5C1 Video Processing : Assignment II

Oscar the Cat

1 Jan 2020

---

1 Introduction

1. What is *adaptive streaming* and why is it necessary?

   Adaptive streaming can automatically adjust video quality based on user' bandwidth and CPU capacity. By detecting the condition of end devices in real time, the source content which is encoded at at multiple bitrates and resolutions is delivered, and the player switches between streaming the different encodings.

   Adaptive streaming provides the users of streaming media with better experience since the media server can automatically adapt their network and device conditions and transmit videos with best-possible quality and avoid playback interruptions.

2. List typical resolutions used in modern video streaming. Why are multiple resolutions used in Adaptive streaming?

   Typical resolutions: 360P (480×360), 480P (858×480), 720P (1280×720), 1080p (1920×1080), 2K (2560×1440), 4K(4096×2160).

   Typically, a higher resolution means a higher bitrate and bandwidth usage. Transmitting videos with different resolutions can effectively change the bitrate. Besides, in some bitrate ranges, low resolution shows higher PSNR and thus changing the video resolution can improve video quality and save bandwidth as well.

3. Describe the Netflix "Per title encoding algorithm" and explain its advantages and disadvantages.

Per title encoding algorithm is designed to find the optimal bitrate-resolution pair. In order to achieve this, we can draw RD curves to depict the bitrate-quality relationship for any video source encoded at different resolutions, and it is shown that each resolution has a switching point where the curve starts flattening out. If these points are fitted with a curve, they collectively form a boundary called convex hull. The optimal bitrate-resolution pair is the point that is as close to the convex hull as possible.

Advantages: Per title encoding can give users better video quality based on complexity of titles.

Disadvantages: It needs to encode different bitrates and resolutions, and therefore the encoding costs are relatively high.

2 Part II

1. Describe, explain and justify the algorithm you use to create the representations. Explain how you generate the PSNR that is used to compare representations and make decisions.

The aim of the algorithm is to find the optimal representation for video encoding. Three resolution, 720p, 360p and 180p, is used to select the best resolution. As described above, we first need to draw the RD curve of each resolution, and the diagram is as shown in Figure 1.

The original size of frames is 1280×720. In order to encode 360P/180P videos, the original frame sequence need to be downsampled to the target size and encoded by H.264 codec. After that, the encoded 360P/180P video files are decoded and upsampled to original size so as to make sure the PSNR between raw 720P frames and decoded frame can be calculated correctly. The bitrate-PSNR pairs of 180P, 360P and 720P are drawn in Figure 2 (left).

In order to analysis the optimal representation mathematically, polynomial fitting is utilised to construct mathematical function of these curves (shown in Figure 2 (left)), and fitted functions are as follow:

$$\begin{aligned} f_{180P}(x) = &- 4.439 \times 10^{-12}x^6 + 3.198 \times 10^{-9}x^5 - 9.179 \times 10^{-7}x^4 \\ &+ 0.000134 \times x^3 - 0.01054x^2 + 0.4421x + 11.25 \end{aligned} \tag{1}$$

$$\begin{aligned} f_{360P}(x) = &- 5.07910 \times 10^{-16}x^6 + 1.747 \times 10^{-12}x^5 - 2.375 \times 10^{-9}x^4 \\ &+ 1.627 \times 10^{-6}x^3 - 0.0005962x^2 + 0.1174x + 12.63 \end{aligned} \tag{2}$$

$$\begin{aligned} f_{720P}(x) = &- 6.357 \times 10^{-12}x^4 + 2.8 \times 10^{-8}x^3 - 4.724 \times 10^{-5}x^2 \\ &+ 0.03972x + 13.13 \end{aligned} \tag{3}$$

where $f_{\Delta}(x)$ is the estimated PSNR and x represents the bitrate. By deriving the second order derivative of functions, which measures the rate of change of slope, the **local maximum** within the constrained bitrate range is the switching bitrate. It turns out that the switching bitrates of 180P, 360P and 720P are 45Kbps, 180Kbps, 927Kbps respectively. Furthermore, the function of Convex Hull passing through three switching points can also be fitted by a fourth degree polynomial:

$$\begin{aligned} f_{CH}(x) = &- 1.312 \times 10^{-11}x^4 + 4.28 \times 10^{-8}x^3 - 5.152 \times 10^{-5}x^2 \\ &+ 0.0316x + 17.51 \end{aligned} \tag{4}$$

Ideally, we want to operate exactly at the convex hull, but due to practical constraints a finite number of resolutions available, we would like to select bitrate-resolution pairs that are as close to the convex hull as possible. Given bitrates, by measuring the distance between RD each curve and convex hull, 88 Kbps and 431 Kbps are chosen as thresholds. When the bitrate is greater than 88 Kbps, the performance of 360p streaming is better than 180P, and 720P streaming should be used when the bitrate is greater than 431 Kbps due to higher PSNR than that of 360P. Apart from that, 5Mpbs should be the maximum value of bitrate since there is no significant improvement on PSNR when the bitrate exceeds it.

