International Journal of Engineering Trends and Technology (IJETT) - Volume 68 Issue 4 - April 2020

Generating High-Resolution Video Using Single Camera in Android

Sunil Kumar[#1] Digvijay Pandey[#2] Nitin Setia[#3]

Oppo Research & Development
Hyderabad, 50081 India

Abstract

This paper talks about generating a multi frame real time video from smart phone using single camera. There are many existing solutions that talk about the same phenomenon using multiple cameras. At the time of video recording, there are multiple frames we want to switch and highlight but using phone's single camera video recording is not possible. This is generally generated using multiple camera's and combining their outputs. Existing process is also manual and high chances of miss the complete scene that user want to shoot. We can do the same using single camera image segmentation. Preview resolution is set at background with full FOV (Field of View) and divided in to the multi FOV's, based on the scene, objects will be decided. Weights are given based on algorithmic relevance. In the same stream, the bounding box of the image having maximum weight will be shown to fit in the display without loss of any pixel information.

1 Introduction

In the era of information technology, the use of an android camera is much common. Especially young minds are very fond of an android phone camera. There are many uses of android and good camera quality is one of them. This paper mainly focusing on video streams. In video mode, the main concern is to show streams without compromising quality as well as information. Getting a high-resolution video is user demand nowadays. A real-time system like surveillance and sports, it is often desirable to zoom and pan into the generated video. Zooming technique commonly used in the past, but it degrades the resolution. It also needs manpower and a number of different use-cases. Sometimes even multiple cameras are not beneficial. Up-scaling captured images may cause pixel aberration. There is always a trade-off between zoom and aberration. This paper will provide a different technique to overcome all these problems.

2 Related Work

Vamsidhar Reddy Gaddam *et.al* [1] presented a model that works on the principle of cylindrical panorama as an intermediate representation. In this paper, they have used additional hardware. This paper has optimized the Zooming technique. The paper relies on interactive end-to-end real-time system zoom and the same zoom/crop pans into the high-resolution panoramic videos. Limited to the panorama video shooting this paper basically talks about virtual zoom using GPUs.

Yasuo Ariki *et.al* [2] presented a paper in which they propose an automatic production system of commentary soccer video by digital shooting techniques based on the event/image recognition.

W. Andrew *et.al* [3] worked on the Automatic extraction of secondary video streams. The automatic generation of one or more secondary video streams based on an input primary video stream may use analytic performed on the video to provide information on targets, events, and/or areas of interest to permit the one or more secondary video streams to concentrate on one or more targets, events, and/or areas of interest.

Amtrup *et.al* [4] has filed a patent name 'Systems and methods for detecting and classifying objects in video captured using mobile devices'. A method includes invoking an image capture interface via a mobile device; and analyzing video data captured via the capture interface. The analysis includes determining whether an object exhibiting one or more defining characteristics is depicted within the viewfinder; and if so, whether that object satisfies one or more predetermined quality control criteria. The method further includes displaying an indication of success or failure to satisfy the predetermined control criteria on the mobile device display. Where the object depicted within the viewfinder satisfies the one or more predetermined quality control criteria, the method also includes: displaying an indication that the object depicted in the viewfinder exhibits the one or more defining characteristics; automatically capturing an image of the object; and/or automatically storing to memory one or more of the frames in which the object is depicted in the viewfinder. Systems and computer program products are also disclosed.

3 Methodology

Overview of android camera is illustrated in Figure 1:

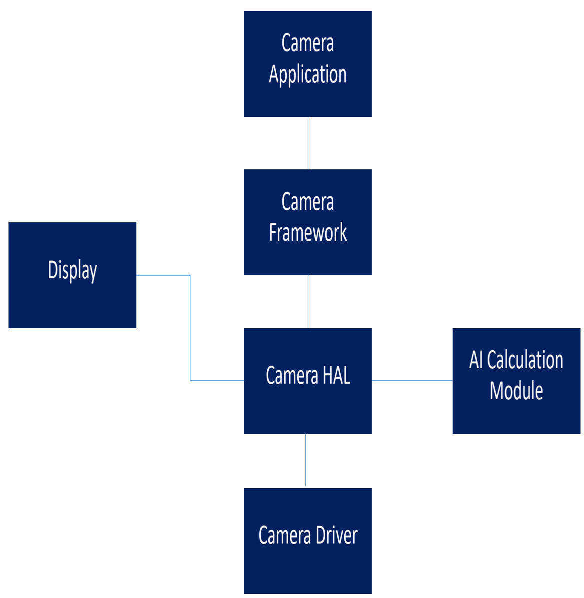

Figure 1: Android Camera Overview

First, Camera is opened in AI Recording mode. Surface Texture is created for displaying preview. The resolution of the above created surface texture is selected by application among the given supported resolutions by Camera HAL. Preview resolution is set at background and divided in grids. The whole surface texture is divided into equal number of grids having same resolution as surface texture. Image analysis is done on preview to find objects and region of interest. Area if interest is evaluating intelligently by AI module based on the scenes. Some of the examples are shown in experimental results. If there is user intervention for region of interest then that is taken into consideration on priority, The display is changed in cropped frame from overall (Field of View) FOV which contains region of interest and the same buffer is sent for video encoding. Each Small grid has same aspect ratio as full preview. We are selecting the main preview resolution with same aspect ratio grid and sending those blocks as one frame for the video encoding. The division of preview into grids, image analysis grid and the selection of final grids (after object detection using AI) are our own.

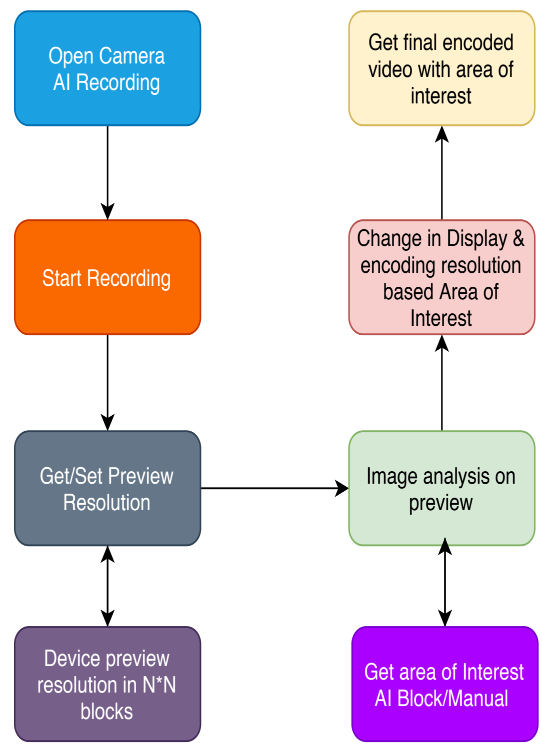

Figure 2: Work Flow

Algorithm 1 High Resolution Video Generation Using YOLOv3

1: **procedure** MYVIDEO
2: Turn android camera on
3: Select usecase
4: Divide the display in grids of same resolution
5: Detect Scene
6: AI based object identification for Scene
7: Select bounding box(es) of interested object
8: Only show region of interest (ROI) on full screen
9: **end procedure**

4 Experiment

Real-Time Scene detection is important for proposed solution. For that Deep Learning model is used. The details of model is illustrated in Figure 1:

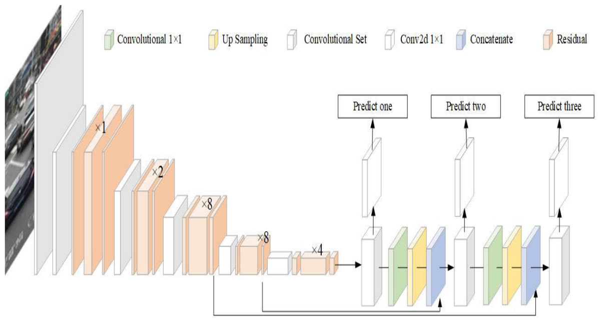

Figure 3: YOLOv3 Architecture

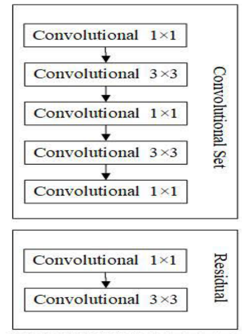

Figure 4: YOLO Output Layers

YOLOv3 [5] testing process is as follows:
Step 1: Take input from user and make the size of input as standard size.
Step 2: Split the input image into the chunks of 13×13, 26×26, 52×52 grids. Wherever central point falls, that grid will be responsible for object detection.
Step 3: Bounding-Box of each grids is decided by Unsupervised Machine Learning algorithms.
Step 4: Input the image into the networks of grid for feature extraction. A small scale feature map is produced by the model. The size of feature map is 13×13.
Step 5: The smallest feature map which is having size of 13×13 is passed to the convolution set. After this the twice sampling is applied on the same. The result is then passed onto the bigger feature map having size 26×26. The output of this feature map will predict the result.
Step 6: Repeat step 5 for for feature map of size 26×26. Output would be passed to bigger feature map for result prediction.
Step 7: Output of all three feature maps combined together for final prediction. Apply threshold for non-maximal suppression and also remove predictions having very low score or having scores which is less than a fixed threshold. Afterward, using a probability score as a threshold to filter out most anchors with low scores.
Step 8: Run YOLO v3 on GPUs to get high performance result. It will consume power but in real time scenario we need accurate result for each frame. This will give bounding box of object.

5 Result

Here are some experimental results. The bounding boxes of region of interest is also shown in diagrams. While recording videos of events like birthday party, wedding etc., user wants main specific object should be focused and be in limelight. Usually user needs to zoom manually and operate it by himself. Due to this user may not be able to enjoy the real moment. This paper can solve this problem as user can live the moments at the same time object focused video is being recorded.

5.1 *Celebration*

In this very first experiment, scene is detected as birthday celebration and object is identified as birthday girl. The grids in which object(girl) falls identified and passed on to encoder for recorder instead of full FOV.

Figure 5: Region of interest is birthday girl

5.2 *Recording Conferences*

In case of conferences, tagged speaker/object can be focused and previewed while recording. It will enhance the user experience.

Figure 6: Region of interest is speaker

5.3 *Live Games*

This solution can be very helpful in case of recording games. If scene is detected as a Game then it is narrowed down to a particular game on the basis of various objects parameters present in the scene. In the experiment scene is detected as Cricket Game and from current FOV object is detected as fielder.

Figure 7: Region of interest is the player catching the ball

5.4 *Wild-Life video recording*

This solution can be used in wildlife video recording when wild animal is far away and be in action like hunting, chasing the prey. When tagged object is in action, it is automatically focused and previewed.

Figure 8: Region of Interest is lion

6 Conclusion

In this paper a new model is presented that can shoot video using a single camera. The model also focuses on user perspective. This paper has discussed a new algorithm for getting a better video resolution video from an android camera. Existing solutions are using multi-camera or zooming techniques to implement the same functionality. This paper is neither using multi-camera nor the zooming mechanism. This paper has reduced the cost by removing additional camera(s), and not using a zooming tool enhances the video quality. From a user experience point of view, before this paper, there exist numerous mechanisms (solution), but none of them has given the user flexibility to enjoy the scene at the time of video shooting.

7 Acknowledgement

Thanks to Oppo India Camera Team.

References

[1] V. R. Gaddam, R. Langseth, H. K. Stensland, P. Gurdjos, V. Charvillat, C. Griwodz, D. Johansen, and P. Halvorsen, "Be your own cameraman: Real-time support for zooming and panning into stored and live panoramic video," in *Proceedings of the 5th ACM Multimedia Systems Conference*, pp. 168–171, 2014.

[2] Y. Ariki, S. Kubota, and M. Kumano, "Automatic production system of soccer sports video by digital camera work based on situation recognition," in *Eighth IEEE International Symposium on Multimedia (ISM'06)*, pp. 851–860, IEEE, 2006.

[3] W. A. Scanlon, A. J. Chosak, J. I. Clark, R. A. Cutting, A. J. Lipton, and G. W. Myers, "Automatic extraction of secondary video streams," Sept. 30 2014. US Patent 8,848,053.

[4] J. W. Amtrup, J. Ma, and A. Macciola, "Systems and methods for detecting and classifying objects in video captured using mobile devices," Nov. 11 2014. US Patent 8,885,229.

[5] Q. Wang, S. Bi, M. Sun, Y. Wang, D. Wang, and S. Yang, "Yolov3-tiny architecture.," Jun 2019.