Nature Inspired Meta-Heuristics for JPEG Quantization Table Optimization

Pedro Henrique Guimarães Ferreira[1]

[1]Departamento de Engenharia Elétrica
Universidade de Brasília

Trabalho de Conclusão de Curso - Bacharelado em Engenharia Elétrica

Table of Contents

Table of Contents

Introduction

- Combined, motion and still pictures respond for over 80% of Internet's traffic, which makes image compression paramount.
- The image compression scenario has been dominated by the Joint Photographic Experts Group (JPEG) compression standard in the last 25 years.
- The widespread adoption of JPEG provides both advantages and disadvantages: While it guarantees an almost universal availability of dedicated hardware, ensuring very fast JPEG encoding and decoding cycles, this extensive hardware support also shields JPEG from the competition since it turns very hard to any other bare standard outperform a dedicated-hardware implementation of JPEG.
- Therefore, a promising way to propose image compression enhancements is to develop modifications of the top of JPEG, thus being able to benefit from the extensive existing hardware capabilities.

Introduction

- Since JPEG has support to user-provided Quantization and Encoding Tables as native features, a lot of research has been conducted in the generation of improved tables.
- As the Quantization step is the only lossy stage in JPEG, most of the compression is achieved through it. Hence, designing improved quantization tables raises as the leading approach to improve JPEG's performance.
- Finding an improved JPEG quantization table can be seen as a non-smooth multi-objective optimization problem. Therefore, numerical optimization techniques, such as evolutionary computing and swarm intelligence, are suitable for determining an optimized quantization table.
- This work lies in the application of two nature-inspired numerical optimization meta-heuristics, Particle Swarm Optimization (PSO) and Dual Simulated Annealing (DSA) to determine image-specific quantization tables that optimize the rate-distortion compromise for a given image.

Table of Contents

Image Compression

Figure: A RGB image can be decomposed on three channels or bands: a red, a green and a blue

Figure: A YC_rC_b image also can also be decomposed on three bands: one of luminance and two of chrominance

Image Compression

- Both in RGB and YC_rC_b color spaces, a naive representation of a colored digital image would take 3 values per pixel. For a classic bit depth of 8, a rate of 24 bits per pixel would be achieved, implying in a file size of approximately 5.93 MB for a Full HD (1920x1080) image, or a 4.2 GB for a 24 fps 30s Full HD video.
- This representation would also be able to describe $2^{24} = 16,777,216$ different colors, much more than the human eye can properly distinguish.
- Therefore, smaller image file sizes can be achieved both through more efficient implementations and redundancy removal, fostering the need for image compression techniques.

The JPEG Still Picture Compression Standard

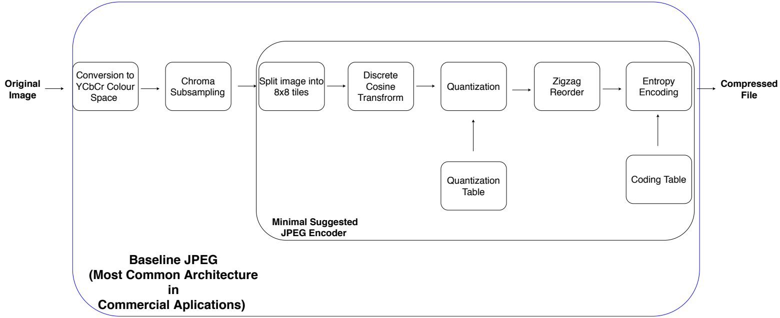

Figure: A Block Diagram representation of the JPEG Encoding Process

- The JPEG compression process is composed of five mandatory and two optional but usual steps

The JPEG Still Picture Compression Standard

- The mandatory steps are:
  - Splitting the image into 8×8 tiles
  - Applying a Discrete Consine Transform (DCT) in each time
  - Quantizing each tile, according to the provided Quantization Table
  - Reordering the quantized coefficients in the ZigZag Scan order
  - Entropy Encoding the reordered coefficients, according to the provided coding table
- The facultative steps, applied before submitting the image to the mandatory stages, are:
  - Converting the image to the YC_rC_b color space
  - Performing Chroma sub-sample
- This work employs the IJG JPEG's implementation as reference implementation, therefore always performing all the optional stages, encoding the YC_rC_b information in the 4:2:2 format.

