



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
FACULDADE DE ENGENHARIA DA COMPUTAÇÃO E
TELECOMUNICAÇÕES *se der.*

**Quantização inter adaptativa baseada no sistema visual humano:
análise e aplicações no MPEG-1.**

Luan Assis Gonçalves

BELÉM - PARÁ

2016



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
FACULDADE DE ENGENHARIA DA COMPUTAÇÃO E
TELECOMUNICAÇÕES

re der

Luan Assis Gonçalves

**Quantização inter adaptativa baseada no sistema visual humano:
análise e aplicações no MPEG-1.**

Trabalho de Conclusão de Curso apresentado para
obtenção do grau de Engenheiro em Engenha-
ria da Computação, do Instituto de Tecnologia,
da Faculdade de Engenharia da Computação e
Telecomunicações.

BELÉM - PARÁ

2016

**Quantização inter adaptativa baseada no sistema visual humano:
análise e aplicações no MPEG-1.**

Este trabalho foi julgado adequado em xx/xx/2016 para a obtenção do Grau de Engenheiro da Computação, aprovado em sua forma final pela banca examinadora que atribui o conceito _____.

Prof. Dr. Ronaldo de Freitas Zampolo
ORIENTADOR

Prof. Dr. Evaldo Gonçalves Pelaes
MEMBRO DA BANCA EXAMINADORA

Prof^a. Dr^a. Valquíria Gusmão Macedo
MEMBRO DA BANCA EXAMINADORA

Prof. Dr. Francisco Carlos Bentes Frey Müller
DIRETOR DA FACULDADE DE ENGENHARIA
DA COMPUTAÇÃO E TELECOMUNICAÇÕES

Aos anjos em minha vida
Helena e Antônio.

Agradecimentos

*“Qualquer um que nunca tenha cometido um erro nunca
tentou algo novo.”
Albert Einstein*

Resumo



Palavras-chave: Compressão, quantização, codificação perceptual, qualidade visual, MPEG-1.

Abstract



Keywords: Compression, quantization, perceptual coding, visual quality, MPEG-1.

Lista de Figuras

2.1	Reduções da fonte [1].	19
2.2	Codificação da fonte reduzida [1].	19
2.3	Fluxograma do codificador JPEG.	22
2.4	Fluxograma do decodificador JPEG.	22
2.5	Macrobloco dizimado no padrão 4:2:0.	25
2.6	Compensação de movimento pra frente.	25
2.7	Compensação de movimento bidirecional.	26
2.8	Fluxograma do codificador MPEG.	29
2.9	Sequência de imagens dentro de um GOP.	30
2.10	Fluxograma do decodificador MPEG.	30
3.1	Distribuição da frequência espacial no domínio da DCT.	36
3.2	Visão perspectiva da superfície de limiar espacial e temporal [2]. Cada curva representa a resposta a frequência espacial em uma frequência espacial fixa. . .	37
4.1	Comparativo dos valores de MSSIM e PSNR pelos codecs padrão e perceptual.	41
4.2	Comparativo entre os codecs padrão, (a) e (b), e perceptual, (c) e (d).	42
4.3	Comparativo dos valores de MSSIM e PSNR pelos codecs padrão, padrão adaptado e perceptual.	43

Lista de Tabelas

2.1	Exemplo de codificação aritmética.	20
2.2	Tabela de quantização padrão (50%).	24
2.3	Padrão zigzag.	24
2.4	Exemplo de subimagem.	24
2.5	Subimagem da tabela 2.4 quantizada.	24
2.6	Tabela intra.	27
2.7	Tabela inter.	27
4.1	Descrição dos vídeos.	39
A.1	Categorias de coeficientes de codificação JPEG.	47
A.2	Códigos JPEG DC.	47
A.3	Códigos JPEG AC.	48
A.4	Códigos JPEG AC (continuação.)	49
A.5	Códigos JPEG AC (continuação).	50
B.1	Códigos de tamanho variável para os vetores de desloaento.	52

Lista de Abreviaturas e Siglas

CD Compact Disc

DCT Discrete Cosine Transform

DPCM Differential Pulse-Code Modulation

DVD Digital Versatile Disc

HDTV High-Definition Televisio,

HVS Human Visual System

IDCT Inverse Discrete Cosine Transform

ITU International Telecommunication Union

JPEG Joint Photographic Experts Group

MSE Mean Squared Error

MSSIM Mean Structural Similarity Index

MOS Mean Opinion Score

MPEG Moving Picture Experts Group

PSNR Peak Signal-to-Noise-Ratio

SNR Signal-to-Noise Ratio

SSIM Structural Similarity Index

TV Television

Sumário

1	Introdução	14
1.1	Motivação	14
1.2	Visão geral do trabalho	14
1.3	Organização do trabalho	15
2	Fundamentos da compressão de imagens e vídeos	16
2.1	Introdução	16
2.2	Redundância	16
2.3	Teoria da Informação	17
2.3.1	Primeiro teorema de Shannon	18
2.3.2	Entropia	18
2.4	Alguns métodos básicos de compressão	18
2.4.1	Métodos	19
2.5	Os padrões JPEG e MPEG	21
2.5.1	JPEG	22
2.5.2	MPEG-1	23
3	Qualidade visual	31
3.1	Introdução	31
3.2	Artefatos provenientes do processo de compressão	32
3.3	Avaliações subjetiva e objetiva	32
3.4	Classificação das métricas de avaliação de qualidade visual	33
3.4.1	Não perceptuais	34
3.4.2	Perceptuais	35
3.5	Quantização inter adaptativa baseada no HVS	36

4	Procedimento experimental e resultados	39
4.1	Implementação do sistema	39
4.2	Conjunto de teste	39
4.3	Procedimento experimental	40
4.4	Resultados	40
5	Considerações finais	44
5.1	Conclusões	44
5.2	Trabalhos futuros	45
A		46
A.1	Tabelas de codificação para compressão JPEG	46
B		51
B.1	Tabelas de codificação para compressão MPEG-1	51
	Referências Bibliográficas	53

Capítulo 1

Introdução



Motivação

Com a expansão de tecnologias como a TV digital de alta definição (também conhecida como HDTV), streaming de vídeo e Blu-Ray, a compressão de vídeo tornou-se elemento indispensável para as transmissões broadcast e mídias de entretenimento. Neste contexto, pesquisas relacionadas à área de compressão de informações visuais foram potencializadas a fim de otimizar a utilização da banda de transmissão, resultando no surgimento de novos padrões de compressão de imagens e vídeos.

