

Estudo e Comparação de Técnicas de Compressão de Imagens Baseadas em Transformadas Discretas

Adinan Alves de Brito-Filho
Universidade Federal do ABC
Santo André - São Paulo, Brasil
E-mail: adinan.brito@aluno.ufabc.edu.br

Kenji Nose-Filho
Universidade Federal do ABC
Santo André - São Paulo, Brasil
E-mail: kenji.nose@ufabc.edu.br

Resumo—O objetivo do trabalho foi estudar e comparar métodos, níveis e formas de compressão de imagens baseados no mascaramento e quantização da transformada utilizada pelo padrão JPEG, a Transformada Discreta de Cosseno, do inglês *Discrete Cosine Transform (DCT)*. Para isto, foram realizados experimentos com imagens distribuídas em cinco categorias: paisagem, retrato, tipografia, padrões geométricos e contexto social. A efetividade da compressão e a qualidade da imagem foram avaliadas em termos de métricas estatísticas como Entropia, PSNR e UIQI, de forma a se obter a melhor configuração de processamento para cada categoria de imagem. Durante os testes, constatou-se que o método de compressão DCT via tabelas de quantização obteve um desempenho superior ao mascaramento de coeficientes DCT. Com base nisto, propomos um novo conjunto de tabelas de quantização para a DCT intitulado KDN, que obteve o melhor desempenho geral, superando inclusive as tradicionais tabelas de quantização JPEG standard.

Abstract—This work aims to study and compare methods, levels and forms of image compression based on the masking and quantization of the transform used by the JPEG standard, the Discrete Cosine Transform (DCT). For this, experiments were performed with selected images from five categories: landscape, portrait, typography, geometric patterns and social context. Compression effectiveness and image quality were evaluated in terms of statistical metrics such as Entropy, PSNR and UIQI, in order to obtain the best processing configuration for each image category. During the tests, it was found that the DCT compression method via quantization tables performed better than the masking of DCT coefficients. Based on this, we propose a new set of quantization tables for the DCT called KDN, which obtained the best overall performance, even outperforming the traditional JPEG standard quantization tables.

I. INTRODUÇÃO

A Transformada Discreta de Cosseno (*Discrete Cosine Transform, DCT*), permite a transformação de sinais e imagens para o domínio das frequências, possibilitando a representação da informação contida nos sinais e imagens através de um número limitado de coeficientes [1]. Logo, dado que a DCT concentra a informação em poucos coeficientes, infere-se que seja possível desprezar parte dos coeficientes sem efeito perceptível na qualidade do sinal reconstruído [2].

Devido a sua alta capacidade de compactação [3], a DCT tem sido amplamente utilizada como um dos algoritmos padrões para compactação de imagens e vídeos, em resposta à urgência tecnológica por rapidez e qualidade na transmissão da informação. A transformada é utilizada tanto em formatos

tradicionais, como o JPEG e o MPEG [4], como nos padrões mais recentes de codificação, como o HEVC [5] e o HEIF [6].

Se valendo dos atributos e da versatilidade da DCT, estudamos quais as melhores maneiras de se obter uma ótima compressão de imagens por meio da transformada, objetivando a comprovação experimental de que certas classes de imagens podem ser mais comprimidas do que outras, sem que haja a deterioração da informação ali presente. Para esse estudo foram considerados os métodos mais tradicionais de compressão de imagens via DCT atualmente: o mascaramento de coeficientes e o arredondamento por tabelas de quantização. Os métodos foram aplicados sob diversas configurações de compressões em um *dataset* contendo 60 imagens, classificadas em 5 grupos: paisagem, retrato, padrões geométricos, tipografia e contexto social.

Satisfatoriamente, a hipótese inicial foi comprovada, evidenciando que imagens de paisagens e retratos podem ser mais comprimidas do que imagens tipográficas e de padrões geométricos. Além disso, criamos novas tabelas de quantização para arredondamento dos coeficientes DCT que obtiveram ótimo desempenho durante os experimentos, superando em taxa de compressão, para um mesmo nível de qualidade, as principais tabelas de quantização atualmente, as tabelas JPEG [7].

