Automatic JPEG Compression Using a Color Visual Model

Xiao-fan Feng and Scott Daly
Sharp Laboratories of America, Inc
Camas, Washington

Abstract

JPEG compression is extensively used in digital cameras, internet, and image databases. The amount of compression can be adjusted by scaling the quantization table or Q-table. In many cases, an iterative process is used to achieve optimum compression having a smaller file size, but still visually lossless at the intended display and viewing distance. This process is very time consuming for large image databases since human observers are used to judge the image quality.

We present an automatic method to achieve the optimum compression using a color visual difference model (CVDM). The CVDM output is a map of the visible differences between reference and distorted images. In order to use the model in automatic compression, a single number JPEG artifact score was derived from the visual difference map to be used as a merit function. A subjective experiment was conducted to find the best merit function for JPEG artifacts. The subjective experiment also derives an acceptance criterion for JPEG artifacts. We found that the 99-percentile provides the best correlation with the subjective results; thus it was used as the JPEG artifact score in the automatic compression. In the compression process, the compressed images were evaluated with the visual model. Based on the predicted artifact score, the selected Q-table was scaled up or down so that the artifact score was close to the acceptance criteria derived in the subjective experiment.

Introduction

All lossy image compression algorithms usually use quantization to tradeoff between file size and compressed image quality. For example, JPEG compression is extensively used in digital cameras, internet, and image databases. The amount of compression can be adjusted by scaling the quantization table or Q-table. A Q-table with larger values can lead to smaller file size, but more compression artifacts. In many cases, an iterative process involving image viewing and quantization adjustment is used to achieve optimum compression: having a smaller file size, but still visually lossless at the intended display and viewing distance. This process is very time consuming for large image databases since human observers are used to judge the image quality. The process is subjective and in many cases, since the display is not available, may not account for the display difference between the intended display (portable screen) and the observer's display (most likely a computer monitor).

The color visual difference model (CVDM)[1] combines the multi-resolution and masking components of monochromatic visible difference predictor (VDP)[2] and single channel CSF based S-CIELAB.[3] The CVDM not only models the overall spatial frequency response of the visual system, but also masking between patterns of similar orientations and spatial frequencies We also improved the originally published CVDM in several ways.[4]

In this paper, we described an automatic compression method in which the human observer is replaced with a color visual difference model and the human adjuster is replaced with an iterative algorithm that adjusts the parameters of compression to achieve best compression ratio while compression artifact is still below the threshold. The new method can also account for the display by the use of a display model.

Color Visual Difference Model

Key to the auto image compression is the color visual difference model which simulates the visual perception of human eye. It is a detection and appearance visual model that collapses to CIELAB for large patch color. It is calibrated so that the threshold occurs at $\Delta E = 1.0$, regardless of frequency and local image background. Figure 1 shows the process to evaluate a JPEG compressed image. Both the original and compressed images are input to the visual model. Based on the viewing condition and display characteristics, the model calculates the visibility of the differences as a function of location in the image.

In the previous implementations, the masking signal from a sinusoidal grating is also sinusoidal, thus there is no masking near the zero crossing. This causes false detection in certain areas. The VDP[3] addressed this problem by first rectifying the masking signal and then low-pass filtering it by the next lowest DOM filter. However, there is no physiological basis for such action. We now use the more physiologically based method of using even and odd channels[5] where the odd channel cortex filter has a π phase shift in II and III quadrants. Figure 2 shows the convolution kernel of the even and odd cortex filters. The outputs of even and odd filters add together in mean square sense.

The other changes to the model include an OTF in addition to the CSF, a modification of the low frequency attenuation of the CSF to better fit color patch data. Another key difference from the VDP model is that we now use the simpler L* nonlinearity rather than the local cone model, so that the results collapse to CIELAB for solid color patches.[4]