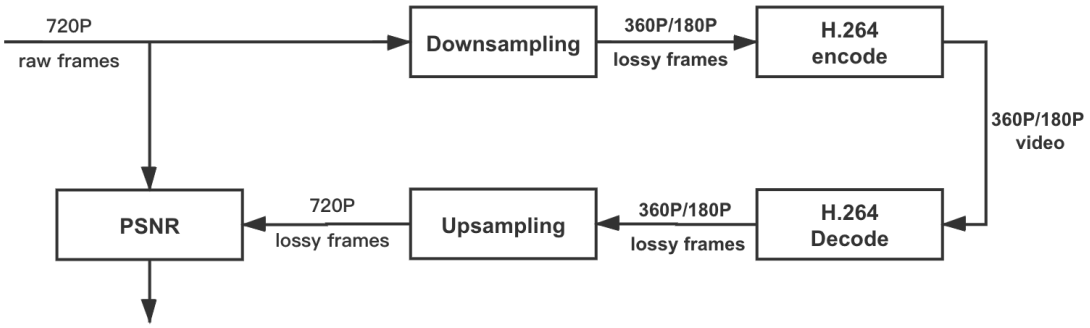

Figure 1: The diagram of measuring PSNR of different representations

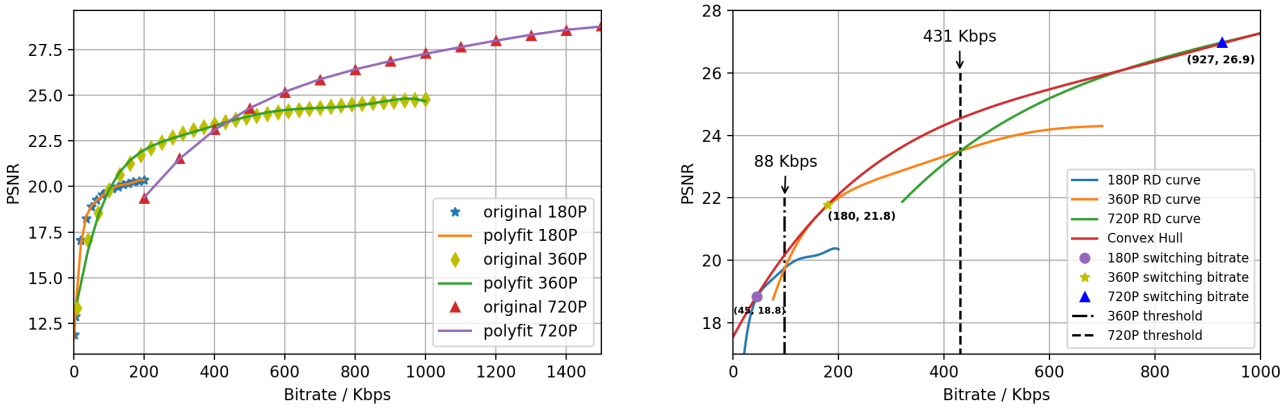

Figure 2: *Left: The original bitrate-PNSR pairs and fitted polynomial functions; Right: Zoomed RD Curves for three representations. 88Kbps and 431Kbps are selected as switching thresholds between 180P, 360P and 720P representations.*

2. Show on a single R/D plot in Figure 2, the RD curves for each of your representations. Place appropriate legends and label your axes appropriately. Show on the plot your estimate for the crossover bitrates.
3. Estimate the bitrate at which the quality of 360p is better than 720p. 88Kbps - 431Kbps
4. Estimate the bitrate at which the quality of 180p is better than 360p. 0Kbps - 88Kbps
5. Explain any observations about these thresholds or the difficulty in establishing them.

We can observe that in the interval less than the threshold 431Kbps, 720P encode may produce a quality lower than 360P by encoding at the same bitrate. This is because encoding more pixels with lower precision can produce a worse picture than encoding less pixels at higher precision combined with upsampling and interpolation. Besides, the encoding overhead associated with every fixed-size coding block consumes more bitrate instead of the consumption by encoding the actual signal.

There are some difficulties in establishing the thresholds. For instance, it is not feasible to trancode the video by every bitrate in practice, so choosing a bitrate stepsize of just-noticeable-difference is a challenging task.

6. The LalaHead representation at 360p has bitrate B_l, and PSNR = 259.5Kbps, 22.5dB
7. Comment on the difference between the LalaHead representation at 360p and your chosen representation properties.

In the case of stable network bandwidth, our algorithm will choose the same representation as LalaHead. However, this strategy is not robust enough for bandwidth changing, and fluctuation of available bandwidth will lead to a decrease in coding efficiency. If the available bandwidth is less than 88Kbps or greater than 431Kbps especially, using a fixed 360P resolution will result in a significant drop in video quality, and our algorithm will switch the resolution to 180P or 720P to provide better picture quality.