The JPEG Still Picture Compression Standard

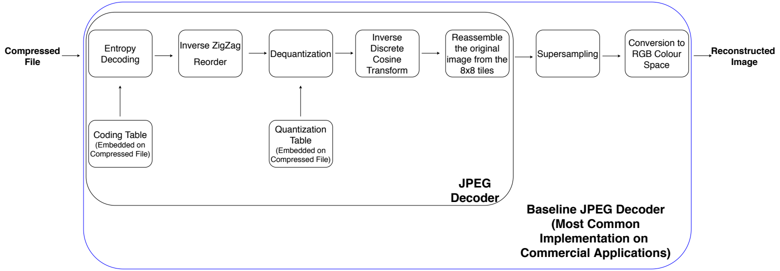

Figure: A Block Diagram representation of the JPEG Decoding Process

- The JPEG decoding process is also composed of five mandatory and two optional but usual steps

The JPEG Still Picture Compression Standard

- The mandatory steps are:
  - Entropy decoding, using the embedded coding table
  - Inverse ZigZag reordering
  - Dequantization of the tiles, using the embedded quantization table
  - Inverse Discrete Cosine Transform (IDCT) on each tile
  - Reassembling the original image from the 8x8 tiles
- The facultative steps, on the other hand, are applied if the encoding facultative steps were applied and after the mandatory stage. They are:
  - Chroma supersampling
  - Converting the image back to its original color space (usually, RGB)

The Rate-Distortion Trade-Off

- Although it is desirable to minimize both rate and distortion, it is usually not possible, since they are positively correlated.
- Since we aim to obtain a image that conciliates high-quality with small file sizes, it is pretty neat that minimizing just the image rate or just the intended distortion is pointless, forcing us to look for the best compromise of both instead.

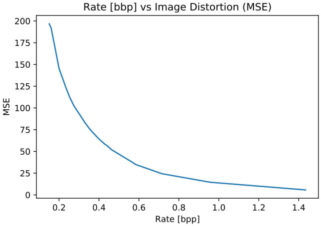

Figure: Rate vs Image Distortion

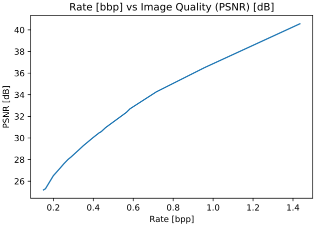

Figure: Rate vs Image Quality

Table of Contents

Particle Swarm Optimization (PSO)

- The PSO is a derivative-free, population-based meta-heuristic that was originally intended to simulate the behavior of a flock of birds. Like a bird in a bird flock, a particle in a particle swarm both benefits from and helps to build social knowledge, allowing the entire swarm's to converge to the best position - i.e, the position that maximizes a given reward, such as the survival chance for a bird flock.
- In formal terms, PSO is defined by the following equations :

$$\begin{aligned} V_i^{t+1} &= C_0 V_i^t + C_1 R_1 \left(P_i - X_i^t\right) + C_2 R_2 \left(G - X_i^t\right) \\ X_i^{t+1} &= X_i^t + V_i^{t+1} \end{aligned} \tag{1}$$

where X is the position vector of the candidate solution, V is the velocity of the particle, P and G are particle's and swarm's best-known positions, R_1 and R_2 are random values sampled from a uniform distribution, and C_0, C_1 and C_2 are constants.

Dual Simulated Annealing (DSA)

- The Dual Simulated Annealing (SA) is a stochastic global optimization meta-heuristic based in emulating the process of internal energy minimization in physical systems.
- For this purpose, in each iteration, a random point is chosen, drawn from a probability distribution where the probability decreases as the distance from the current point increases, and its fitness is evaluated. If the new point has better fitness than the current point, it surely becomes the next point. Else, it has a small probability to become the next point. This probability, denominated the acceptance probability of a transition, is determined both by the fitness values of the current and the new position and by a time-varying hyper-parameter called Temperature.

Table of Contents

Quantization Table Optimization

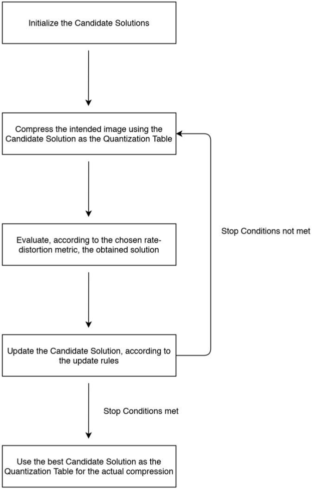

Figure: Fluxogram of the nature inspired metaheuristic behaviour on optimizating JPEG quantization tables

Fitness Functions

- For optimization algorithms, the fitness function works as a proxy to the optimization target. Since there is an implicit trade-off between rate and distortion, what this work aims to optimize is the rate-distortion compromise rather than just rate or distortion.
- For modeling the rate-distortion compromise, the classic approach in the literature is to resort to the Lagrangian rate-distortion function.
- The Lagrangian cost function establishes a static, linear exchange rate between image rate and image distortion, intrinsically assuming that a change in the image rate should produce a proportional change in the introduced distortion among all the support of possible image rates, which is mostly not true.