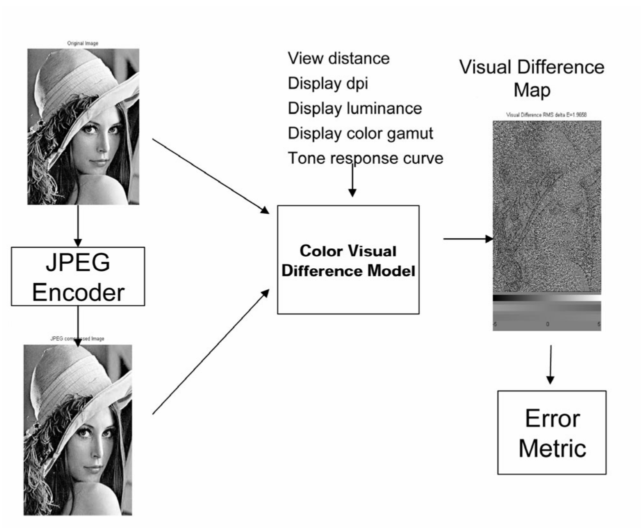

Figure 1. JPEG evaluation using CVDM

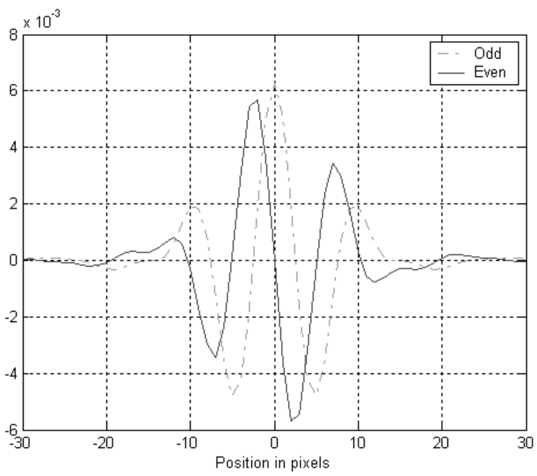

Figure 2. Convolution kernels of even and odd filters

Psychophysical Experiment to Derive Error Metric

We conducted a psychophysical experiment to establish the relationship between visual model output, which is a map, and a singular subjective quality per image. The experiment consists of two observer studies. The first one is a paired comparison experiment with reference. An observer sees three side by side images with the reference (original image) in the middle and two compressed images on the left and right. Observers are asked to pick the better image: left or right. A LCD display is used in the experiment. The display resolution is 90 dpi and visual distance was fixed at 16" (that was dictated by the product application—a Sharp Picture Dictionary model PW-C5000). Four images were used with 4 repetitions and each image is compressed with five different levels resulting a total of 384 comparisons. Five observers participated the test, and the results were pooled to generate a matrix of preference matrix. A psychophysical scale is derived using Thurston's law of comparative study.[6] In the second part of the psychophysical experiment, we try to establish the acceptance criteria for compression artifacts. Two images are display on the screen side by side: one is the original image, and the other is the compressed image. Each observer was asked to judge whether the compressed image is acceptable.

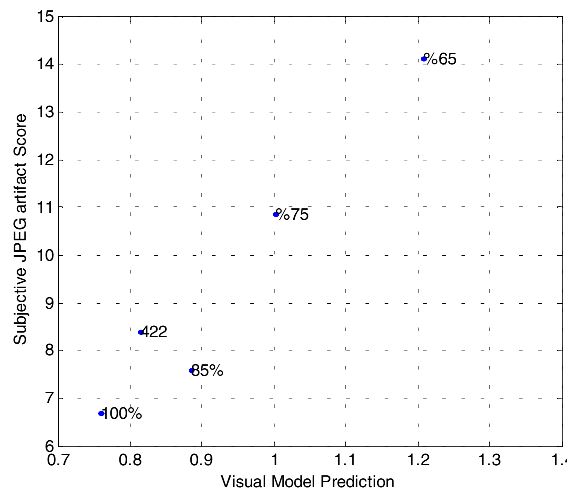

Figure 3. Correlation between subjective scale and 99%

We evaluated a few metrics such as root mean square (RMS), median, 90 percentile, and 99 percentile and found that the 99 percentile provide the best correlation with the subjective experiment, and thus it was used as an error metric for auto JPEG compression. Figure 3 shows the relationship between the 99 percentile from the visual difference map and subjective experiments for 5 compression rates averaged over 4 images. Based on the acceptance experiment, we set the threshold (THD) to 1 ΔE unit, which is approximately the visual detection threshold. This threshold can be adjusted higher for applications that quality is not critical and storage is at premium. The threshold can also be adjusted lower for applications that quality is very critical, or the JPEG images may be viewed at a closer distance.

