ClickTheBest: An AI Based Enhanced Burst Capture for Android Camera

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

Oppo Research & Development
Hyderabad, 50081 India

Abstract

This paper talks about clicking multiple pics of dynamic object using burst mode of android camera. Idea is to generate burst images intelligently and save only those which are currently required. Burst photo is user-initiated photo capture where user long press the capture icon and capture continuous images (mostly 20-30 continuous images). Usage of burst capture is where a moving object or Action is present, and user is not sure which frame user wants or which frame would be the best suited for a specific environment. To overcome this, end user captures continuous photographs (which are basically called burst mode capturing) and then based on some heuristics. select the best picture out of all captured pictures. Selection of Heuristics and its criteria is decided by Deep learning model. Apart from automatic selection this paper also includes manual selection of objects. End user has been given flexibility to choose their own scene, object, moment which they want to shoot.

1 Introduction

In the era of information technology, the use of an android camera is much common. Idea is to generate burst images intelligently and save those which are currently required. Burst photo is a user-initiated photo capture where user long press the capture icon and captures continuous images (mostly 20-30 continuous images). Usage of burst capture is where a moving object or Action is present and user is not certain which frame user wants, so the user captures continuous photographs and select the best one. There are some drawbacks of existing Burst mode:

1. User is capturing those images also, which he does not require, and these images occupy space in his phone memory. One or two out of twenty images are useful and the rest are not. Over the period if there are many burst captures, it occupies large phone memory
2. As it is fast capture, and after first focus before start capturing, there is no Refocus. It may happen that the image user wants is out of focus
3. An end user needs to press capture button continuously

The idea behind this research paper is to create AI based solution to improvise the existing burst capture that solves above three problems. Object is tagged with the help of AI module before triggering burst mode even if it is not in FOV (Field of View). If tagged object starts moving or if it enters the frame (if object was not in frame initially) of camera, burst mode will automatically get triggered and real moments are captured. Another mode *i.e.* Manual Mode is also given to enhance the level of customization. In manual mode user has given flexibility to select object and it will supersede the environment-based object tagged by AI module.

2 Related Work

Vinayak J *et.al* [1] filled a patent name: A method of providing an optimum image of a scene in a digital camera. In this patent they've talked about plurality of camera frames and setting. Some other methods (features) like method of comprising setting a plurality[2] of camera settings, capturing a plurality of images of said scene in a burst mode, selecting automatically a first image in said plurality of images, forming a plurality of frames, extracting a corresponding plurality of image parameters and calculating a corresponding metric value[3] for each plurality of frames also got mention substantially.

Nobori *et.al* [4] filled a patent. In the same patent they talked about an image capturing apparatus according to one aspect of the present disclosure is an image capturing apparatus that captures a plurality of images, calculates a three-dimensional position[5] from the plurality of images that have been obtained, and outputs the plurality of images and information about the three-dimensional

position, the image capturing apparatus including an image capturing unit, a camera parameter storage unit, a position calculation unit, a position selection unit, and an image complementing unit. The image capturing unit is configured to generate and output the plurality of images by performing image capturing using at least three cameras, the cameras being arranged such that optical axes[6] thereof are substantially parallel to one another and the cameras are close to one another.

According to the S Thrun *et.al* [7] invention, enables a user to define tags that label objects in panoramic images and video. Objects may be stationary in the 3-D world, or they may be moving. It also gives a user flexibility to define the time-space coordinates of tagged objects in a user-friendly way, by manually labelling one or more panoramic images. They also enable a user to attach further information to a tag, such as textual information, images, or specific actions. Embodiment according to the present invention make it possible to visualize tags when viewing panoramic images and video. They enable a user to execute specific actions through selecting a tag. And finally, embodiment according to mentioned invention make it possible to textually index a panoramic image database[8], by linking textual annotations to specific objects in the video stream.

Kevin Flory *et.al*[9] proposed a solution. According to mentioned invention camera has a continuous full-resolution burst mode wherein a sequence of full-resolution images are captured, images are processed by a pipeline of dedicated hardware image processing engines, is zoomed by a zoom engine, is compressed by a compression engine, and is stored into non-volatile storage as a sequence of discrete files.

