Optimizing bit allocation accross long sequences of compressed video

Paul Cockshott Allin Cottrell David Breslin [1]

1 Introduction

This paper concerns the optimal allocation of bits to individual video frames when compressing a video sequence, where the average number of bits per frame is given as a prior constraint. The optimization algorithm developed here is suitable for offline compression; it requires multiple passes through the video sequence.

The raw information content of frames in a video sequence differs. Some frames have high redundancy, showing large expanses of uniform colour. Others show lots of detail, and as such, inherently require more information to encode them. The conditional information of frame n in a sequence given frame $n-1$ also varies. It is high when a sharp scene change occurs or where there is a lot of rapid motion, lower for fades and wipes, and almost zero for continuous and unchanging scenes. The general problem we are addressing in this paper is how to optimally adjust the bit rate of a compressor to accomodate these changes in conditional information accross a complete video.

Video compression is used in a number of distinct fields, in digital television broadcasting, in video-phone applications, in web-video streaming and in the preparation of recorded video sequences on CDs and DVDs. These applications differ both in bandwidth and in their temporal characteristics.

Television broadcasting is high bandwidth, but has continuous delivery, but need not be real time. Many programs are pre-recorded and are compressed off-line.

Video phone applications are low bandwidth and conversational. This requires real-time compression and low latency.

Web video is low bandwidth, has bursty delivery, and is generally pre-recorded and compressed off-line.

CD and DVD video recording is medium to high bandwidth, always compressed offline, and stored on drives which typically have a sustained delivery rate higher than the bandwidth of the video signal.

The bandwidths and temporal characteristics of the delivery vehicles enforce constraints on how a compressor/decompressor pair can interact with the varying information content of the input signal. MPEG I and II, designed in the high bandwidth context of broadcast digital television attempt to deal with the variation by the use of key-frames and inter-frames. This is a local optimization strategy constrained by the limited buffer capacity of the decoder. It accomodates adequately to scene changes but does not allow the transfer of bits from long sequences with relatively little change, to other more demanding parts of a video.

Conversational video-phones have to have low latencies to allow fluent interaction between users. In these circumstances the delivery of a constant number of bits per frame may be the best trade-off between the conflicting requirements for quality and of latency.

Web video along with video distributed on CDs offers a wider range of rate control choices. Such sequences are typically delivered at a mean bitrate that is small compared to the buffering capacity of the RAM on a modern PC, or are delivered from a drive whose peak transfer rate is several times the mean bitrate of the compressed signal. This means that the rate-control algorithms used are less constrained and can attempt more global optimisations.

Our algorithm was developed in the context of the Strathclyde Compression Transform (SCT) [4, 5], a hierarchical vector quantisation codec. To date the SCT has employed a rate control that minimizes end-to-end delay rather than maximizing perceptual quality. On the basis that it takes a finite time to transmit each bit in a "real-time" mode of operation, each frame is compressed up to the point where it utilizes the bits available per second (bandwidth) divided by the frames encoded per second (frame rate). While suitable for applications such as video phones where delay is critical, the minimal delay rate control is inappropriate for off line operation where file size rather than transmission delay is an issue.

This research was motivated by the hypothesis that a significant improvement in quality ought to be obtainable by varying the allocation of bits to frames in the light of a well-defined optimization procedure, based on the eco-

[1] Department of Computing Science, University of Glasgow, Department of Economics, Wake Forest University, and Department of Computer Science, University of Strathclyde, respectively. The authors gratefully acknowledge support from the UK Engineering and Physical Sciences Research Council.

nomic principle of equalization of marginal return. For reasons given in section 2.1 below, the SCT is particularly suitable for such a form of rate control.

Section 2 presents the relevant theoretical argument, describes the optimization algorithm, and discusses the properties of the algorithm. Section 3 presents experimental results, and section 4 concludes.

2 Error and marginal error

As progressively more bits are allocated to the compressed representation of a given video frame the representation should, converge towards the original image. This is illustrated in Figure 1.