Embora o poder de processamento computacional, armazenamento e a largura de banda disponível para transmissão de dados tenham evoluído, ainda assim, a necessidade de métodos que possam prover taxas consideráveis de compressão e baixas taxas de distorção é um desafio. Esta realidade é evidenciada em trabalhos como [3], [4], [5] e [6]. *simônimo*

A situação apresentada anteriormente **potencializa-se** com a expansão dos dispositivos móveis. Em [7], é feita uma projeção de que nos próximos 2 anos o planeta terá mais dispositivos móveis conectados à internet do que habitantes, gerando um fluxo de cerca de 10 imagens e 1 vídeo por habitante ao dia.

Visão geral do trabalho

Na literatura há trabalhos que abordam de maneira profunda as bases teóricas da compressão de imagens e vídeos, enquanto outros ^{tem} como foco principal os aspectos práticos e de implementação. Na fase de graduação observa-se que os discentes muitas vezes são expostos a

conceitos teóricos de teoria de informação e processamento de sinais e poucas vezes tem contato com suas aplicações. Por isso, um dos objetivos deste trabalho é abordar teoria e prática, fundamentando os conceitos de compreensão de imagens e vídeos através de implementações simplificadas dos padrões *JPEG* (~~do inglês~~ Joint Photographics Experts Group) e *MPEG-1* (~~do inglês~~ Moving Picture Experts Group).

Outro aspecto ~~que não interessa~~ ^{de interesse} está relacionado com a qualidade perceptual dos vídeos recuperados através do processo de decodificação. Indiscutivelmente, a melhor forma de avaliar a qualidade de perceptual de informações visuais é através da consulta ao próprio usuário final (avaliação subjetiva), no entanto este processo custa muito tempo para que se possa obter uma avaliação com valor estatístico. A fim de contornar este problema, trabalhos como [8], [9] e [10] apresentam propostas para substituir a avaliação subjetiva pela objetiva, que consiste na utilização de funções matemáticas que simulam a percepção visual humana.

Portanto, este trabalho objetiva abordar a teoria e a prática do processos de compressão e melhoria de qualidade visual, comparando o codificador padrão MPEG-1 com o mesmo acrescido do modelo de quantização perceptual inter adaptativa baseada no sistema visual humano, apresentado em [11], através da avaliação objetiva de qualidade perceptual.

Organização do trabalho

O restante deste documento está organizado da seguinte forma:

- Capítulo 2: são apresentados conceitos básicos relacionados à compressão de imagens e vídeos, com ênfase nos padrões *JPEG* e *MPEG-1*;
- Capítulo 3: neste capítulo, o foco é a avaliação da qualidade visual de imagens e vídeos depois do processo de descompressão bem como ~~avaliar~~ ^{descrever} um método que visa a melhoria da qualidade perceptual, através da geração dinâmica das tabelas de quantização baseadas no sistema visual humano, proposto em [11].
- Capítulo 4: o procedimento experimental é definido e, por fim, os resultados obtidos são apresentados e discutidos;
- Capítulo 5: a partir da análise detalhada dos resultados, serão apresentadas as conclusões da técnica analisada, destacando ~~suas~~ ^{observadas} vantagens e desvantagens.

Capítulo 2

Fundamentos da compressão de imagens e vídeos

Introdução

Neste capítulo abordam-se as aspectos teóricas do processo de compressão de imagens e vídeos. Primeiramente, o conceito de redundância, presente em arrays 2D, será abordado, fazendo-se a relação com a teoria da informação. Em seguida, é discutida uma possível classificação dos tipos de compressão com base na preservação do sinal original. Por fim, analisa-se o funcionamento dos padrões JPEG e MPEG-1.

Redundância

O processo de compressão de dados consiste em reduzir a quantidade de bits necessária para representar uma dada informação. Neste contexto, os conceitos de dados e informação são diferentes, em que os dados são os meios pelos quais as informações são transmitidas [1].

Seguindo esta linha de raciocínio, uma informação pode ser representada de infinitas maneiras. Desta forma, o questionamento a ser respondido quando se objetiva a compressão de dados é: qual representação forneceria o menor volume de dados sem que haja perda de informação?

A compressão é obtida através da eliminação dos dados redundantes presentes na representação de uma informação. Em se tratando de arrays bidimensionais, os principais tipos de redundância são:



- *Redundância de codificação*: ocorre quando a quantidade de bits utilizada para representar os símbolos de uma determinada informação é superior à quantidade necessária. Este tipo de de redundância é muito comum em que trabalham com a atribuição de palavras código de tamanho fixo.
- *Redundância espacial e temporal*: devido à grande parte dos pixels presentes em um array 2D estarem espacialmente correlacionados, surge a redundância espacial. Já as sequências de vídeo estão sujeitas a mais outro tipo de redundância, a temporal, em que os pixels de quadros vizinhos encontram-se correlacionados, devido à grande semelhança entre eles. Isso significa dizer que em compressão de vídeos é possível alcançar maiores taxas de compressão ~~do~~ que em imagens estáticas.
- *Redundância psicovisual*: é originada a partir das características do sistema visual humano. Sua resposta aos estímulos visuais é uma função não linear de grandezas físicas, como intensidade luminosa e cores. Neste contexto, pesquisas a respeito do funcionamento e comportamento do sistema visual humano são de grande importância para que possam ser gerados modelos matemáticos que ~~descrevam a sensibilidade~~ *representem características* do sistema visual humano.

A quantificação do volume de dados redundantes presente em uma representação de imagem é necessária ~~x~~ para que se possa avaliar a compressão obtida. Sendo assim, assumindo que b e b' são, respectivamente, o volume de dados presentes na representação real de uma imagem e o volume de dados presentes em uma representação comprimida da mesma, a *redundância relativa* R é dada por *[ref]*

$$R = 1 - \frac{1}{c} \quad (2.1)$$

em que c é a *taxa de compressão*, definida como sendo

$$c = \frac{b}{b'} \quad (2.2)$$

Teoria da Informação

Durante a década de 40, no período da Segunda Guerra Mundial, o processo de troca de informações tornou-se fundamental. Dessa forma, surgiu a necessidade de estabelecer um limite mínimo de volume de bits necessário para a transmissão de uma determinada informação, a fim de otimizar a utilização do canal disponível.

Neste contexto, Claude Elwood Shannon ficou conhecido com o “pai da teoria da informação” ao propor com sucesso uma maneira de medir a incerteza sobre espaços desordenados.

Primeiro teorema de Shannon

Através do teorema da codificação sem perda [12] pode-se provar que é possível representar a saída de um sistema sem memória com uma média H de unidades de informação por pixel,

$$\lim_{n \rightarrow \infty} \left[\frac{L_{avg,n}}{n} \right] = H \quad (2.3)$$

em que, $L_{avg,n}$ é o tamanho médio dos códigos necessários para representar todos os grupos de n símbolos e H é a entropia. *→ conferir*

Entropia

A entropia, equação 2.5, é uma medida utilizada em diversas áreas do conhecimento, como na química e física. Em se tratando de informações, a entropia representa o grau de incerteza de uma fonte.