The prior art(s) is only capturing the images in burst mode with full resolution as a sequence of images. Also, optimal image is found out of captured burst images. The prior art approach is totally manual and does not guarantee to capture accurate moments of a moving object. User needs to trigger the burst mode manually in aforementioned prior art approaches.

3 Methodology

Overview of android camera is illustrated in Figure 1:

Idea behind burst shot is that there is one moving subject which needs to be captured properly. In our AI burst shot, user needs to tag object/person that needs to be captured and focus on the frame in which object needs to be captured; once the object is in that frame and started moving, it will first refocus, then burst capture will initiate, it will select the best capture frame and auto discard others.

Whole process is divided into four steps:

- **Initiate:** Tag the object that need to capture, making at least one known object from other unknown objects
- **Start:** Trigger point to start Burst Capture
  - User
  - Automatic
- **Process:** Capture burst images
- **End:** End of Burst capture based on below parameters
  - Maximum number of images captured in one burst (currently 20)
  - Object out of scene
  - Object stop moving

Object can be tagged by two different ways:-

- If object is in frame, object can be tagged by long press
- If object is not in frame and going to arrive, it can be tagged by entering text, for example, type dog and dog will be tagged, once dog arrives in frame burst mode starts capturing.

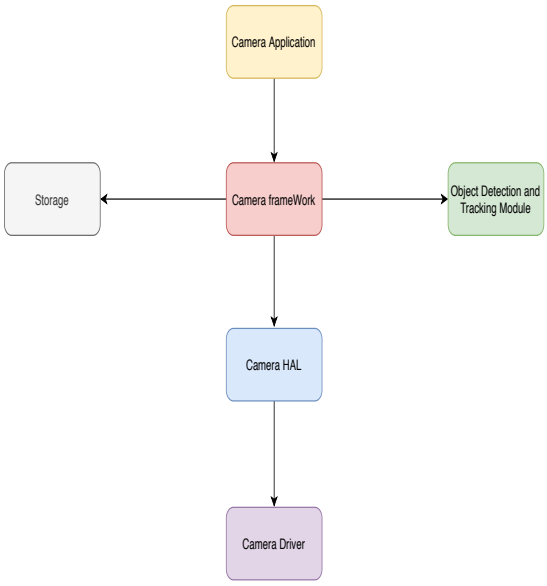

Figure 1: Android Camera Overview

A pre-existing module is used for object detection and tagging. Burst mode will be triggered either by user as

an existing method, long press capture button or in automatic mode. In automatic mode, user will click capture button once after tagging the object and the condition for AI module capturing burst mode given in algorithms. Working of AI module is given below:-

Algorithm 1 High Resolution Video Generation Using YOLOv3

procedure MYVIDEO
    Turn android camera on
    Select usecase
    Divide the display in grids of same resolution
    Detect Scene
    AI based object identification for Scene
    Select bounding box(es) of interested object
    Only show region of interest (ROI) on full screen
end procedure