As more bits are devoted to the current frame the total error falls and the marginal error per bit sent rises. The marginal error is the increment in the total error associated with the addition of one bit to the compressed representation of the given image. It is negative since the addition of bits reduces total error. On different frames the total and marginal error curves may approach zero at different rates, as shown in Figure 2.

2.1 Marginal principle

Consider first the allocation of bits between two frames. Given two marginal error curves, e'_1 and e'_2, for frames 1 and 2 respectively, an optimal allocation of bits between frames, (b_1, b_2), must satisfy the condition that

$$e'_1(b_1) = e'_2(b_2).$$

Were this not the case the pair of pictures could be improved by shifting bits from the frame with the greater (absolute) marginal error to the frame with the lower marginal error.

This principle is a commonplace of economics, where it appears in many guises, notably neoclassical production theory. In this theory the alternative inputs to production are viewed as continuous variables. The production function itself is viewed as a continuous function, differentiable in all of its inputs. In addition all of the alternative production processes to which resources could be allocated are taken to be differentiable functions. The condition for optimal resource allocation is then that all resources are so distributed that each faces the same marginal return in every utilized production process.

We view each frame of the video sequence as a production process; the bits allocated to the frame are the continuously distributable resource. This is an idealization because the encoding process allocates an integral number of bits to each frame. The granularity of bit allocation is, however, sufficiently small that the abstraction of continuity remains reasonable.

Marginalist economic techniques have already influenced the literature on rate control. Fox [2] first introduced the marginal principle to problems of discrete optimization. Shoham and Gersho [3] extended to rate control the principle that, given a finite set of quantizers, the optimal bit allocation involves all quantizers being at a common slope λ of a rate–distortion curve.

Gersho's work borrows from economic optimization theory [1] the technique of Lagrange multipliers. The use of Lagrangians has continued in recent literature [6, 7, 9] addressed to the problem of rate control in DCT type codecs. These codecs differ from the SCT in that the within-frame compression algorithms typically do not follow a "best first" approach. The best first algorithm sorts possible differential updates that could be applied to construct frame t from frame $t-1$ in descending order with respect to the reduction in frame error they will produce. To a first approximation, this gives a monotonic marginal error curve of the sort shown in Figure 2. (but see the further discussion in section 2.4.1). It is this property of the SCT that makes it a particularly fruitful field for the application of neoclassical optimization techniques. It may be noted that Fox's original presentation of the application of marginal analysis to problems of discrete optimization essentially involved the use of best-first.

2.2 The algorithm

Returning to the two-frame example, in addition to the equalization of marginal error we know that the sum of the bits allocated to the two frames, $b_1 + b_2$, must not exceed B, the total available bits. The optimality condition is therefore fully constrained, and may be readily extended to more than two frames. All frames must have the same marginal error on the last bit sent for that frame and the sum of allocated bits must be less than or equal to the total bit budget. (In a practical situation it is likely that the optimal solution will involve exhausting the bit budget.) An algorithm to arrive at such an allocation is set out below.[2]

1. Make an initial guess at the threshold marginal error to be used as a cutoff, i.e. the common value of e'. Call this T_0.
2. Compress all frames in the sequence using T_0 as a cutoff for the allocation of bits to frames. While compressing each frame record the number of bits allocated.
3. At the end of the run, find the total number of bits allocated to the whole sequence, $\hat{B}_0$.

[2]The idea of using an iterative algorithm to arrive at an optimum allocation goes back a long way in the economics literature—see the notion of "tâtonnement" in Leon Walras [8] (first French edition Lausanne, 1874).

4. If $\hat{B}_0$ differs significantly from the available bit budget, B, set a new threshold using negative feedback from the bit allocation error:

$$T_{j+1} = \left(1 - \alpha \frac{B - \hat{B}_j}{B}\right) T_j \qquad (1)$$

where j indexes the steps of the iteration and α (> 0) is an "acceleration" parameter.