In this description, we will use JPEG as an example compression technique. The same framework can be directly applied to other lossy compression techniques, e.g., JPEG-2000. Figure 5 shows the flowchart of the auto JPEG compression process. For the compression of many images such as the database applications, this process is repeated for each image. In the JPEG compression process, an image was first compressed with a default compression ratio using several Q-tables. These Q-tables are optimized for different image types (e.g., text images, graphic images, cont-toned images, and etc.). The compressed images were evaluated with the visual model. The Q-table with the lowest predicted artifact score was chosen for that image. The selected Q-table was then scaled so that the artifact score (as predicted by the visual model) was close to the acceptance criteria derived in the subjective experiment.

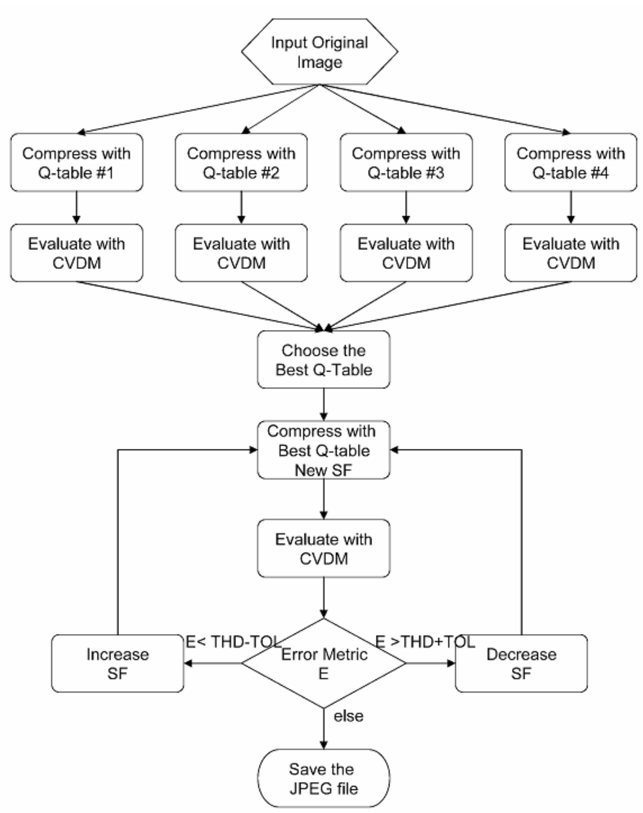

Figure 4. Flowchart of Auto JPEG Compression Process

Q-Table Generation

The quantization table (Q-table) is used to quantize discrete cosine transform (DCT) coefficients. Larger values in a Q-table cause smaller compressed file size, but larger compression artifacts. In most applications, quantization table is fixed. In our optimized compression, four Q-tables were generated using the human visual contrast sensitivity function (CSF) with 4 different viewing distances (11, 14, 17, and 19 inches).[7,8] The display resolution and MTF were also taken into consideration. Closer viewing distances yield flatter tables (since they are in the frequency domain), while further distances yield steeper tables in which the high order DCT coefficients are quantized more aggressively. We found that the best Q-table is not only viewing condition dependent, but also image dependent. This is because of the global level of masking in a given image. If an image has a high level of masking everywhere, then the thresholds are elevated everywhere to a near uniform level[7,9-11] for busy images having uniform distributions of masking, a flat table will be best, though with much higher values than a table for visually lossless results at close distances. Fortunately, the close distance tables can coincidentally be used for very busy images, albeit with a different scale factor.

Q-Table Optimization

For each image to be compressed, all four Q-tables are used to compress the image to the same compression ratio. These compressed images are compared with the original using the visual model. The best Q-table is the one with the lowest artifact score and is chosen to compress the image.

