are passed through a deduplication filter on the frame selection module, also illustrated by number 3, which removes similar images based on certain differences (*e.g.*, facial expression difference, camera orientation difference). The remaining YUV images after the deduplication filter are encoded by a high-resolution video encoder and output as an MP4 file. The cascaded three-step frame selection (RAW frame selection, YUV frame selection, YUV frame deduplication filter) allows the system to select the best images out of a large number of frames with low power and low memory consumption.

The video stream and the high-resolution images can be saved in the same MP4 file. The video frames from the video ring buffer are stabilized, encoded by a video encoder, and saved to the MP4 file. Only the frames which need to be saved will be stabilized and encoded to save power. The video track and the audio tracks are stored in the MP4 file. However, the high-quality images are embedded in a metadata track of the MP4 file instead of a video track due to limitations on most existing media players to support multiple video tracks. The MP4 file also contains per-video metadata and per-frame metadata. The per-video metadata track contains information for the entire video (*e.g.*, timestamps and quality scores of all of the high-quality images). The per-frame metadata track contains information for each video frame (*e.g.*, the quality score of each video frame).

In conclusion, utilizing the techniques and processes described herein, a user can acquire high-quality images automatically when the user is taking a video in real-time with low memory and power requirements for an imaging device.

References:

[1] Patent Publication: US 20030068100 A1. Automatic selection of a visual image or images from a collection of visual images, based on an evaluation of the quality of the visual images. Priority Date: July 17, 2001.

[2] Patent Publication: US 20110261217 A1. Image processing architecture with pre-scaler. Priority Date: April 21, 2010.

[3] Patent Publication: US 20050099494 A1. Digital camera with panoramic image capture. Priority Date: November 10, 2003.