Sendo (a_1, a_2, \dots, a_J) o conjunto de símbolos (eventos) emitidos por uma determinada fonte e $P(a_1), P(a_2), \dots, P(a_J)$ suas respectivas probabilidades de ocorrência, a informação contida em cada símbolo é

$$I(a_i) = \log_b \frac{1}{P(a_i)} \quad (2.4)$$

e a entropia é dada pela informação média associada a cada símbolo. *?*

$$H = - \sum_{j=1}^J P(a_j) \log_b P(a_j) \quad (2.5)$$

No caso de imagens digitais, em que a unidade de representação é o *bit*, *tem-se* ~~temos que~~ $b = 2$ na equação 2.5.

Alguns métodos básicos de compressão

De maneira geral, existem dois tipos de compressão, com e sem perda, com as seguintes características:

? mão senior?
dados?

- Sem perda: objetiva comprimir uma determinada informação sem que a mesma seja afetada. Para isso, códigos diferentes do código natural são atribuídos aos símbolos, objetivando reduzir a redundância de codificação.
- Com perda: objetiva alcançar um maior nível de compressão através da eliminação de elementos sem que a informação seja fortemente afetada, de forma que a mesma possa ser entendida.

Métodos

Existem vários métodos de codificação com e sem perda. Dentre eles, os mais utilizados são:

1. Sem perda:

- Huffman: a codificação de Huffman [13] é um método de codificação de tamanho variável que consiste na atribuição de palavras código menores para símbolos mais frequentes e maiores para os símbolos menos frequentes.

O primeiro passo é reorganizar as probabilidades dos símbolos da fonte em ordem decrescente e fazer reduções sucessivas agrupando os símbolos de menor probabilidade de ocorrência, como na figura 2.1.

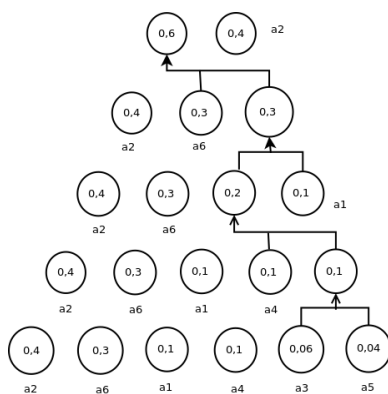


Figura 2.1: Reduções da fonte [1].

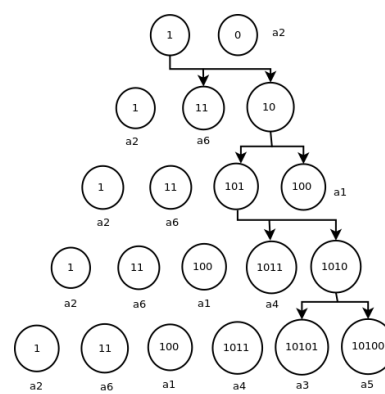


Figura 2.2: Codificação da fonte reduzida [1].

Por fim, deve-se codificar cada fonte reduzida segundo o sentido inverso das reduções atribuindo 0 para as menores probabilidades e 1 para as maiores, ou vice-versa, como na figura 2.2. Desta forma cada símbolo terá uma única palavra código

~~correspondente~~ cujo tamanho é inversamente proporcional a sua probabilidade de ocorrência.

- Codificação aritmética: é um método de codificação de tamanho variável que trabalha de maneira bem diferente em relação a codificação de Huffman. Sua abordagem atribui faixas de valores, entre 0 e 1, para cada símbolo e por fim atribui palavras código aos intervalos [14].

Considerando o alfabeto $s = (s_1, s_2, s_3, s_4)$, suponhamos que se queira codificar a mensagem $s_1s_2s_4$. primейamente calculam-se as probabilidades dos símbolos-fonte, as quais irão ocupar seguimentos proporcionais do intervalo $[0, 1]$. Depois devemos fazer sucessivas divisões proporcionais dos símbolos dentro dos seguimentos, seguindo a ordem dos símbolos da mensagem, como na tabela 2.1.

s	$p(s)$	Faixa	s_1	s_1s_2	$s_1s_2s_4$
s_1	0,2	$[0,0; 0,2)$	$[0,0; 0,04)$	$[0,04; 0,048)$	$[0,072; 0,0736)$
s_2	0,2	$[0,2; 0,4)$	$[0,04; 0,08)$	$[0,048; 0,056)$	$[0,0736; 0,0752)$
s_3	0,4	$[0,4; 0,8)$	$[0,08; 0,16)$	$[0,056; 0,072)$	$[0,0752; 0,0784)$
s_4	0,2	$[0,8; 1,0)$	$[0,16; 0,2)$	$[0,072; 0,08)$	$[0,0784; 0,08)$

Tabela 2.1: Exemplo de codificação aritmética.

Por fim, escolhe-se um número dentro do intervalo atribuído para uma determinada mensagem que deverá representar a mesma. No caso da mensagem $s_1s_2s_4$, foi atribuída a faixa $[0,072; 0,08)$ e cada valor dentro da mesma poderá ser escolhido para representar esta mensagem.

- Run-length: inicialmente produzida para a compressão para ser utilizada na tecnologia de FAX, cujas imagens são binárias.

A codificação run-length é executada linha a linha começando com o valor inicial (0 ou 1) seguido pelo número de repetições sucessivas. Quando houver a mudança de valor basta acrescentar o números de repetições sucessivas, pois sabe-se que o próximo valor é a negação do anterior.

2. Com perda:

- DPCM: Considerando que uma determinada amostra pode ser representada por

$$f(n) = \hat{f}(n) + e(n) \quad (2.6)$$

em que $\hat{f}(n)$ é uma aproximação da amostra original e $e(n)$ é erro associado à mesma, o erro médio quadrático entre f e \hat{f} pode ser minimizado através de uma melhor aproximação do sinal original.

Na codificação DPCM (~~do inglês~~ Differential Pulse-Code Modulation), se $e(n) \rightarrow 0$, temos que a aproximação do sinal original pode ser representada por uma combinação linear descrita ~~em~~ *por*

$$\hat{f}(n) = \sum_{i=1}^m \alpha_i f(n-i) \quad (2.7)$$

em que os coeficientes devem ser calculados através da minimização da expressão

2.8.

juntar

$$E\{e(n)^2\} = E \left\{ \left[f(n) - \sum_{i=1}^m \alpha_i f(n-i) \right]^2 \right\} \quad (2.8)$$

- Codificação baseada em transformada de blocos: é uma técnica de compressão que consiste em dividir uma imagem em blocos não sobrepostos de tamanhos iguais (geralmente 8×8). Uma transformada linear reversível, como a transformada de Fourier e a transformada cosseno, é utilizada para mapear estes blocos no domínio da frequência que por fim serão submetidos a um processo de quantização [1].