Optimization Process

Once the optimal Q-table is chosen, the image is compressed with a scaled version of this Q-table. A scale factor (SF) is used to scale the Q-table, thus controlling the JPEG compression quality. The compressed image is evaluated using the visual model, and a visual error metric (E) is derived from the visual difference map. The visual error metric is compared to a pre-defined threshold (THD), and tolerance (TOL), which is around 5% of the threshold. If the error metric is greater than the threshold plus tolerance, the compressed image is not acceptable; thus the scale factor is reduced. If, on the other hand, the error metric is less than the threshold minus tolerance, the compression artifact is below the visual threshold at the specified display and viewing condition. We can compress more to further reduce the compressed image size by increasing scale factor. With the new scale factor, a new compressed image is generated. This new compressed image is then evaluated with the visual model to determine the quality, and, if the error metric is not within the threshold and tolerance, the scale factor is modified as before. After a few iterations, the visual error metric converges to the visual threshold within the specified tolerance. With this method, the compressed image is still visually lossless at the smallest possible file size. The compressed image is then saved and the next image in the database is processed.

Conclusion

We have presented a human visual model-based automatic compression algorithm that can select compression quality parameters to match the threshold of the human visual system and the image characteristics. It can save time (human observer) and improve the image quality. Key to the auto compression method is the color visual difference model. We also described improvements to the color visual difference model for more accurate masking prediction.

This method was successfully used to compress an image database of more than 10,000 images.

We notice that 99% metric is more sensitive to ringing than blocking artifact, so we may modify the metric for future usage. One possibility is to have a spatial pooling stage in the detection map before applying the metric.[4]

References

1. Jin, E., X.F. Feng, and J. Newell, The development of color visual difference model (CVDM), in Proc. of IS&T PICS, pp. 154. (1998).
2. Daly, S. The visible differences predictor: an algorithm for the assessment of image fidelity, in Digital Images and Human Vision, A.B. Watson, Editor. MIT Press: Cambridge, MA. pp. 179-206, (1993).
3. Zhang, X. and B.A. Wandell, "A spatial extension of CIELAB for digital color image reproduction", SID Symposium Technical Digest, 27, pp. 731-734 (1996).
4. X. Feng, J. Speigle, and A. Morimoto, "Halftone image quality evaluation using color visual models", Proc. Of PICS 2002, p5-10. (2002)
5. J. Lubin "A visual discrimination model for imaging system design and evaluation", Ch. 10 in Vision Models for target Detection and Recognition, ed E. Peli. World Scientific Press. (1995)
6. Thurstone, L. L. The Measurement of Values. Univ. of Chicago Press, Chicago (1959).
7. S. Daly "Application of a noise-adapative contrast sensitivity function to image data compression", Optical Engineering, V. 29 (8), 977-987, (1990)
8. S. Klein, et al, "Relevance of human vision to JPEG-DCT compression", Proc of SPIE Vol. 1666, p200-215, 1992
9. A.B. Watson "DCTune : A technique for visual optimization of DCT quantization matrices for individual images ", (1993)
10. M.A. Georgeson and G.D. Sullivan, "contrast constancy: deblurring in human vision by spatial frequency channels", J. Physiol., V 252, 627-656. (1975)
11. A. R. Biondini and M.L.F. Maffei "Suprathreshold contrast perception at different luminance levels", Vis. Res. V. 25 pp1-9. (1985)

Biographies

Xiao-fan Feng received the B.S. degree in 1983 from Zhejiang University in Hangzhou, China, ME degree in electro-optics from the Chinese Academy of Science in 1986, MS and Ph.D. degrees in imaging science from Rochester Institute of Technology in 1990 and 1995 respectively. From 1993 to 1997, he worked at Xerox Corporation in Webster NY. Since 1997 he has worked at Sharp Labs of America on image processing algorithms, display quality optimization, and visual modeling.

Scott Daly received a B.S.E.E degree from North Carolina State University, worked at Photo Electronics Corporation in Florida, and then obtained an M.S. in Bioengineering from the University of Utah in 1984. He spent 1985 to 1996 at Eastman Kodak doing image compression, image fidelity, and watermarking. Currently in the Center for Displayed Appearance at Sharp Laboratories of America, he is now applying visual models towards digital video and displays. He is currently a member of IEEE, SPIE, and SID.