July 2023

On-device Frame Interpolation for High Frame Rate Video Conferencing

Lijun Li

On-device Frame Interpolation for High Frame Rate Video Conferencing

ABSTRACT

This disclosure describes techniques to use frame interpolation to provide high quality, high frame rate video streams displayed on a mobile device in applications such as a video conference with low power consumption. A video stream having a low frame rate is received at a mobile device. On-device frame interpolation is performed using techniques such as motion estimation and motion compensation, voxel flow, or machine learning to determine a new frame between consecutive frames of the video stream. The video stream including the interpolated frames is displayed on the mobile device at a higher frame rate, providing smooth motion.

KEYWORDS

- frame interpolation
- video stream
- optical flow
- voxel flow
- super resolution
- motion estimation
- motion compensation
- high frame rate
- on-device machine learning
- video conference

BACKGROUND

Video conferences are a popular method of communication using mobile devices. However, multi-party video conferencing provides challenges for mobile devices related to power consumption and availability, as well as thermal limits of the devices. For example, multi-party video conferencing over a WiFi or a cellular network has high demand for communication bandwidth. A device participating in a video conference encodes and uploads a video stream while receiving and decoding multiple video streams from the cloud. In addition, features such as blurring video background, relighting a video, modifying displayed faces, providing realistic audio, etc. available in many video conferencing applications have high computing requirements for CPU (central processing unit), GPU (graphics processing unit), and machine learning units.

The power consumption of a device during a video conference may depend on the bit rate of displayed video and voice traffic on WiFi and cellular networks. Video conferencing features tax the battery of mobile devices and cause significant power drain. Mobile devices have only limited battery life, which is quickly consumed with the high amount of processing required by the mobile device to process and display multiple video streams. In addition, high temperature operation of the mobile device resulting from video conference processing may cause a poor experience for the user holding the mobile device during the video conference.

DESCRIPTION

This disclosure describes techniques for displaying high frame rate video on a mobile device without correspondingly high power consumption. These techniques can enable a multi-party video conference to be implemented on a mobile device with reduced power usage and heat emission. The described techniques reduce the frames in transmitted and received video streams and interpolate frames in video to maintain visual quality. When software decoding is

used for multiple video streams, reducing the bandwidth of each stream can reduce CPU power consumption significantly. Moreover, video stream bandwidth reduction can also reduce the amount of network traffic. Thus, reducing video bandwidth in video conferencing can enable substantial power usage reduction on a mobile device.

The described techniques can be implemented on any suitable portable device, e.g., a smartphone, smartwatch, headset, etc. The user is provided with options to enable or disable described techniques. The user can permit specific applications, uses, data, or types of data to be processed with described techniques and can deny such processing for other uses or data.

Interpolating video frames in a video stream for display on a mobile device

During a video conference, a mobile device may receive multiple video streams, each stream being provided by a device of a different participant in the video conference. Each video stream may be simultaneously displayed on a screen of the mobile device, e.g., such that all (or a subset) of the participants in the video conference are displayed. At the same time, the mobile device captures image frames of the user of the mobile device, converts the frames into a video stream, and sends the video stream to a server or other participant devices such that the video stream can be displayed by the other devices in the video conference.

According to described techniques, frame interpolation is performed on a mobile device to increase a low frame rate of a received video stream to a higher frame rate when the video is displayed by the mobile device. The frame interpolation inserts a new frame between one or more frame pairs (that include consecutive original frames) in the video stream. This allows a lower base frame rate to be used in transmitting the video stream, while providing smooth motion in the displayed video stream at the receiving device. For example, a base frame rate of

12 fps in a video stream can be increased to 24 fps when displayed on a screen of the mobile device.