5. Repeat steps 2 to 4 until the total number of bits used is acceptably close to B.[3]

The first step of such an iteration is shown in Figure 3. The parameter α is set to 1.0 (as was the case in our experimental work, so that equation (1) simplifies to $T_{j+1} = (\hat{B}_j/B)T_j$.

Two questions arise in relation to this algorithm. Does it necessarily converge? And if it does converge, is it guaranteed to find the optimal allocation of bits between frames?

2.3 Convergence of the algorithm

The algorithm as stated above is not guaranteed to converge for all video sequences. Consider the case of $\alpha = 1.0$ in (1), and suppose that at iteration step j the bit budget is in surplus ($\hat{B}_j < B$) to the extent of 5 percent of B. In that case the marginal error threshold will be adjusted upward by 5 percent at step $j + 1$. Suppose the curvature of the aggregate marginal error schedule is such that this throws the bit budget into deficit to the extent of 10 percent of B. Then the threshold will be adjusted downward by 10 percent at step $j + 2$, overshooting its value at step j and starting a divergent movement.

To preclude such divergence the rule given above is modified as follows. Writing $\hat{T}_{j+1}$ for the threshold value derived by application of (1), the actual threshold is limited thus:

- If $\hat{B}_j < B$ and $\hat{T}_{j+1} > T_{\max}$ then $T_{j+1} = (T_{\max} + T_j)/2$.
- If $\hat{B}_j > B$ and $\hat{T}_{j+1} < T_{\min}$ then $T_{j+1} = (T_{\min} + T_j)/2$.
- Else $T_{j+1} = \hat{T}_{j+1}$.

$T_{\max}$ is initialized to zero at the start of iteration, and $T_{\min}$ to a large negative number. Thereafter $T_{\max}$ is updated to equal the lowest (i.e. largest absolute) T_j value for which $\hat{B}_j$ is found to exceed B, while $T_{\min}$ is updated to equal the highest (i.e. smallest absolute) T_j for which $\hat{B}_j$ is found to fall short of B.

[3] If the sequence of frames to be compressed is large and there is a concern with the speed of the computation of the optimal allocation it may make sense to operate with a sample of the frames rather than the entire set, at least in the earlier rounds of the iteration.

2.4 Optimality of the algorithm

Given convergence, the following two conditions are jointly sufficient for the algorithm set out above to produce the optimal allocation of bits between frames.

1. The marginal error curve for each frame is everywhere monotonically increasing.
2. There is no interdependency among frames, in the sense that the marginal error curve for frame t is unaffected by the allocation of bits to the build of frame s, $s \neq t$.

Let us consider these conditions in turn.

2.4.1 Monotonicity of marginal error curve

If the condition of monotonicity of the per-frame marginal error curve is not met, that means that the build of a given frame may be cut off, by the threshold condition, when there remain unexploited opportunities for making improvements with a greater absolute marginal error than the chosen threshold.

As mentioned above, the SCT uses a best-first algorithm when selecting the next incremental improvement to make to the representation of any given frame. It would seem that this ought to ensure that the marginal error curve is strictly monotonic but that is not the case, for two reasons.

First, the marginal unit of information, so far as the SCT's best-first algorithm is concerned, is not the individual bit but rather the "packet" of bits required to code some definite improvement to the image, chosen from its repertoire of vector quantization and motion compensation. These packets are not all of the same size; typically they vary in the range of 15–32 bits but some packets may fall outside of that size range. When calculating the marginal error for the purposes of optimization, however, what matters is the improvement per bit. If the marginal error curve is monotonic when expressed on a per-packet basis, it need not be monotonic when expressed in per-bit terms.

This might seem to be a readily remediable weakness in the SCT, but matters are not so simple. It is cheaper, in terms of the information required for addressing, to code improvements in the neighbourhood of existing "build". If best-first is set to operate strictly on a per-bit basis this tends to produce spatial clustering of high-frequency detail in the compressed image, which is perceptually inferior to a relatively even spread of detail. In principle it ought to be possible to overcome this unwanted side effect by amending the objective function appropriately. For instance, instead of using the simple minimand of mean squared error, pixel by pixel, one could use a combination of mean squared error and a measure of the dispersion of error such as variance and/or spatial autocorrelation. This may be a topic for future work.