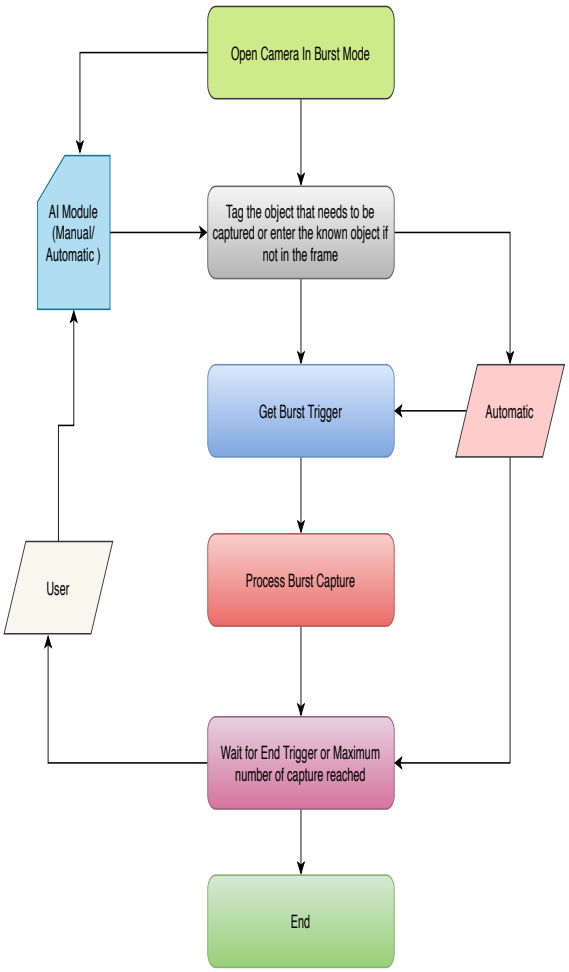

Figure 2: Work Flow

Algorithm 2 Preconditions for Capturing Burst

Result: Set flag for Capturing Burst event
while *Wait till below conditions are not true* **do**
    1. Object is in frame and starts moving. Object is focused at his position and better result is seen.
    2. Object is not in frame and once it arrives in frame, burst capture starts, before first capture, lens moved and object refocused and burst started
    3. Object/Environment color is changing
    4. Object shape is changing
end
Trigger Burst Shot
while *Wait till below conditions are not true* **do**
    1. If object moved out of frame
    2. Object movement is stopped
    3. Object/Environment color is stable and not changing
    4. Object is stopped changing shape
    5. Maximum number of frames are captured
    6. User released the capture button (user triggered burst mode)
end
Stop Burst shot

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:

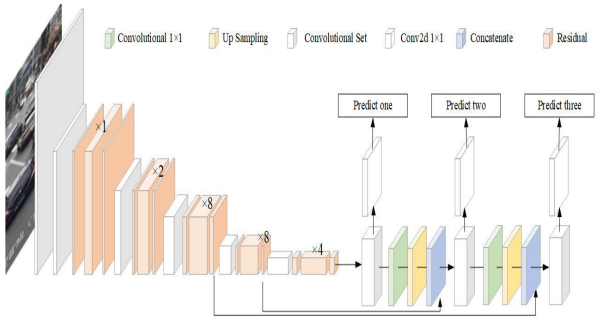

Figure 3: YOLOv3 Architecture

YOLOv3 [10] 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

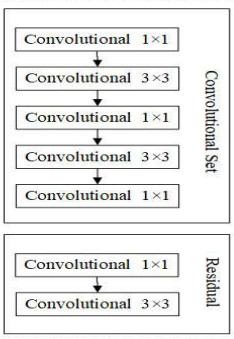

Figure 4: YOLO Output Layers

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 tagged object and the actual event is also shown in below diagrams. For android camera/computer vision, YOLOv3 yields the desired result. Selection of AI module is still an open problem. We can also change the result based on how much time and android camera is taking to capture an event. This paper also solves the problem of redundant images as well as automating the process using AI module. Sometimes object detection may not return the same result as used wants. It totally depends on AI module that we used here.

5.1 *A Baseball scene*

In the game of baseball, a player is hitting the ball with his bat and the time ball reaches to bat is an event. End user is interested to capture those shots preciously.

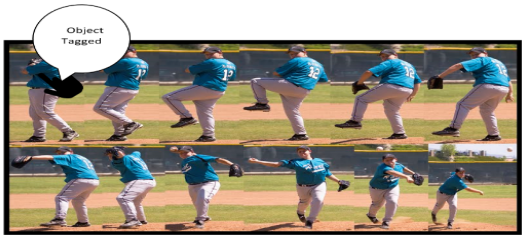

Figure 5: A Baseball Match

5.2 *Live Games*

In this image user wants a pic such that cyclist is at hill-top. Objects are cyclist and hill-top. Once these two object identified by AI module burst shot will be triggered. Timing and customization is decided by AI module.

Figure 6: Cycling

5.3 *Hand-Shake*

The below image captured when Trump and Modi met at the white house. Here the focus is on the very moment both celebrities shake their hands. The AI module first identify the environment and then objects accordingly.

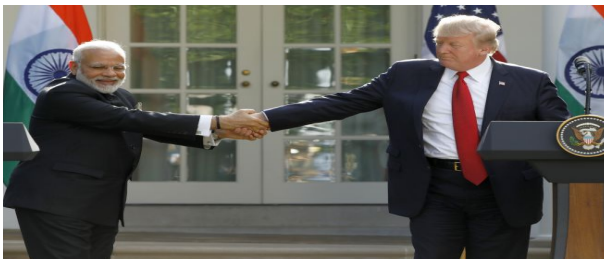

Figure 7: Trump-Modi Hand Shake [11]

5.4 *Capturing User selected moment*

This paper talks about automated burst capture as well as capturing based on user intention. If end user wants to capture a particular frame/shot he is allowed to do that as manual mode feature is also there.