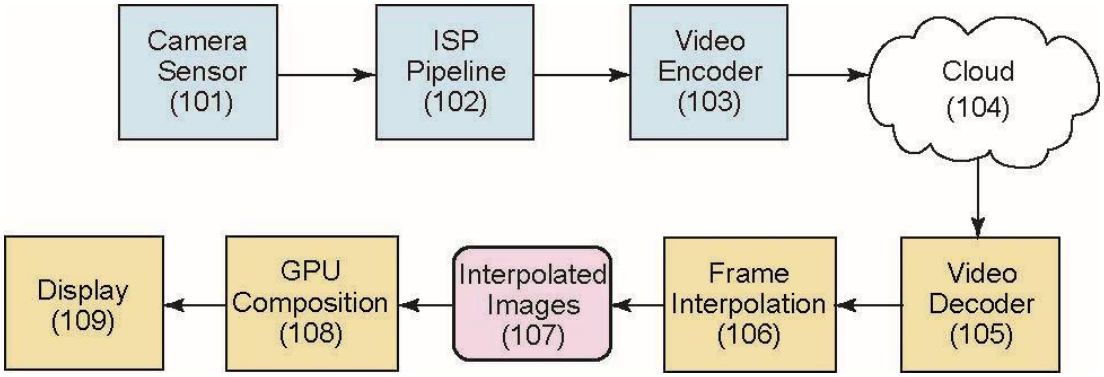

Fig. 1: Video conference with frame interpolation

Fig. 1 shows an example system for sending and receiving/displaying video streams with frame interpolation. The system includes a sending device and a receiving device (or multiple sending/receiving devices). Each of these devices can be a mobile device, e.g., mobile phone, tablet, smartwatch, headset, etc. A camera sensor (101) of the sending device captures frames for a video. The video frames are captured at a particular image resolution and a low frame rate, e.g., 720P at 12 fps (frames per second) in this example.

The sensor provides captured video frames to an image signal processor (ISP) pipeline (102) of the sending device, which performs processing for the video frames, e.g., noise reduction, gamma and color correction, color interpolation, etc. The ISP pipeline provides the video frames to a video encoder (103) of the sending device that encodes the video frames into a video stream. The video encoder sends the video stream to the cloud (104), where a server sends the video stream to other devices, e.g., devices participating in the video conference.

A receiving mobile device, connected to the video conference, receives the encoded video stream from a server in the cloud. A video decoder (105) of the mobile device decodes the

video stream and sends it to a frame interpolator (106). The frame interpolator interpolates new frames between original frames of the video stream. Some example techniques that can be used to interpolate the new video frames are described below.

The frame interpolator sends interpolated image frames (107) to a graphical processing unit (GPU) composition block (108). The composition block interleaves the interpolated frames with the original frames of the video stream to produce an interpolated video stream. The interpolated video stream is sent to a display device (109) for display, e.g., a display screen of the mobile device. In the display of the video stream, the original and interpolated video frames are displayed alternately to provide a higher frame rate than the frame rate of the original video stream received from the cloud.

Multiple video streams can be received by the device, each video stream simultaneously being processed from the cloud similarly as the received video stream described above. For example, each other participant in the video conference may send a video stream such that the multiple video streams are received by the mobile device and the multiple participants are simultaneously displayed.

The frame interpolator can use any suitable technique to interpolate the new video frames. An interpolated frame is determined between first and second original frames in the received video stream. For example, a first original frame in the video stream at time 0 is followed by a second original frame at a time 1, and an interpolated frame may be inserted at a time t, halfway between times 0 and 1.

In some devices, motion estimation and motion compensation (MEMC) techniques can be used for the interpolation. Motion estimation is used to determine block-wise or pixel-wise motion vectors between the two original frames. For example, an optical flow technique can be

used to estimate movement of pixels between the two frames, e.g., in forward flow from the first original frame and/or backward flow from the second original frame and determine the new intermediate frame from these estimates. Block-based techniques can calculate motion vectors for blocks of pixels moving between the first and second original frames.

Other frame interpolation techniques can also be used. For example, a machine learning (ML) model can be used, e.g., a pre-trained ML model specific to video content. Real-time inference of interpolated frames can be performed using the ML model with model optimization provided on the mobile device. For example, a lightweight ML model can be provided that executes efficiently on devices with limited computing and memory resources. Pruning and weight quantization can be used to provide an ML model that is suited to mobile devices.