Os padrões JPEG e MPEG

O processo evolutivo da espécie humana deu-se de forma que a visão foi o sentido que mais se desenvolveu: cerca de 80 – 90% dos neurônios estão relacionados com o processamento de informações visuais [15]. Dessa forma, não é de se surpreender que imagens e vídeos sejam cada vez mais explorados digitalmente.

Seguindo essa tendência, intensificaram-se as buscas por métodos capazes de otimizar a utilização da banda de transmissão sem que a informação seja prejudicada. Desta forma surgiu a necessidade de padronização de métodos de compressão de imagens e vídeos, dando origem ao JPEG [16] e MPEG-1 [17], [18], [19].

JPEG

Em meados da década de 80, a União Internacional de Telecomunicações (ITU, do inglês International Telecommunication Union) concentrou seus esforços para a criação de um padrão de compressão de imagens estáticas. Desta forma deu-se a origem do JPEG.

Este padrão consiste em uma combinação de duas técnicas de compressão, com e sem perda (quantização e codificação de entropia). Como pode ser notado nas figuras 2.3 e 2.4, o codificador e o decodificador, respectivamente, do *JPEG baseline* [1] são destacados pela cor azul.

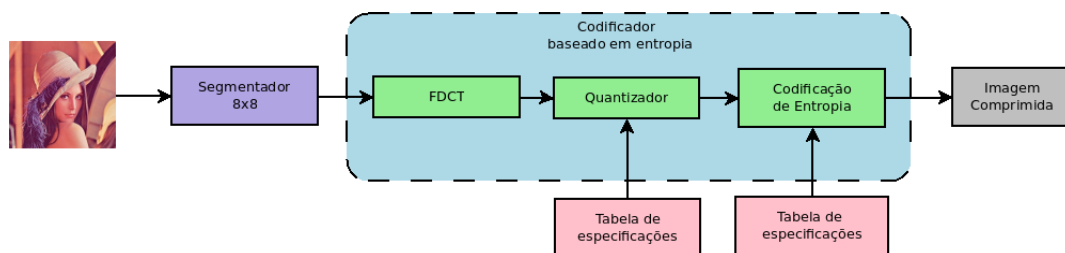


Figura 2.3: Fluxograma do codificador JPEG.

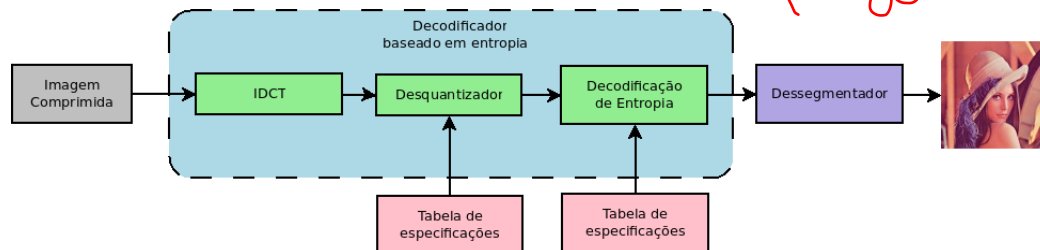


Figura 2.4: Fluxograma do decodificador JPEG.

está certa a ordem dos blocos?

JPEG baseline

No sistema “baseline” [1] o processo de compressão é composto por quatro passos sequenciais: segmentação, cálculo da transformada cosseno discreta, quantização e determinação dos códigos de tamanho variável para cada símbolo.

Inicialmente, a imagem é subdividida em blocos 8×8 . Depois que os blocos são encontrados, seus valores são deslocados, subtraindo 2^{k-1} unidades, em que 2^k é o número máximo

de níveis de intensidade. Então, aplica-se a transformada cosseno, onde $p(x, y)$ é o valor do pixel na posição (x, y) e N é a ordem do bloco (oitava ordem)[20],

$$D(i, j) = \frac{1}{\sqrt{2N}} C(i) C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} p(x, y) \cos \left[\frac{(2x+1)i\pi}{2N} \right] \cos \left[\frac{(2y+1)j\pi}{2N} \right] \quad (2.9)$$

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } u = 0 \\ 1 & \text{if } u > 0 \end{cases} \quad (2.10)$$

seguida pelo processo de quantização,

$$\hat{T}(u, v) = \text{round} \left(\frac{T(u, v)}{Z(u, v)} \right) \quad (2.11)$$

em que $\hat{T}(u, v)$ é o bloco quantizado (Tabela 2.5), $T(u, v)$ é o bloco transformado e $Z(u, v)$ é a tabela de quantização padrão (Tabela 2.2) multiplicada por um fator de qualidade.

Através de experimentos subjetivos a respeito da percepção visual humana obteve-se a tabela de quantização padrão do JPEG. A utilização da mesma gera uma qualidade visual de 50%, que pode ser alterada multiplicando-a por um fator de qualidade: se a qualidade desejada for superior a 50% o fator usado deve ser $(100 - \text{qualidade})/50$, caso contrário o fator é $50/\text{qualidade}$.

Por fim, os coeficientes dos blocos quantizados são reordenados no padrão zigzag (Tabela 2.3), resultando em um vetor,

$$[-26 -3 1 -3 -2 -6 2 -4 1 -4 1 1 5 0 2 0 0 -1 2 0 0 0 0 0 -1 -1 \text{EOB}]$$

e codificados com base nas tabelas de Huffman pré definidas, as quais agrupam derivações dos métodos DPCM e run-length¹, como descrito no Apêndice A.1.

O processo de decodificação é obtido através da simples inversão da ordem das operações.

MPEG-1

O padrão H.261 [21] é um método versátil de compressão de vídeos com perda, pois pode ser aplicado a uma grande variedade de formatos de entrada. Porém, foi otimizado para aplicações que suportam taxas contínuas de transferência de bits de 1,5 Mbps, *sum italiano*

Como mencionado na Seção 2.2, os vídeos estão sujeitos à redundância temporal devido à alta semelhança entre imagens vizinhas. Por isso, antes de falar do MPEG-1, é conveniente abordar o processo de estimação e compensação de movimentos baseados em blocos, a fim de eliminar a redundância temporal.

¹ A palavra código EOB significa que os coeficientes são iguais a 0 daquele ponto até o último coeficiente AC.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

Tabela 2.2: Tabela de quantização padrão (50%).