Nas próximas seções são apresentados a fundamentação teórica, a metodologia experimental para os comparativos, os resultados e as considerações finais a respeito do estudo.

II. FUNDAMENTAÇÃO TEÓRICA

Uma imagem pode ser definida matematicamente como uma função bidimensional, $f(x, y)$, em que x e y são coordenadas espaciais de um plano, e a amplitude de f em qualquer par de coordenadas (x, y) é chamada de intensidade [8]. Quando x , y e os valores de intensidade de f são quantidades finitas e discretas possuímos então uma imagem digital, onde cada par ordenado representa um pixel e sua amplitude representa a intensidade de cor, variando a cor a ser representada conforme o sistema de cores utilizado na imagem.

A. O padrão JPEG

O padrão JPEG é o principal padrão de codificação de imagens digitais atualmente [9]. O JPEG define vários processos de compressão para imagens estáticas. Dentre esses,

seu método padrão e mais utilizado se trata de um método de compressão com perdas baseado na DCT.

Neste trabalho, demos enfoque aos processos de codificação de uma imagem JPEG referentes à compressão com perdas, efetuada por meio de operações matemáticas na Transformada Discreta de Cosseno da imagem.

B. A Transformada Discreta de Cosseno

As Transformadas são operadores matemáticos utilizados com o objetivo de facilitar a busca de soluções para problemas não triviais. A Transformada Discreta de Cosseno expressa uma sequência finita de dados em termos de uma soma de funções de cosseno que oscilam em diferentes frequências [10].

Seja um sinal $F(x, y)$, com $x = 0, 1, \dots, 7$ e $y = 0, 1, \dots, 7$, a transformada discreta do cosseno $C(u, v)$ para duas dimensões, com $u = 0, 1, \dots, 7$ e $v = 0, 1, \dots, 7$, é dada por:

$$C(u, v) = g(u)g(v) \sum_{x=0}^7 \sum_{y=0}^7 F(x, y) \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right). \quad (1)$$

Sendo que:

$$g(k) = \begin{cases} \frac{1}{2} \frac{1}{\sqrt{2}} & \text{se } k = 0, \\ \frac{1}{2} & \text{se } k \neq 0. \end{cases} \quad (2)$$

Essencialmente, a DCT leva a informação do domínio espacial (amplitudes) para o domínio das frequências. Através dela, a informação presente em cada subimagem, resultante do particionamento da imagem original em blocos de 8 por 8 píxeis não sobrepostos, passa a ser representada pelo coeficiente de cada um dos 64 blocos de interpolações de cossenos, também chamados de funções-base ou primitivas da DCT [11]. A posição do bloco é determinada pelas frequências que geram o seu conteúdo espectral, ou seja, quanto maior o índice do bloco, mais alta a frequência.

C. Métodos de compressão na DCT

A compressão com perdas na DCT se caracteriza como qualquer alteração nos coeficientes que leve à perda de informação. Abaixo estão listados os principais métodos de compressão de imagens por meio da DCT:

1) *Mascaramento*: Consiste no descarte dos coeficientes de alta frequência da DCT, onde a taxa de compressão se dá pela quantidade de elementos da matriz DCT de cada subimagem a serem zerados.

2) *Tabelas de Quantização*: Os coeficientes DCT são mapeados para um número finito e reduzido de níveis, gerando um efeito comparável ao de um arredondamento, em termos gerais. Neste procedimento, os coeficientes $C(u, v)$ são divididos, elemento a elemento, por seu correspondente na tabela de quantização $Q(u, v)$.

O produto da divisão elemento a elemento é então arredondado, gerando a matriz de coeficientes DCT quantizados. A quantização, de um modo geral, é realizada da seguinte forma:

$$D(u, v) = \text{round}\left(\frac{C(u, v)}{kQ(u, v)}\right), \quad (3)$$

onde k é o chamado fator de compressão.

D. Métricas de avaliação

Neste estudo foram utilizadas três métricas para a avaliação dos resultados: a Relação sinal-ruído de pico (*Peak signal-to-noise ratio, PSNR*) [12], para avaliação do nível de degradação da imagem comprimida - cujo unidade de medida é o decibel (dB); o Índice Universal de Qualidade de Imagem (*Universal Image Quality Index, UIQI*) [13], para quantificação do grau de similaridade entre a imagem original e a imagem comprimida; a Entropia, para estimação da taxa de compressão.