In some cases, deep voxel flow can be used, in which a deep network is trained to synthesize video frames by flowing pixel values from existing pixels. Adjacent frames of the video stream can be input to a convolutional encoder-decoder, which determines voxel flow between the two frames and produces a synthesized and interpolated frame. In another example, a neural network can directly estimate intermediate flows from the original frames, and a fusion process can combine the frames to reconstruct the interpolated frame.

Frame interpolation can be performed in an adaptive manner based on power consumption. For example, frame interpolation may be disabled for a two person video conference to provide high quality video by allowing the camera sensor to use a high frame rate, e.g., 24 frames per second. The described techniques can be implemented in any video conferencing application that executes on a mobile device such as a smartphone.

Features can provide real time frame interpolation on a mobile device with low power consumption, allowing a lower frame rate of the video stream that is received and processed by

the mobile device. In some examples, in order to have the best picture quality, the camera sensor can have a high image resolution and a low frame rate (e.g., a resolution of 720P and a frame rate of 12 fps). The reduction of frame rate allows the bandwidth requirements to be reduced in other components of the system such as the ISP pipeline and video encoder of a sending device, in communication traffic through the cloud to the receiving device, and in the video decoder of the receiving device. After the video stream is decoded at the receiving device, a new frame is interpolated between every two frames to provide smoother motion in the display of the video stream at the receiving device.

The frame interpolation can be adaptive to the power consumption of the mobile device. For example, the frame interpolation may be disabled on the mobile device for two person video conferences in which only one video stream is received, where power consumption is lower than in multi-party video conferences having multiple received video streams. In such two-person video conferences, the camera sensor can sample frames at a higher rate than video conferences with more persons (e.g., 24 fps instead of 12 fps with more persons) to provide higher video quality without interpolation.

Some devices can also upscale low resolution frames for display at the receiving device, as well as interpolating frames. This allows frames of reduced data to be transmitted, encoded, and decoded, reducing power consumption, network transmission, and processing time. While the foregoing description refers to the received device performing frame interpolation, frame upscaling/ interpolation can also be performed by a sender device, depending on the configuration. The described techniques can improve battery life through power reduction for applications such as gaming, video conferencing, video recording/playback, real-time map display, etc.

CONCLUSION

This disclosure describes techniques to use frame interpolation to provide high quality, high frame rate video streams displayed on a mobile device in applications such as a video conference with low power consumption. A video stream having a low frame rate is received at a mobile device. On-device frame interpolation is performed using techniques such as motion estimation and motion compensation, voxel flow, or machine learning to determine a new frame between consecutive frames of the video stream. The video stream including the interpolated frames is displayed on the mobile device at a higher frame rate, providing smooth motion.

REFERENCES

1. Jiang, Huaizu, et al. "Super slomo: High quality estimation of multiple intermediate frames for video interpolation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
2. Huang, Zhewei, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. "Real-time intermediate flow estimation for video frame interpolation." In *European Conference on Computer Vision*, pp. 624-642. Cham: Springer Nature Switzerland, 2022.
3. Romano, Yaniv, John Isidoro, and Peyman Milanfar. "RAISR: Rapid and accurate image super resolution." *IEEE Transactions on Computational Imaging* 3, no. 1 (2016): 110-125.
4. “Time-stretching and time-remapping” available online at https://helpx.adobe.com/in/after-effects/using/time-stretching-time-remapping.html accessed July 2, 2023.

5. “Transforming Standard Video Into Slow Motion with AI | NVIDIA Technical Blog” available online at https://developer.nvidia.com/blog/transforming-standard-video-into-slow-motion-with-ai/ accessed July 2, 2023.
6. “Optimal Video Viewing Experience with AMD Radeon™ Software” available online at https://www.amd.com/en/support/kb/faq/dh2-006 accessed July 2, 2023.
7. “Trimension - Wikipedia” available online at https://en.wikipedia.org/wiki/Trimension accessed July 2, 2023.
8. “Asynchronous Timewarp Examined” available online at https://developer.oculus.com/blog/asynchronous-timewarp-examined/ accessed July 2, 2023.
9. “Asynchronous SpaceWarp | Oculus Developers” available online at https://developer.oculus.com/documentation/native/pc/asynchronous-spacewarp/ accessed July 2, 2023.