Lagrangian Cost Function

$$CF = MSE + \lambda R$$

Fitness Functions

- One possible way to cope with the non-linearity in the rate-distortion relation is to assume that the rate and distortion have a fixed exchange rate just around a given operating point. To generalize this idea, defining a self-adaptive metric that requires very few or actual no hyper-parameter tuning, we have proposed a new metric called The Fixed Quality Expected Rate Gain (FQ-ERG), defined by the following equations:

The Fixed Quality Expected Rate Gain (FQ-ERG)

$$ERG = \frac{R}{E(D)}$$

$$FQ - ERG = \begin{cases} ERG, & if \quad |D - D^*| \leq \epsilon \\ ERG + |D - D^*| \cdot P, & if \quad |D - D^*| \leq \epsilon \end{cases}$$

Fitness Functions

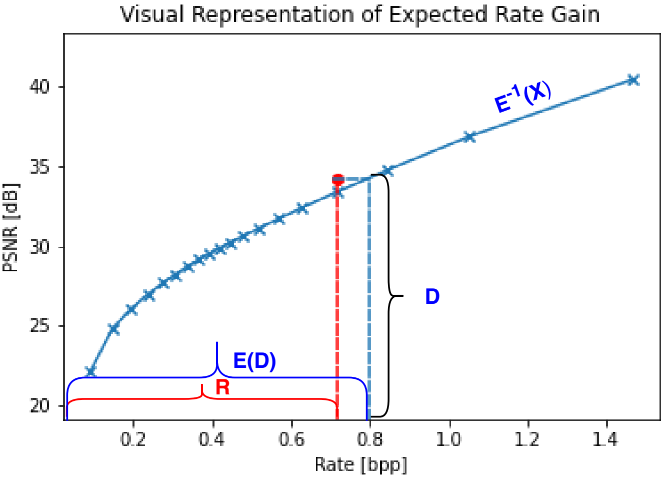

Figure: Graphical Representation of the Expect Rate Gain for a given operation point: The ERG corresponds to the ratio between the actual point rate, R, and

Implementations

- In this work, the Independent JPEG Group's JPEG still image codec v9 encoder is employed as our reference JPEG encoder implementation.
- For Particle Swarm Optimization, the DEAP library implementation is adopted as reference.
- On its turn, the Scipy implementation of Dual Simulated Annealing is chosen as our DSA reference implementation.

Hyper-Parameters

- For the Particle Swarm Optimization, the following hyper-parameters were set:
  - Population Size = 20
  - Maximum Number of Generations = 50
  - Maximum Local Update Factor = 2.0
  - Maximum Global Update Factor = 2.0
  - Minimum Speed = -3.0
  - Maximum Speed = 3.0

  .
- In Dual Simulated Annealing, all parameters but the maximum number of iterations are kept as default. The maximum number of iterations is set to 1,000.

Setting

- To verify the proposed concept, both Particle Swarm Optimization and Dual Simulated Annealing are employed to generate custom, image-specific quantization tables for every image on the *Kodak Image Dataset* using both the Lagrangian Cost Function and the Fixed Quality Expected Rate Gain as fitness functions.
- The *Kodak Image Dataset*, presented in the next slide, is composed by 24 uncompressed, 768 x 512 or 512 x 768 true color images.
- For each image, custom quantization tables were generated for 19 different target quality factors, ranging from $q = 5$ to $q = 95$ in steps of 5
- When optimizing the Lagrangian Cost Function, the value of the Lagrangian multiplier λ for each quality factor and image was determined through a grid search.