The second reason for non-monotonicity in the marginal error curve relates to interdependencies in the build process for a given frame. Best-first selects the option that makes the largest possible improvement to the frame, given the menu of compression possibilities open at that particular stage of the build. Due to phenomena such as block occlusion in the context of variable dimension vector quantization, however, it may be that while improvement I is the best available at step s of the build, carrying out I then makes possible a larger improvement at step $s+1$.

At any rate, while the SCT does not produce a strictly monotonic marginal error curve, it does produce a reasonably close approximation—close enough for the bit-allocation algorithm to be effective, if not demonstrably optimal. For practical purposes, to avoid the situation where the build of a given frame is cut off prematurely owing to a local upspike in the marginal error curve, we specify the cutoff condition as follows: rather than building to the point where a single improvement step yields a marginal error of smaller absolute value than the threshold, we require two consecutive such steps before cutting off.

2.4.2 Interdependency between frames

If marginal error curves are monotonic and there is no interdependency between frames—i.e. if the allocation of extra bits to frame t leaves the marginal error curve for frame s, $s \neq t$, unaffected—then we can be sure that compression of all frames up to a common marginal error threshold produces the lowest possible error, for any given bit budget, for the sequence as a whole. In general, interdependence removes this certainty. Take the simple case of two frames, F_1 and F_2, and suppose these frames are compressed up to a common marginal error, at which point the bit allocation (b_1, b_2) just exhausts the bit budget B. Now consider the effect of moving Δb bits from F_2 to F_1. Given independence this is sure to increase the aggregate error, since the reduction in error for F_1 will be more than offset by the increase in error for F_2. If, however, devoting additional bits to F_1 shifts the marginal error curve for F_2 it seems we can no longer be sure that the effect of the reallocating of Δb bits must be to raise the aggregate error.

It is clear that the independence condition is not in fact satisfied by the SCT. Various subtle forms of dependence may be present, but the most obvious form derives from the fact that the compressed representation of frame $t-1$ forms the starting point for the build of frame t. Thus if frame t does not differ radically from $t-1$, allocating more bits to frame $t-1$ is likely to reduce the starting value of the error in the build of t, and hence is also likely to raise (reduce the absolute value of) the marginal error schedule for frame t.

Fortunately, this sort of interdependence is not damaging to the procedure advocated above. Note that dependency is both uni-directional and localized: uni-directional, because it is in the nature of the SCT's linear procedure that the degree of build for a given frame cannot have any effect on the marginal error curve for previous frames; and localized, because the slate is wiped clean, so to speak, at every scene change. Frames are interdependent only to the extent that they share visual elements. This means we can partition each frame into two components, the elements that are shared with the previous frame and those that are new: each of these components will have an associated marginal error curve. For the shared elements, the marginal error curve in frame t will be a continuation of that in frame $t-1$. If a threshold marginal error T cut off construction of the shared elements after b_{t-1} bits in frame $t-1$, then that same threshold will cause no bits to be allocated to handle the shared elements in frame t. It follows that all bits used in frame t will relate to the building of non-shared elements, which are independent of the previous frame, and hence the marginal principle retains its validity.

As with the issue of monotonicity of the marginal error curve, therefore, our view is that while the independence condition is not strictly satisfied, neither is it violated in a way that jeopardizes the effectiveness of the proposed procedure.

2.5 Mean Squared Error and PSNR

Optimization via equalization at the margin is a very general method; it can be used to find the extremum of any chosen figure of merit within a budget constraint. Our work has concentrated on marginal error defined as the change in Mean Squared Error (MSE), and hence on the minimization of MSE for the sequence of frames as a whole. It is also possible to work in terms of PSNR; in this case one compresses each frame in the sequence up to a common threshold value of marginal PSNR, which will produce the effect of maximizing the mean of the per-frame PSNRs across the sequence as a whole. The latter approach is of some interest, particularly since PSNR is the most commonly quoted figure of merit in the video compression literature, and we show results of this sort below. We are doubtful, however, that this approach will give results as good to the eye as those obtained via minimization of MSE.

