Probability Modelling of Intra Prediction Modes

Jean-Marc Valin

July 6, 2015

1 Introduction

Modern video codecs and still image codecs make use of intra prediction where a region (or block) of the image is predicted based on its surrounding. There are usually multiple intra predictor *modes*, each preforming a different kind of prediction. For example, some modes may predict along a particular direction, in which case the selected mode would typically represent the direction of the patterns in the region being coded. The mode is typically selected by the encoder and transmitted to the decoder. The cost of coding the mode can be large for small block sizes, so it is important to efficiently encode the information using entropy coding [1, 2]. The following describes an efficient way of modelling mode probabilities.

2 Probability Modelling

Let $m_{i,j}$ be the id of the intra prediction mode selected for block (i,j). It is useful to know the probability $p\left(m_{i,j}\right)$ to make optimal use of entropy coding when encoding the selected mode in the bitstream. In particular, coding can be made more efficient by making use of the context: modes selected for causal neighbors that are already encoded in the bitstream. For example, it is desirable to estimate $p\left(m_{i,j}\left|m_{i-1,j-1}, m_{i,j-1}, m_{i-1,j}\right.\right)$. Let M be the number of possible modes and N be the number of neighbors considered, then explicit modelling of the conditional probabilities using an explicit table requires $M^{(N+1)}$ entries, which rapidly becomes prohibitive. For example, with 10 modes and using the left, up-left and up blocks as context requires a table with 10,000 entries. This is one reason why the VP8 codec only uses the left and up blocks as context.

It is possible to reduce the size of the context by considering only whether the neighboring blocks use the same mode or a different mode, i.e. modelling the probability $p\left(m_{i,j}\left|m_{i-1,j-1}=m_{i,j},m_{i,j-1}=m_{i,j},m_{i-1,j}=m_{i,j}\right.\right)$ instead of $p\left(m_{i,j}\left|m_{i-1,j-1},m_{i,j-1},m_{i-1,j}\right.\right)$. Because the conditional parameters are now binary (equal or not equal), then a lookup table only requires $M \cdot 2^N$ entries, e.g., only 80 entries when using 3 neighbours. One drawback of this approach is that

$$S_p=\sum_{k=0}^{M}p\left(m_{i,j}=k\left|m_{i-1,j-1}=k,m_{i,j-1}=k,m_{i-1,j}=k\right.\right)\neq 1 \tag{1}$$

so some form of renormalization is required.

The simplest way to renormalize probabilities is to multiply each of them by the same factor such that the sum equals 1. If the entropy coder, such as a classic non-binary arithmetic coder or a range coder, is set up to code symbols using cumulative frequency counts, then simply declaring S_p to be the *total frequency count* allows the entropy coder to perform the renormalization process automatically as part of the coding process. This method is the simplest, but sometimes does not accurately model probabilities when one of the probabilities is close to unity.

A slightly more bit-efficient renormalization procedure is to first "amplify" probabilities that are close to unity:

$$p_k^{'}=p_k\cdot\frac{\left(\sum_j p_j\right)-p_k}{1-p_k}\,, \tag{2}$$

followed by renormalizing the $p_k^{'}$ to have a sum of 1. The step in 2 results in probabilities close to 1 being less affected by the renormalization and slightly improves coding efficiency over the simple renormalization.

2.1 Adaptation

There are two possible ways of adapting the probability model to individual images being coded. The first is to compute an online probability $p_0\left(m\right)$ of mode m being selected when none of the neighbours use mode m. Then, $p_0\left(m\right)$ is used as a floor probability for each mode m. The purpose of this floor is to ensure that when a particular mode is highly used in an image, its cost goes down.

The second way of adapting the probability model is to compute a statistic on a large number of previously encoded modes. For example, this can be the most often used mode in the frame, the percentage of non-directional modes used, the most common direction. This statistic can be used as an additional condition on the probability. In that case, the storage requirement becomes $M \cdot 2^N S$ where S is the number of possible discrete value for the statistic.

References

[1] Moffat, A., Witten, I.H., “Arithmetic coding revisited”, *ACM Transactions on Information Systems (TOIS)*, Vol. 16, Issue 3, pp. 256-294 , July 1998.

[2] Stuiver, L. and Moffat, A., "Piecewise Integer Mapping for Arithmetic Coding", *Proc. of the 17th IEEE Data Compression Conference (DCC)*, pp. 1-10, March/April 1998.