Para a PSNR, o valor do índice pode variar entre 0 e 100. Quando não houve compressão da imagem, o valor obtido é 100, e quando a imagem de saída apresenta apenas ruído, o valor é 0. Já para o UIQI, o valor do índice varia entre 0 e 1, sendo 1 o melhor valor e, portanto, qualidade indistinguível entre as imagens comparadas.

Para uma imagem, a Entropia representa a quantidade de aleatoriedade (ou incerteza) presente no sinal [14]. A Entropia de um canal de cor de uma imagem, calculada através de sua DCT, é dada por:

$$H(x) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i), \quad (4)$$

sendo x a variável referente ao coeficiente DCT quantizado, N a quantidade de valores possíveis para o coeficiente DCT quantizado e $p(x)$ a probabilidade de ocorrência de um valor de coeficiente DCT.

Calculando a Entropia de uma imagem original e de sua versão comprimida, é possível obter uma estimativa da taxa de compressão efetuada sobre a imagem comprimida. A taxa de compressão de uma imagem T é calculada da seguinte forma:

$$T = 1 - \frac{\frac{1}{N_c} \sum_{i=1}^{N_c} H_i(y)}{\frac{1}{N_c} \sum_{i=1}^{N_c} H_i(x)}, \quad (5)$$

onde x e y representam os coeficientes das matrizes DCT tridimensionais da imagem original e comprimida, respectivamente; N_c representa o número de canais de cor das imagens; i representa o canal de cor; H_i representa a Entropia da matriz DCT de um canal de cor da imagem.

III. METODOLOGIA

Para a realização dos experimentos entre os métodos de compressão na DCT, fora construída uma aplicação utilizando a linguagem de programação *Python*. Definiu-se o recorte de estudo dos efeitos da DCT e seus métodos de compressão em cinco classes de imagens, sendo elas: paisagem, retrato, tipografia, padrões geométricos e contexto social. Tais classes

foram definidas sobretudo por sua relevância no âmbito fotográfico. Com base nisso, selecionamos 60 fotografias com permissão para edição e uso não comercial da plataforma de imagens Flickr.

Confeccionado o *dataset*, deu-se início à programação do projeto. Primeiramente, foi necessário redimensionar todas as imagens presentes no *dataset*, para que ambas as dimensões horizontais e verticais fossem múltiplas de 8, visando a facilitação dos processos seguintes. Para isto, definiu-se a resolução de 600x400, sem distorção da proporção original da imagem.

Após o redimensionamento, a imagem foi convertida do espaço de cores RGB para YCbCr [15], visando a facilitação dos processos de compressão a seguir. Logo após, as matrizes de cada canal (Y, Cb e Cr) foram particionadas em blocos de 8x8 píxeis, resultando em três matrizes de quatro dimensões, de ordem 50x75x8x8.

Com as matrizes de canais de cores particionadas e prontas para o processamento, realizamos a transformação da imagem para o domínio das frequências, segundo a DCT, e calculamos a entropia dos coeficientes DCT resultantes da transformação sem compressão. Após a aplicação da transformada, foi definido o método de compressão DCT utilizado, sendo ele o mascaramento ou a quantização via tabelas de compressão.

A. Métodos de mascaramento estudados

Neste projeto, foram utilizados dois métodos de mascaramento:

- Mascaramento Quadrado: Os coeficientes desprezados formam um quadrado na matriz DCT. Quanto maior o tamanho do quadrado de zeros, maior a compressão.
- Mascaramento Triangular: Os coeficientes truncados se posicionam formando uma matriz triangular inferior, no mesmo sentido da diagonal secundária da matriz DCT. Quanto maior a ordem da matriz triangular, maior a compressão.

A Figura 1 ilustra os mascaramentos utilizados nos testes deste projeto.