Tabela 2.3: Padrão zigzag.

52	55	61	66	70	61	64	73
63	59	66	90	109	85	69	72
62	59	68	113	144	104	66	73
63	58	71	122	154	106	70	69
67	61	68	104	126	88	68	70
79	65	60	70	77	63	58	75
85	71	64	59	55	61	65	83
87	79	69	68	65	76	78	94

-26	-3	-6	2	2	0	0	0
1	-2	-4	0	0	0	0	0
-3	1	5	-1	-1	0	0	0
-4	1	2	-1	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Tabela 2.4: Exemplo de subimagem.

Tabela 2.5: Subimagem da tabela 2.4 quantizada.

Estimação e compensação de movimentos baseados em blocos

Objetivando-se reduzir a redundância temporal entre imagens consecutivas poder-se-ia pensar em armazenar apenas a diferença entre duas imagens, porém este processo pode ser otimizado através da utilização de macroblocos. Por isso pode-se dizer que este método é uma variação da codificação DPCM.

Macroblocos são compostos por 4 blocos $n \times n$, na maioria das vezes 8×8 . Na ~~figura~~ ^F 2.5 é representado um macrobloco dizimado no padrão 4:2:0 [22].

Utilizando macroblocos para a encontrar a diferença entre imagens consecutivas torna-se possível a minimização do erro médio absoluto (MAE, ~~do inglês~~ ^{do inglês} mean absolute error)

$$MAE(i, j) = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - \underline{R}(x+i+k, y+j+l)| \quad (2.12)$$

O que são \underline{C} e \underline{R} ?

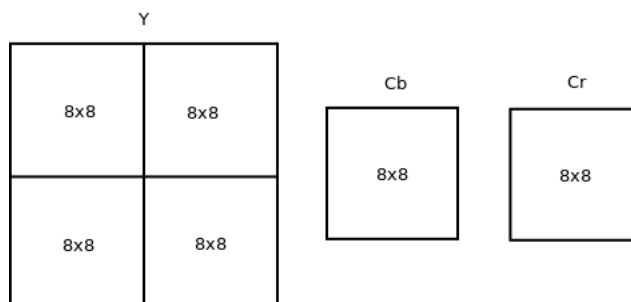


Figura 2.5: Macrobloco dizimado no padrão 4:2:0.

da mesma. Dessa forma obtendo uma maior redução de volume de dados.

Inicialmente considera-se uma imagem atual em que se tem um macrobloco de dimensões 16×16 e uma imagem anterior com um macrobloco de mesmas dimensões com a menor diferença possível. Estes dois macroblocos possuem um deslocamento relativo designado por “vetor de deslocamento” e esta diferença é designada por “erro de predição”. A este processo dá-se o nome de compensação de movimento pra frente (Fig. 2.7).

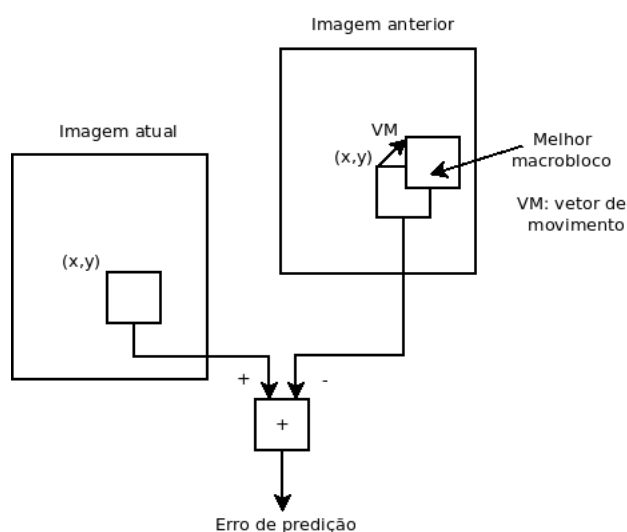


Figura 2.6: Compensação de movimento pra frente.

Também pode-se estender este raciocínio, tomando como base uma imagem atual, uma anterior e uma posterior a fim de obter a menor diferença possível. Este processo chama-se compensação de movimento bidirecional (Fig. 2.7).

A diferença fundamental entre os dois modelos de compensação de movimento é que o bidirecional trabalha com três erros de predição (pra trás, pra frente ou interpolativo) e não um,

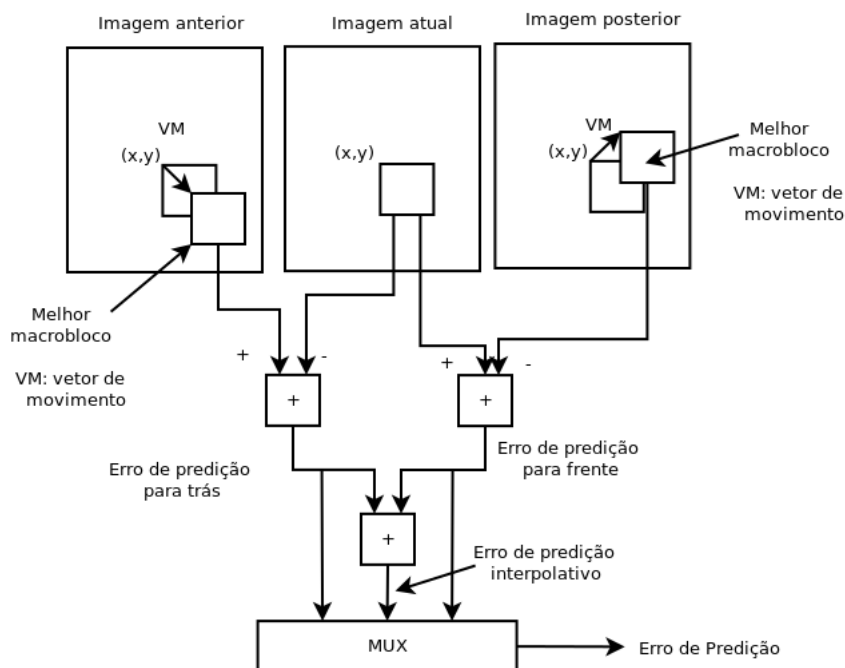


Figura 2.7: Compensação de movimento bidirecional.

como na predição para frente.

Métodos de busca

Há muitas formas de estimar os vetores de deslocamento dos macroblocos, algumas possibilitam uma maior precisão ao encontrar o melhor candidato, outras possibilitam uma maior velocidade na determinação do mesmo.

Dentre os métodos básicos, podemos citar [14]:

1. **Busca completa:** apesar de ser o método mais simples, garante que o melhor candidato seja encontrado, obtendo o menor valor possível do MAE.