The reason for this doubt is illustrated in Table 1. Consider two compressed images with current MSE values of 100 and 1000 respectively, relative to their uncompressed counterparts. We have Δb additional bits to allocate to these images and we wish to decide which frame should get them. Adding the extra bits to frame 1 would reduce its MSE to 90, while adding the same number of bits to frame 2 would reduce its MSE to 980. According to the criterion of minimum MSE, the bits should clearly go to frame 2, where they achieve a greater reduction in squared error.

The formula for frame PSNR is $10\log(k/\mathrm{MSE})$, where the log is to the base 10 and k is a constant that depends on the number of bits per pixel. The change in PSNR that results from devoting extra bits to a frame, and hence lowering the frame's MSE from MSE_0 to MSE_1, is then

$$\Delta\mathrm{PSNR} = 10\log\left(\frac{k}{\mathrm{MSE}_1}\right) - 10\log\left(\frac{k}{\mathrm{MSE}_0}\right)$$

$$= 10(-\log\mathrm{MSE}_1 + \log\mathrm{MSE}_0) = 10\log\left(\frac{\mathrm{MSE}_0}{\mathrm{MSE}_1}\right)$$

The numerical results of this calculation for the example are shown in Table 1: the criterion of maximum PSNR will lead to the allocation of the extra bits to frame 1. This seems wrong: the bits are going to where they achieve a smaller reduction in MSE, and to a frame that is already relatively "good" in the sense that its MSE is low, rather than being used to achieve a bigger error reduction for a frame that is substantially in error. Mathematically, the reason why the maximum PSNR criterion gives the bits to frame 1 is that they achieve a larger *percentage* reduction in MSE there, but we are sceptical that this is the right thing to do on perceptual grounds.[4]

	Frame 1	Frame 2
MSE_0	100	1000
MSE_1	90	980
ΔMSE	-10	-20^*
ΔPSNR	$10\log\left(\frac{100}{90}\right)$ $= 0.458^*$	$10\log\left(\frac{1000}{980}\right)$ $= 0.088$

Table 1: Minimization of MSE versus maximization of PSNR

3 Experimental results

We show the results of applying the algorithm in compressing 100 frames from Strathclyde University's "Tour of Glasgow" sequence. Here we used a tight bit budget of 2400 bits per frame. The frame rate is 12 frames per second. The three lines in Figure 4 show the PSNR for each of the 100 frames, for each of three compression variants: a constant bit rate, compression to a common threshold for marginal squared error, and compression to a common threshold for marginal PSNR.

Summary statistics relating to the same three compression runs are given in Table 2.

Perhaps the most striking difference made by rate control based on the equalization of marginal squared error across frames is the big reduction in the variance (or standard deviation) of the PSNR. This is evident both from Table 2 and Figure 4. It is also quite clear from watching the respective sequences: the rate-controlled version looks much smoother. The tendency for a sharp deterioration in perceived quality at each scene change, characteristic of constant bit-rate compression, is substantially mitigated. This comes at the cost of a somewhat lesser degree of build of detail (lower PSNR) for frames appearing at a later stage in each scene, but this is much less noticeable to the eye. These preliminary results are quite encouraging.

4 Conclusion

We believe that the algorithm described here is of quite general use in the optimisation of pre-recorded video sequences for distribution either on optical media or, with suitable pre-buffering in the receiving PC, for webcasts. Whilst we have demonstrated it in the context of a vector quantisation based codec, the basic principles are independent of the codec used. The same basic algorithm could be applied to streams compressed by wavelets or the DCT.

The algorithm borrows from well established principles of neo-classical economic theory, just one example of the fruitful transplant of ideas from another discipline to computer engineering.