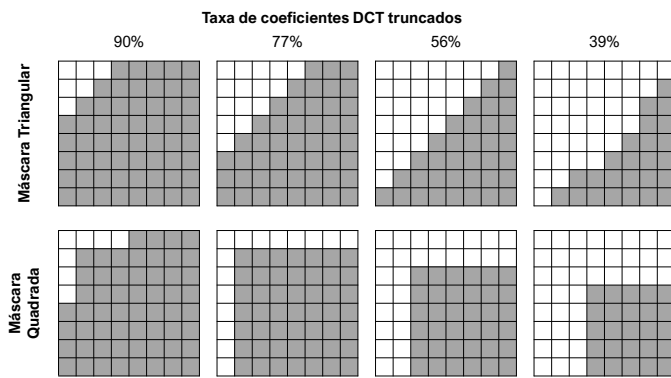


Fig. 1. Tipos de mascaramento e taxas de truncamento utilizadas nos testes. Os elementos de cor branca simbolizam os coeficientes DCT preservados. Os elementos de cor cinza representam os coeficientes DCT desprezados.

B. Tabelas de Quantização estudadas

As tabelas de quantização utilizadas nos experimentos foram as tabelas JPEG *standard* de luminância e crominância [9], bem como as tabelas criadas pelos autores, para luminância e crominância, denominadas como KDN (referência aos primeiros nomes dos autores, Kenji e Adinan).

As tabelas KDN foram formuladas a partir da consideração de que os coeficientes DCT de baixas frequências constituem a maior parte da informação presente em uma imagem. Logo, as altas frequências são menos importantes para a representação dessa, e a diminuição de suas magnitudes não alteraria significativamente a imagem. Além disso, esse processo geraria economia de espaço para armazenamento da informação. Com base nisso, formulamos dois algoritmos para a confecção das matrizes de luminância e crominância, que levam em consideração o índice do coeficiente na matriz e um aspecto psicofísico do sistema visual humano, abordado a seguir. Os algoritmos podem ser visualizados abaixo.

Algoritmo 1 Matriz proposta para luminância

```

1: for ii = 1 : 8 do
2:   for jj = 1 : 8 do
3:     Q(ii,jj) = (ii + jj)2
4:     if Q(ii,jj) > 99 then
5:       Q(ii,jj) = 99
6:     end if
7:   end for
8: end for

```

Algoritmo 2 Matriz proposta para crominância

```

1: for ii = 1 : 8 do
2:   for jj = 1 : 8 do
3:     Q(ii,jj) = (ii + jj)2.5
4:     if Q(ii,jj) > 99 then
5:       Q(ii,jj) = 99
6:     end if
7:   end for
8: end for

```

Dado que o sistema visual humano é mais sensível a pequenas variações de luminosidade do que de cromaticidade [16], o algoritmo 2 foi elaborado de forma que os valores na matriz de quantização de crominância fossem maiores do que os valores na matriz de quantização de luminância, por meio do expoente na operação de potência da linha 3, que é maior no algoritmo 2. Dessa forma, propicia-se uma maior perda de informação nas matrizes de coeficientes DCT de crominância, pouco significativa ao sistema visual humano.

Para o fator de compressão k empregado nas tabelas de quantização, utilizamos valores obtidos por meio da equação 6, exibida abaixo.

$$k(n) = \text{round}(0.05 \cdot 1.459^{n-1}) \quad (6)$$

A equação 6 foi formulada experimentalmente, com base nos dez níveis de compressão de imagens JPEG do programa *Adobe Photoshop*, de forma que os fatores de compressão $k(n)$ fornecessem taxas de compressão compatíveis com os níveis de compressão utilizados na indústria. Foram utilizados dez níveis de compressão $n = \{0, 1, \dots, 9\}$, de forma que obtivemos os seguintes fatores de compressão $k = \{0.05, 0.07, 0.11, 0.16, 0.23, 0.33, 0.48, 0.70, 1.03, 1.50\}$.

Realizada a compressão, calculamos a entropia dos dados comprimidos a partir da matriz de coeficientes resultante. Em seguida, aplicamos a inversa da DCT para obtê-los no domínio

da intensidade de luz. Realizamos a transformação das matrizes de quatro dimensões para duas dimensões, desfazendo o particionamento, e convertimos novamente o espaço de cores dos dados, dessa vez para o RGB, obtendo assim a imagem comprimida resultante. Após, salvamos a imagem resultante no formato PNG para evitar novas compressões com perdas. Em seguida, calculamos as métricas UIQI (normalizada de 0 a 100), PSNR e a taxa de compressão para cada imagem.