Consiste na definição de uma área de busca de p pixels para os lados, para cima e para baixo a partir do canto superior esquerdo do macrobloco. Para cada um dos $(2p + 1)^2$ pixels da área de busca serão formados os possíveis macroblocos a serem submetidos ao critério de minimização do MAE.

2. **Busca unidimensional paralela hierárquica:** este método não garante o menor valor possível para o MAE, porém apresenta um ganho considerável em velocidade, em

Este algoritmo de busca é descrito da seguinte forma:

comparação ...

(a) Para uma área de busca $[-p, p]$, definida no método de busca completa, define-se $S = 2^{\lfloor \log_2 p \rfloor}$ e assumindo que a posição de origem (d_i, d_j) do macrobloco seja $(0, 0)$.

(b) Em paralelo, deve-se calcular:

- Para o eixo i : encontrar dentre as três posições $(d_i - S, d_j)$, (d_i, d_j) e $(d_i + S, d_j)$, qual delas gera o menor MAE. Por fim substituir d_i pela posição encontrada.
- Para o eixo j : encontrar dentre as três posições $(d_i, d_j - S)$, (d_i, d_j) e $(d_i, d_j + S)$, qual delas gera o menor MAE. Por fim substituir d_j pela posição encontrada e $S = \frac{S}{2}$.

O passo (b) deve ser repetido sucessivas vezes até que $S = 0$. O vetor resultante (d_i, d_j) será o vetor de deslocamento do macrobloco.

Quantização

O processo de quantização utilizado no H.261 é similar ao utilizado no JPEG baseline, diferindo apenas no fato de utilizar duas tabelas de quantização, para as codificações intra e inter imagens (tabelas 2.6 e 2.7).

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Tabela 2.6: Tabela intra.

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Tabela 2.7: Tabela inter.

É importante deixar claro que a necessidade de duas matrizes de quantização está relacionada com o fato de que para as imagens preditas as componentes de alta frequência não são necessariamente frequências espaciais, pois também podem ser provenientes do efeito de bloqueio e das limitações do processo de compensação de movimentos [23]. Portanto, a padronização

de uma tabela de quantização ponderada para as imagens inter codificadas seria uma tarefa muito complexa.

Algoritmo

O padrão H.261 não reconhece entradas entrelaçadas, por isso utiliza-se a denominação de “imagens” e não “frame”. Há três tipos de imagens que podem ser classificadas em dois métodos de compressão:

1. Intra imagem: as imagens I, como são conhecidas, são imagens que são codificadas de forma semelhante ao padrão JPEG. Neste tipo de codificação não há redução de redundância temporal, logo exige um volume de dados considerável para armazenamento.
2. **I**nter imagens: são imagens que passam pelo processo de compensação de movimentos em que as imagens de erro geradas são codificadas de forma semelhante ao JPEG.

Nesta classe se encaixam dois tipos de imagens, imagens P e B, que alcançam diferentes taxas de compressão:

- P (predita): basicamente a imagem é processada com base na imagem I ou P anterior, em que os vetores de deslocamento são parte da imagem anterior para a imagem atual. Estas imagens alcançam maiores taxas de compressão quando comparadas com as imagens I.
- B (bidirecionalmente predita): imagem processada com base na imagem I anterior e na imagem P posterior, ou vice versa. A possibilidade de escolha entre dois erros de predição e o erro interpolativo possibilita que estas imagens alcancem taxas menores de compressão quando comparadas com as imagens P.

confirma?

O processamento de imagens inter codificadas possui ~~em~~ maior complexidade quando comparado com as imagens I, não apenas pela compensação de movimento mas também por casos especiais que ocorrem durante o processo de codificação, como:

1. Anulação de vetores de deslocamento ocorre em situações em que o erro de predição usando um vetor não nulo pode ser muito próximo do obtido através de um vetor nulo, pois neste caso a codificação dos vetores de deslocamento pode gerar a expansão do volume de dados necessários para a representação de um macrobloco.

Este recurso é frequentemente utilizado, a não ser que o erro de predição a partir de vetores não nulos seja no mínimo 1, 1 ^{vezes} menor que o obtido por vetores nulos.

2. Os macroblocos podem ser intra ou inter codificados. Pois há vezes em que a codificação intra é mais vantajosa do que a a inter.
3. Os macroblocos podem ou não ser codificados. A não codificação é utilizada quando todos os coeficientes quantizados da DCT são nulos, neste caso este macrobloco será uma cópia do presente na imagem de referência anterior.

Para o codificador (Fig. 2.8), inicialmente os canais Cb e Cr são dizimados segundo o padrão 4:2:0 e define-se a ordem dos tipos de imagens dentro de um GOP (do inglês group of images), de acordo com as necessidades. A mais comum é mostrada na figura 2.9.

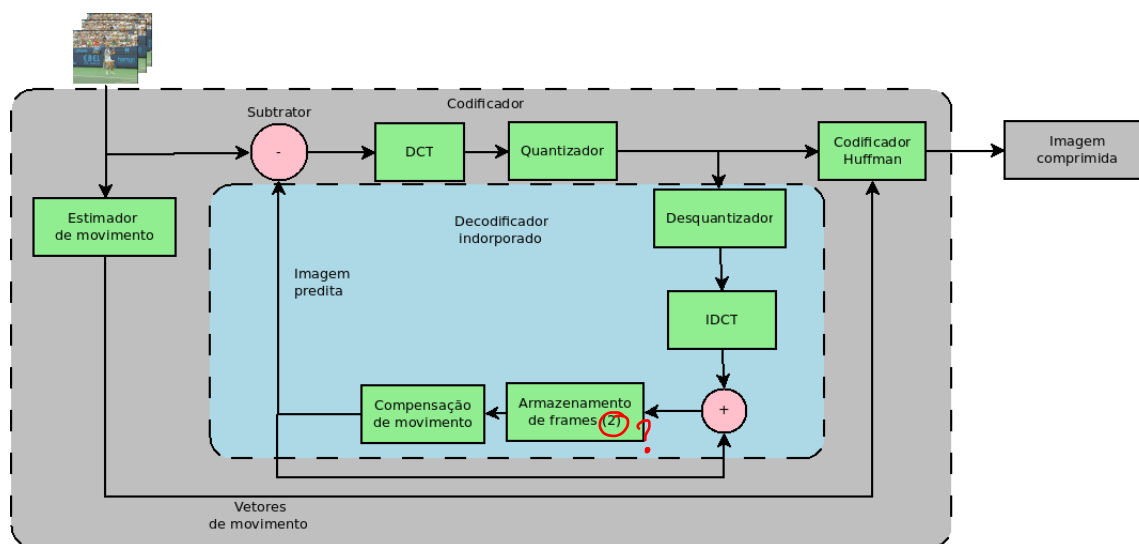


Figura 2.8: Fluxograma do codificador MPEG.