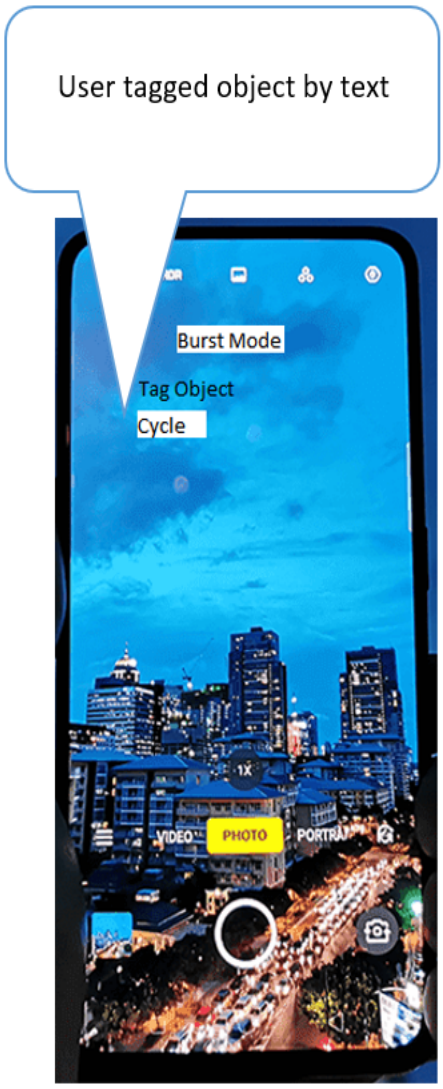

Figure 8: Manual Mode: User is selecting object

6 Conclusion

This paper solves the problem of how an end user needs to track the object before capturing the moving object, also user needs to keep pressing shutter button for capturing burst mode. There is a chance that end user might missed the real shot moment due to manual shooting or because of the fast movement of object it would be difficult for user to handle it properly. This paper also talks about manual mode. Some Benefits also include:

- No need to track the object before capturing the burst mode as it would be dome by AI environmental heuristics
- No need to start and stop burst mode manually
- Better chance of capturing moving object with AI solution as object is being tagged and burst mode has to start when object comes into frame or start moving. In manual burst mode, it is difficult to capture real moving object.
- Memory saving as no unnecessary frame capture.
- Manual Mode shoot. Paper also gives user a flexibility to select his own object and once object is identified according to that burst should be triggered.

7 Acknowledgement

Thanks to Oppo India Camera Team.

References

[1] V. J. Pore, “Selection of an optimum image in burst mode in a digital camera,” Sept. 9 2014. US Patent 8,830,341.

[2] T. Abe, “Digital still camera performing white balance adjustment,” Oct. 18 2001. US Patent App. 09/802,955.

[3] J. I. Woodfill, H. H. Baker, B. Von Herzen, and R. D. Alkire, “Data processing system and method,” Apr. 10 2001. US Patent 6,215,898.

[4] K. Nobori, S. Sato, and T. Azuma, “Image capturing apparatus, monitoring system, image processing apparatus, image capturing method, and non-transitory computer readable recording medium,” Dec. 18 2018. US Patent 10,157,315.

[5] W. T. Cathey Jr and E. R. Dowski Jr, “Extended depth of field optical systems,” May 5 1998. US Patent 5,748,371.

[6] E. Dowski, “Wavefront coded imaging systems,” Aug. 29 2002. US Patent App. 09/747,788.

[7] S. Thrun, “System and method for tagging objects in a panoramic video and associating functions and indexing panoramic images with same,” June 7 2016. US Patent 9,361,943.

[8] R. Szeliski and H.-Y. Shum, “Block adjustment method and apparatus for construction of image mosaics,” Nov. 16 1999. US Patent 5,987,164.

[9] K. Flory, H. Do, W. Stemper, Y. Noguchi, and S. D. Loi, “Continuous burst mode digital camera,” Aug. 31 2010. US Patent 7,787,026.

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

[11] Y. Singh and L. K. Jha, *Modi and Trump: Of hugs and handshakes*, June 27, 2017 16:29 IST.