IV. RESULTADOS

Analisando os dados obtidos, constatou-se que a compressão via tabelas de quantização obteve desempenho muito superior ao mascaramento para ambas as métricas, sobretudo na PSNR.

O desempenho aquém do mascaramento na PSNR pode ser explicado pela própria informação que a métrica avalia: o nível de ruído, que é acentuado no mascaramento devido ao descarte das altas frequências. Tal fenômeno pode ser visualizado na Figura 2. A perda se torna ainda mais acentuada no mascaramento quadrado, que além de desprezar as altas frequências, também acaba desprezando boa parte das médias e baixas frequências. Logo, o mascaramento não é recomendado para nenhuma classe de imagens, sobretudo, classes com maior ocorrência de altas frequências, como tipografias e padrões geométricos.

TABELA I
RESULTADOS DAS TABELAS DE QUANTIZAÇÃO
JPEG E KDN PARA FATORES k NOTÁVEIS

Fator k	Métrica	Tabelas de Quantização	Classes				
			Padrões Geométricos	Paisagem	Retrato	Contexto Social	Tipografia
0,05	$PSNR$ (dB)	KDN	44,16	43,77	44,50	44,02	45,50
		JPEG	44,82	44,31	45,07	44,58	46,14
	$UIQI$	KDN	99,12	98,70	99,05	98,88	99,98
		JPEG	99,18	98,79	99,13	99,01	99,99
	Taxa de Compressão	KDN	28,6%	36,1%	37,8%	35,3%	35,1%
		JPEG	23,5%	30,6%	32,0%	29,8%	29,6%
0,11	$PSNR$ (dB)	KDN	39,00	39,49	40,43	39,77	41,28
		JPEG	39,82	40,21	41,25	40,56	42,14
	$UIQI$	KDN	98,59	98,24	98,77	98,39	99,96
		JPEG	98,70	98,34	98,86	98,56	99,97
	Taxa de Compressão	KDN	46,0%	54,5%	56,0%	53,2%	52,8%
		JPEG	41,2%	49,7%	51,1%	48,4%	47,9%
0,33	$PSNR$ (dB)	KDN	32,49	34,02	35,33	34,28	35,40
		JPEG	33,57	34,94	36,42	35,36	36,57
	$UIQI$	KDN	97,56	97,49	98,21	97,50	99,90
		JPEG	97,76	97,57	98,32	97,66	99,92
	Taxa de Compressão	KDN	65,3%	72,7%	73,8%	71,4%	70,4%
		JPEG	61,3%	69,5%	70,6%	68,0%	67,1%
1,5	$PSNR$ (dB)	KDN	25,30	28,40	30,35	28,66	28,26
		JPEG	26,72	29,39	31,26	29,72	29,77
	$UIQI$	KDN	95,64	96,25	97,06	95,75	99,66
		JPEG	96,12	96,15	96,62	95,73	99,72
	Taxa de Compressão	KDN	82,4%	86,9%	87,3%	85,9%	85,0%
		JPEG	80,0%	85,8%	86,6%	84,7%	83,5%

Os resultados em negrito denotam as situações em que as tabelas KDN obtiveram melhor desempenho tanto em compressão quanto em UIQI.

Observando os resultados contidos na Tabela I, nota-se uma diferença significativa na compressão efetuada pelas tabelas KDN em relação às tabelas JPEG, tornando a vantagem das

tabelas JPEG em PSNR e UIQI irrelevante. Portanto, é mais vantajoso codificar uma imagem utilizando as tabelas KDN, pois a diferença no tamanho do arquivo será mais significativa do que a diferença na qualidade da imagem.

Fig. 2. Imagem da categoria Contexto Social comprimida utilizando Mascaramento e Tabelas de Quantização. O baixo desempenho do mascaramento é notório.