References

[1] H. Everett, "Generalised Lagrange Multiplier method for solving problems of optimum allocation of resources", *Operations Res.*, vol 11., pp. 399–417, 1963.

[2] B. Fox, "Discrete Optimization via Marginal Analysis", *Manage. Sci.*, vol 13, no. 3, pp. 210–216, Nov. 1966.

[3] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers", *IEEE Trans. Acoust., Speech, Signal Proc.*, vol 36, pp. 1445–1453, Sept. 1988.

[4] R. B. Lambert, R. J. Fryer, W. P. Cockshott and D. R. McGregor, "A comparison of variable dimension vector quantization techniques for image compression", *Proc. EC-MAST'96*, vol. II, pp. 655–670, May 1996.

[4] On theoretical grounds the minimization of MSE should also maximize the sequence PSNR, while the maximization of the mean of the per-frame PNSRs will not achieve this effect. If one is calculating percentages globally, rather than frame by frame, the incremental compression step that achieves the greatest reduction in MSE will also achieve the greatest *percentage* reduction in MSE, and hence the greatest increase in PSNR of the sequence.

[5] R. B. Lambert, R. J. Fryer, W. P. Cockshott and D. R. McGregor, “Low bandwidth video compression with variable dimension vector quantization”, *Proc. Advanced Digital Video Compression Engineering*, Cambridge, UK, pp. 53–58, Jul. 1996.

[6] K. Ramchandran, A. Ortega and M. Vetterli, “Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders”, *IEEE Trans. on Image Proc.* 3(5), pp. 533–545, Sept 1994.

[7] Chi-Yuan Hsu and A. Ortega, “A Lagrangian optimization approach to rate control for delay-constrained video transmission over burst-error channels”, *Proc. ICASP’98*, 1998.

[8] L. Walras, *Elements of Pure Economics* (tr. W. Jaffé), Homewood, Ill.: Irwin, 1954.

[9] J. Zdepski, D. Raychaudhuri and K. Joseph, “Statistically based buffer control policies for constant rate transmission of compressed digital video”, *IEEE Trans. Communications*, vol 39, no. 6, pp. 947–957, June 1991.