Kokak Image Dataset

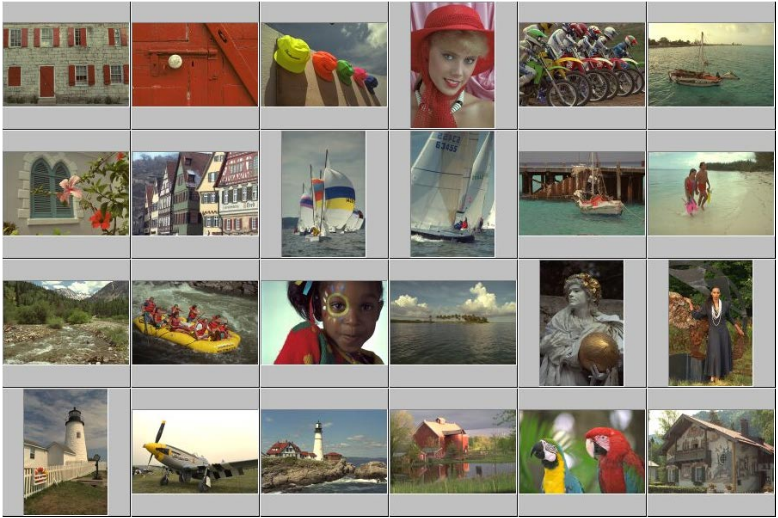

Figure: The Kodak Image Dataset

Table of Contents

Evaluation Metrics

- Since we intend to compare the performance of different JPEG rate-distortion curves, the Bjøntegaard Delta was chosen as our evaluation metric.
- Bjøntegaard Delta Rate (BD-Rate): The BD-Rate compares a candidate method rate-distortion curve against an anchor curve, in our case the baseline JPEG, assessing the percentage increase on the file sizes to obtain the same quality image. A negative BD-Rate implies that the same quality image files produced by the candidate method are smaller than baseline JPEG ones.
- Bjøntegaard Delta PSNR (BD-PSNR): The BD-PSNR compares a candidate method rate-distortion curve against an anchor curve, in our case the baseline JPEG, assessing the difference of PSNRs between the candidate method and JPEG baseline generated images for same-sized files. A negative BD-PSNR implies that the candidate method produces images that have worse quality than baseline JPEG ones.

Results

Method	BD Rate	BD PSNR
PSO FQ-ERG	-8.26	0.51
PSO Lagrangian	-8.37	0.52
DSA FQ-ERG	-7.92	0.49
DSA Lagrangian	-3.90	0.26
Pointwise Best	-9.38	0.59

Table: Average Bjøntegaard Delta values for each approach (the baseline is the Baseline JPEG)

Results

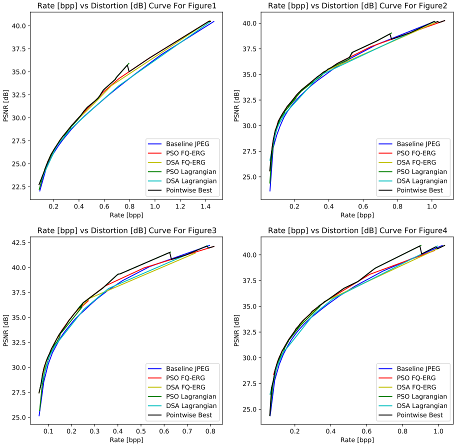

Figure: Rate-Distortion Curve for the Images Kodak1-Kodak4

Significance Test

- Since our methods were evaluated on a limited dataset, it is important to be very careful before claiming that a method outperforms others. To verify if any claim on this direction can be made, we have performed a Pairwise T-Test.
- The null hypothesis of the Pairwise T-Tests is that the two paired samples are significantly different.

Sample 1	Sample 2	t-statistic	P-Value	Decision
PSO FQ-ERG	Lagrangian PSO	-0.56	0.58	Reject the Null Hypothesis
DSA FQ-ERG	Lagrangian DSA	13.76	0.00	Failed to Reject the Null Hypothesis
Lagrangian PSO	DSA FQ-ERG	-1.53	0.14	Reject the Null Hypothesis
PSO FQ-ERG	DSA FQ-ERG	-1.59	0.13	Reject the Null Hypothesis

Table: Pairwise T-Test for the BD Rates of Proposed Algorithms

Coefficient Change - Luminance

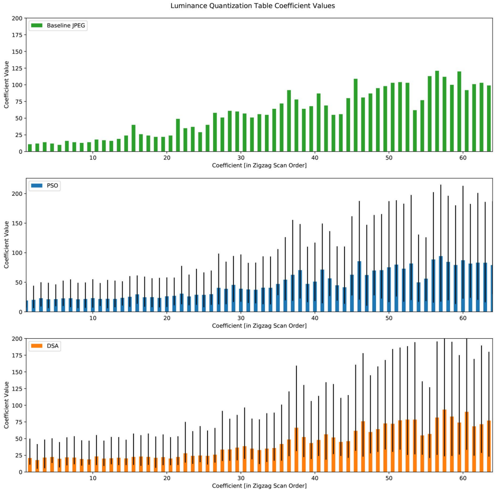

Figure: Average Luminance Quantization Coefficients Values vs Transformed Coefficients in Zigzag Scan Order for the Kodak dataset with quality factor 50.