Depois que cada tipo de imagem foi processada, as imagens I, P e B são codificadas conforme descrito no padrão JPEG, porém as duas últimas utilizam a codificação de Huffman descrita no ^Aapêndice B.1 para armazenar os vetores de deslocamento.

Para o decodificador (Fig. 2.10), deve-se trabalhar por grupos de imagens, recuperando as imagens segundo o decodificador JPEG e recuperando as imagens ($P_1, P_2, P_3, \dots, P_n$) originais na ordem I, P, B, B, I, B e B para, se for necessário, realizar a compensação de movimento.

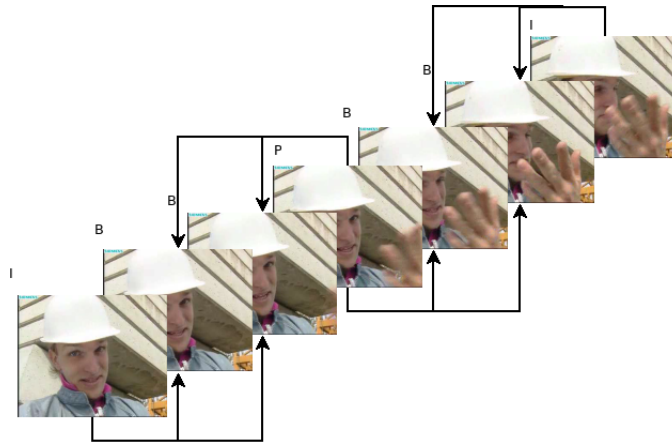


Figura 2.9: Sequência de imagens dentro de um GOP.

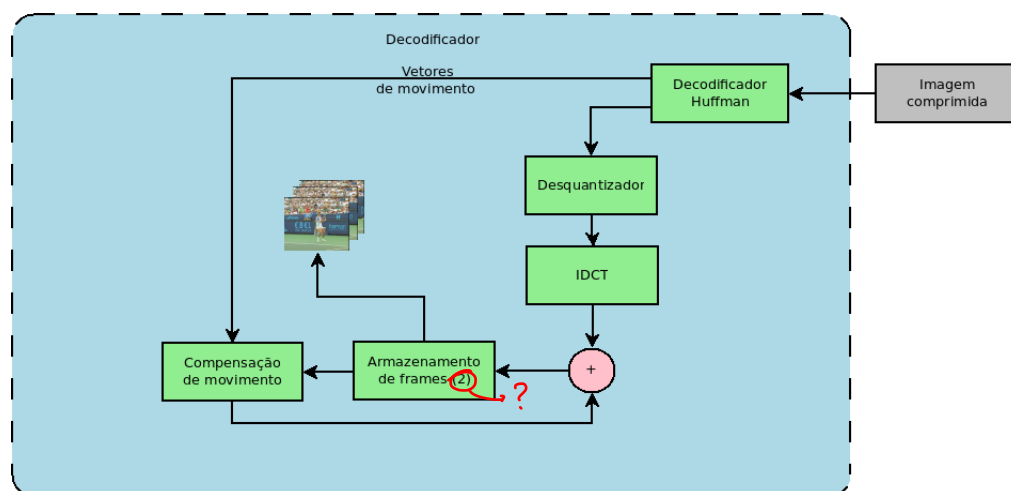


Figura 2.10: Fluxograma do decodificador MPEG.

Diagrama em blocos

Capítulo 3

Qualidade visual

Introdução

“Uma imagem vale mais do que mil palavras”, este ditado expressa a importância que os seres humanos dão às informações visuais. Não é de se espantar que existam grandes esforços voltados para pesquisas que objetivam encontrar métodos capazes de avaliar, bem como melhorar a qualidade visual de imagens e vídeos recuperados através do processo de decodificação. *decodificados*

O conhecimento das características do sistema visual humano (HVS, do inglês human visual system) é fundamental para o desenvolvimento de metodologias eficazes no processo de avaliação e melhoria da qualidade visual de informações visuais. Embora o ~~nosso~~ conhecimento do HVS seja muito limitado, há de se reconhecer que ~~já nos~~ proporcionaram bons resultados, como em [11] e [24], em que são apresentados dois modelos de quantização adaptativa, um com base na percepção visual humana em determinadas frequências espaciais e temporais e outro que utiliza métricas de qualidade perceptual na fase de estimação de movimentos dos macroblocos, ambos alcançando bons resultados visuais em baixas taxas. Neste contexto, os processos de captura, exibição, armazenamento e transmissão deverão ser adaptados a fim de gerar representações mais exatas das imagens reais.

Este capítulo tem como objetivo apresentar os aspectos principais da qualidade visual, bem como apresentar maneiras de quantificá-la, com foco em compressão de imagens e vídeos digitais.

Artefatos provenientes do processo de compressão

Na ~~Seção~~ 2.5 foi mencionado que o JPEG e o MPEG-1 realizam uma quantização dos blocos transformados, a fim de eliminar as redundâncias espacial e temporal. Em alguns sistemas este processo é o principal responsável pelo surgimento de distorções, apesar de não ser o único fator capaz de afetar a qualidade visual. Alguns dos tipos de artefatos mais comuns em sequências de vídeos são listados a seguir:

- Efeito de blocagem: ocorre devido a quantização independente de blocos, gerando descontinuidade ~~nos limites dos blocos adjacentes.~~
na fronteira com
- Embaçamento: consequência da supressão das componentes de alta frequência durante o processo de quantização, manifestando-se através da perda da resolução espacial.
- Vazamento de cores: devido aos processos de subamostragem de crominância seguido pela quantização, ocorre o vazamento de cores entre áreas de crominância muito diferentes.
- Efeito de imagem base da DCT: ocorre quando uma única componente da DCT é dominante em um macro bloco, resultando na ênfase de uma imagem base da DCT.
- Efeito ressonante: está relacionado com o fenômeno de Gibbs [1], logo é mais evidente em áreas de grande contraste. Resultando da irregularidade de reconstrução das componentes de baixa frequência devido ~~o~~ processo de quantização.
- Aliasing: ocorre quando a frequência de amostragem está abaixo da frequência de Nyquist [1], tanto espacialmente quanto temporalmente.

Avaliações subjetiva e objetiva

A melhor forma de conseguir avaliar a qualidade de informações visuais é através da opinião de observadores. A nota média de opiniões (MOS, do inglês Mean Opinion Score), avaliação subjetiva que necessita da opinião de um grupo de observadores, é a melhor forma de avaliar a qualidade de uma imagem se se pretende que essa avaliação seja de acordo com critérios de percepção humana, porém custosa.