5 Figures

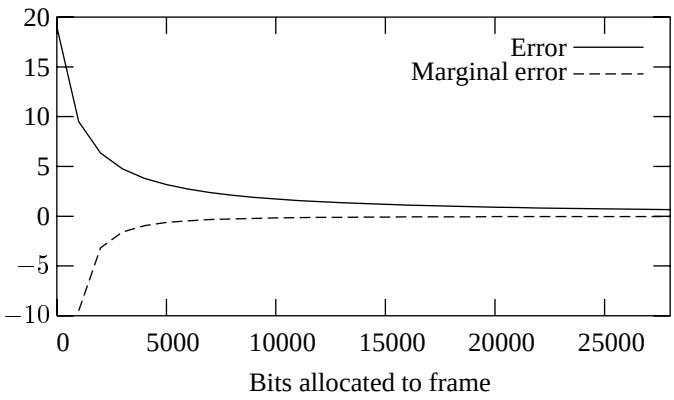


Figure 1: Error and marginal error against bits allocated to a given frame

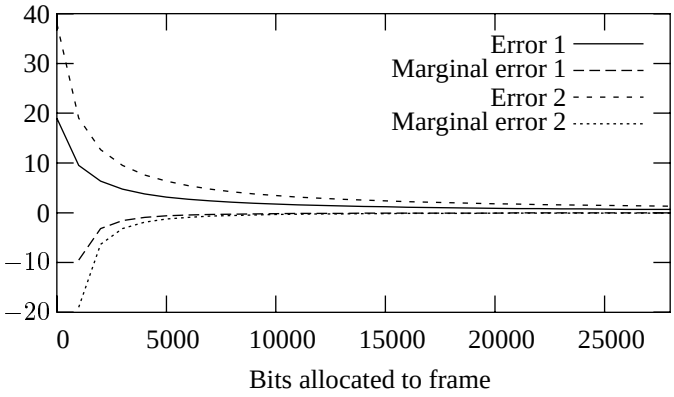


Figure 2: Error and marginal error for two hypothetical frames

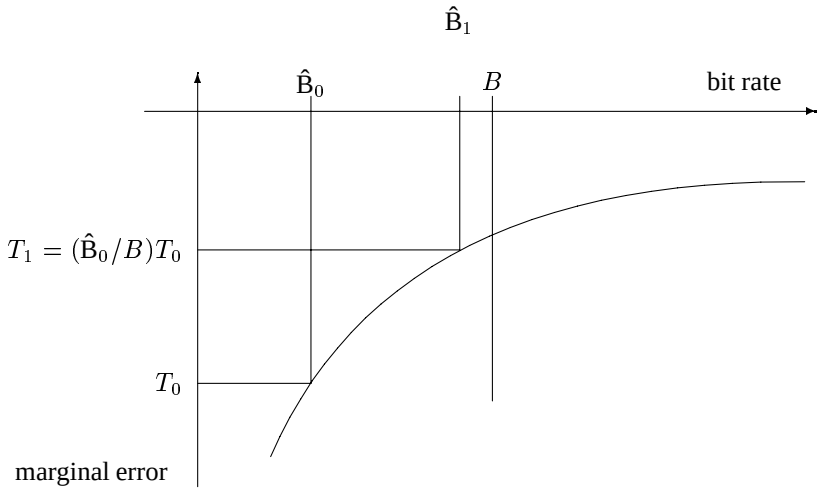
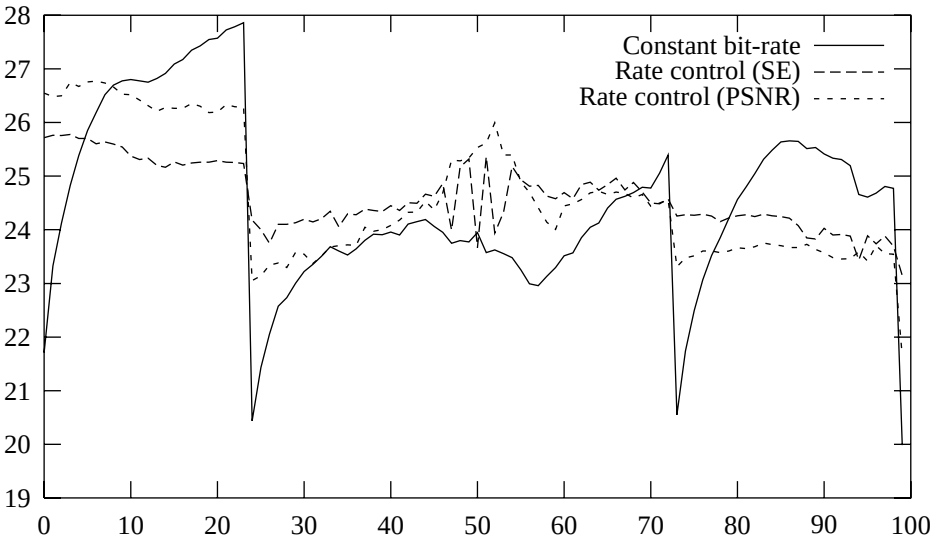


Figure 3: Iteration towards the correct bit allocation

	mean	s.d.	min	median	max	bits used
No rate control	24.497	1.649	19.986	24.172	27.860	240032
Rate control (SE)	24.603	0.607	23.146	24.490	25.775	241345

Table 2: PSNR statistics, with and without rate control



graph showing comparative psnr per frame with and without the equal marginal error algorithm

Figure 4: PSNR Comparison, 100 frames, budget = 2400 bits/frame

Figure 5: The upper pair of images show successive frames with the rate control algorithm described in the paper. The lower pair show the same frames at the same average bit rate of 2400 bits per frame but with a constant number of bits per frame.