Coefficient Change - Chrominance

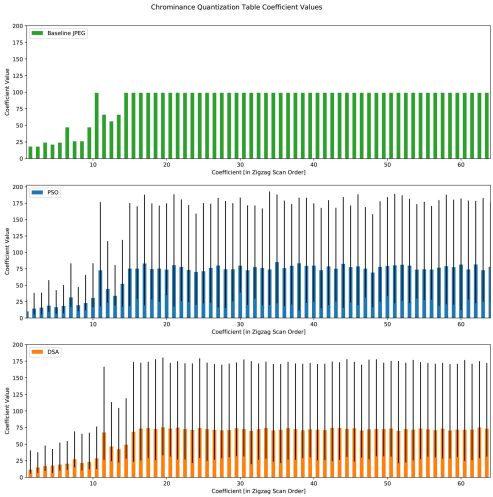

Figure: Average Chrominance Quantization Coefficients Values vs Transformed Coefficients in Zigzag Scan Order for the Kodak dataset with quality factor 50.

Time Performance

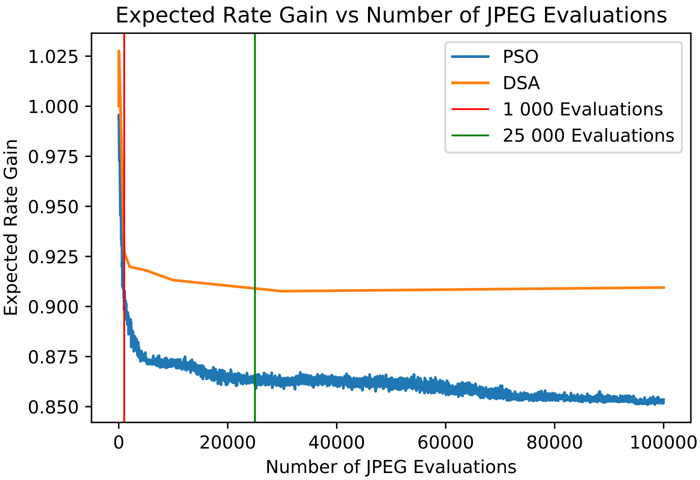

Figure: Average ERG vs Number of JPEG Evaluations

Time Performance

Method	Time	Compression Gain	PSNR Gain
PSO	30	4.6%	0.03dB
PSO	90	8.1%	0.08dB
PSO	300	11.3%	-0.01dB
DSA	30	3.2%	-0.08dB
DSA	90	7.3%	0.13dB
DSA	300	9.2%	0.06dB

Table: Average Performance of a Fixed Time Compression on Kodak Dataset for Quality Factor = 75: Compression and PSNR Gains measured as the average gain across the dataset in comparison with the Baseline JPEG

Table of Contents

Conclusion

- Both DSA and PSO can generate optimized quantization tables that, regardless of the choice of the fitness function, considerably and significantly outperform the vanilla ones from baseline JPEG, being able to generate the same quality images despite producing 8% smaller files.
- The FQ-ERG outperforms the Lagrangian Cost Function in the context of DSA optimization, while no statement can be made for the PSO. Regardless of that, since the FQ-ERG has performed at least no worse than PSO and requires very few parameter tuning, it raises as a suitable alternative for commercial applications built to incorporate by default quantization table optimization in the image compression process.
- This work has also shown that noticeable gain can be achieved even in a modest amount of time and iterations, making this approach flexible and promising even for near real-time or resource-scarce scenarios.

Conclusion and Future Work

- This approach produces a JPEG compliant image, thus preserving backward compatibility and being able to benefit from the existing hardware support.
- The large excursion range in the luminance and chrominance coefficients, combined along with high coefficients of variation, signalizes that we extract a substantial benefit from exploring image-specific information.
- Furthermore, these results work as a showcase of the capabilities of the potential of nature-inspired meta-heuristics for optimization in image compression, making room for its applications as part of the image compression process.
- A short version of this work was submitted as a full paper to the XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT), the leading conference on signal processing in Brazil, and is currently waiting for decision.

The End

- Thank you for your participation!