(a) Mascaramento quadrado com 90% de coeficientes desprezados. Taxa de compressão igual a 73,38%. UIQI = 95,08. PSNR = 21,58 dB.

(b) Tabelas KDN, com $k = 1,5$. Taxa de compressão igual a 85,10%. UIQI = 96,48. PSNR = 25,10 dB.

V. CONCLUSÃO

A hipótese inicial de que existem certos padrões na codificação JPEG que podem ser manipulados de forma a favorecer a compressão de certas classes de imagens foi comprovada. A categoria retrato demonstrou suportar uma maior compressão, ao contrário da categoria padrões geométricos, onde houve demasiada perda de informação. Logo, é possível se valer dessa estratégia para otimizar o armazenamento e a transmissão da informação.

As tabelas KDN demonstraram eficácia como alternativa às tabelas de quantização JPEG, promovendo uma maior taxa de compressão e entregando, em certas condições, uma maior qualidade e maior taxa de compressão. Assim, evidenciou-se sua relevância, bem como a importância de uma codificação inteligente e assertiva, com base na classe da imagem.

REFERÊNCIAS

- [1] A. Bovik, *The Essential Guide to Image Processing*. Elsevier Science, 2009. [Online]. Available: <https://books.google.com.br/books?id=6TOUgytafmQC>
- [2] R. Gonzalez and R. Woods, *Digital Image Processing*. Prentice Hall, 2002. [Online]. Available: <https://books.google.com.br/books?id=738oAQAAIAAJ>
- [3] W.-D. Liang and X.-D. Liu, “Comparison of approximate dct and approximate dtt for image compression,” in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2021, pp. 337–341.
- [4] A. C. Bovik, *Handbook of Image and Video Processing*, ser. Communications, Networking and Multimedia. Elsevier Science, 2010. [Online]. Available: https://books.google.com.br/books?id=UM_GCfJe88sC
- [5] G. J. Sullivan and J.-R. Ohm, “Recent developments in standardization of high efficiency video coding (HEVC),” in *Applications of Digital Image Processing XXXIII*, A. G. Tescher, Ed., vol. 7798, International Society for Optics and Photonics. SPIE, 2010, pp. 239 – 245. [Online]. Available: <https://doi.org/10.1117/12.863486>
- [6] J. Lainema, M. M. Hannuksela, V. K. M. Vadekital, and E. B. Aksu, “Hevc still image coding and high efficiency image file format,” in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 71–75.

- [7] C.-Y. Pang, R.-G. Zhou, B.-Q. Hu, W. Hu, and A. El-Rafei, "Signal and image compression using quantum discrete cosine transform," *Information Sciences*, vol. 473, pp. 121–141, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020025516310970>
- [8] J. C. Russ, *The Image Processing Handbook*. CRC Press, 6ed., 2011.
- [9] International Telecommunication Union, *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines*, 1994.
- [10] A. J. I. Barbhuiya, T. A. Laskar, and K. Hemachandran, "An approach for color image compression of jpeg and png images using dct and dwf," in *2014 International Conference on Computational Intelligence and Communication Networks*, 2014, pp. 129–133.
- [11] G. Stolfi, *COMPRESSÃO DE IMAGENS: PADRÃO JPEG*, Notas de aula da disciplina PTC2547 – PRINCÍPIOS DE TELEVISÃO DIGITAL, da EPUSP, 2016. [Online]. Available: <http://www.lcs.poli.usp.br/~gstolfi/PPT/APTV0616.pdf>
- [12] J. Hwang, *Multimedia Networking: From Theory to Practice*. Cambridge, 2009.
- [13] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [14] D.-Y. Tsai, Y. Lee, and E. Matsuyama, "Information-entropy measure for evaluation of image quality," *Journal of Digital Imaging*, vol. 21, 09 2008.
- [15] E. Hamilton, *JPEG File Interchange Format*, 09 1992. [Online]. Available: <https://www.w3.org/Graphics/JPEG/jfif3.pdf>
- [16] F. Zhang and D. Bull, *Intelligent Image and Video Compression: Communicating Pictures*. Elsevier Science, 2021. [Online]. Available: <https://books.google.com.br/books?id=KiMZEAAAQBAJ>