Frequentemente o erro médio quadrático (MSE, do inglês Mean Square Error) é utilizado pelos ~~métodos~~ ^{de} avaliação de qualidade de imagens, devido a sua fácil implementação. No

entanto, esses métodos apresentam baixo nível de correlação com as avaliações subjetivas, pois o MSE não leva em consideração características espaciais [25].

As pesquisas de avaliação objetiva de imagens buscam contornar os inconvenientes presentes na avaliação subjetiva, através de modelos matemáticos que simulam a percepção visual humana.

As duas abordagens possíveis para o desenvolvimento de simuladores perceptuais: bottom-up e top-down. *?? Certeza?*

A abordagem mais comum é estudar o funcionamento de cada elemento relevante para a percepção do sistema visual humano de forma a combiná-los em um sistema computacional. Esta abordagem é comumente conhecida como bottom-up.

As abordagens diferentes *??* são conhecidas como top-down. Em que o sistema é tido como uma caixa preta onde comportamentos hipotéticos são implementados e ajustados. Os pontos atrativos desta abordagem são que o único conhecimento prévio necessário é a relação entre a entrada e a saída do sistema e a simplicidade de implementação.

Classificação das métricas de avaliação de qualidade visual

segundo a disponibilidade de uma imagem de referência
Para que uma métrica seja realmente útil para a avaliação de qualidade visual é necessário que ela seja capaz de suprir as limitações do MSE. Objetivando-se contornar essa situação, *??* surgiram vários métodos que, a grosso modo, podem ser classificados em três categorias:

- Referência completa: *há uma imagem de referência* utiliza a ~~imagem original~~ (considerada sem distorção) para avaliar uma imagem distorcida. Portanto, proporciona resultados mais precisos em relação a similaridade e fidelidade entre as duas imagens.
- Sem referência: utilizada quando não é possível ter acesso a imagem ~~original~~ *de referência*, logo a avaliação da imagem distorcida deve ser feita às "cegas", o que faz disso uma tarefa difícil. Por isso esta categoria também é conhecida como de referência cega.
- Referência reduzida: neste caso a imagem de referência não está completamente disponível e sim algumas características, que são embutidas no sistema que irá avaliar a imagem distorcida.

Outra classificação possível seria em relação a utilização das características do HVS. Neste caso, existem duas categorias:

- Perceptuais: são formulações matemáticas de certa complexidade inspiradas em características fisiológicas e psicovisuais da visão que representam de forma automática a percepção humana perante uma representação de uma informação visual, com expressiva correlação com a real percepção humana.
- Não perceptuais: não utilizam características do HVS nas suas formulações e têm como virtude a baixa complexidade computacional. Porém, possuem baixa correlação com as avaliações subjetivas.

A seguir serão apresentadas ~~as~~ métricas não perceptuais e ~~depois~~ as perceptuais ^{de especial interesse para este} utilizadas neste trabalho.

Não perceptuais

Erro médio quadrático (Mean Square Error - MSE)

O MSE é uma métrica bastante utilizada ~~que~~ que consiste no valor esperado dos ~~quadrados~~ dos erros,

$$MSE = E [(y(i, j) - x(i, j))^2] \quad (3.1)$$

em que y é a imagem distorcida, x é a imagem ^{de referência} original e (i, j) determinam a posição dos pixels.

Razão Sinal-Ruído (Signal-to-Noise Ratio - SNR)

Quantifica o quanto um sinal foi distorcido através do cálculo da energia do erro da imagem distorcida.

A SNR é comumente medida em dB da seguinte forma,

$$SNR = 10 \log_{10} \frac{\sum_{i,j} [x(i, j)]^2}{\sum_{i,j} [x(i, j) - y(i, j)]^2} \quad (3.2)$$

em que y é a imagem distorcida e x é a imagem original.

Razão Sinal-Ruído de Pico (Peak Signal-to-Noise Ratio - PSNR)

Comumente utilizada para medir a qualidade da reconstrução da imagem ou vídeo após uma compressão com perdas.

A PSNR é comumente medida em dB da seguinte forma,

$$PSNR = 10 \log_{10} \frac{NK^2}{\sum_{i,j} [x(i, j) - y(i, j)]^2} \quad (3.3)$$

onde $x(i, j)$ representa o sinal de referência, $y(i, j)$ representa o sinal de teste, N representa o número total de pixels da imagem e K representa o valor máximo que um pixel pode atingir. No caso de uma imagem de 8 bits/pixel o valor de K é 255.

Perceptuais

Índice de Similaridade Estrutural (Structural Similarity Index - SSIM)

O SSIM talvez seja a função de aproximação perceptual humana mais utilizada, devido a sua baixa complexidade de implementação em relação às demais e obter boas aproximações do HVS. Esta métrica avalia o quanto a estrutura da imagem de teste é diferente da estrutura da imagem de referência.

O SSIM é definido como

$$SSIM = [l(x, y)^\alpha c(x, y)^\beta s(x, y)^\gamma] \quad (3.4)$$

em que $\alpha > 0$, $\beta > 0$ e $\gamma > 0$ são responsáveis pelo ajuste da importância relativa das componentes $l(x, y)$, $c(x, y)$ e $s(x, y)$ correspondem às componentes de luminância, contraste e estrutura, respectivamente, definidas como

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \quad (3.5)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (3.6)$$

$$s(x, y) = \frac{2\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \quad (3.7)$$

em que μ_x e σ_x são a média e o desvio padrão da imagem x ; μ_y e σ_y são a média e o desvio padrão da imagem y ; σ_{xy} é a covariância entre as imagens x e y . Os valores c_1 , c_2 (3) c_3 são inseridos com o objetivo de evitar instabilidades. *Termos*

Esta métrica é aplicada localmente, deslocando, horizontalmente e verticalmente, uma janela de tamanho $B \times B$ sobre a imagem. Pode-se calcular a média do SSIM (MSSIM) para que se tenha uma índice de qualidade geral da imagem, em que M é o número de janelas.

$$MSSIM = \frac{1}{M} \sum_{j=1}^M SSIM_j \quad (3.8)$$

definir

Quantização inter adaptativa baseada no HVS

A grosso modo, o sistema visual humano consegue captar pequenas variações de brilho em áreas relativamente grandes, mas não consegue captar com a mesma facilidade quando estas variações estão presentes em componentes de alta frequência [2]. Os sistemas de compressão obtêm vantagens desse fato ao quantizar fortemente as componentes de alta frequência e de maneira mais branda as componentes de baixa frequência. Como pode ser visto na figura 3.1, a DCT agrupa as componentes de baixa frequência no canto superior esquerdo e as de alta frequência no inferior direito do macrobloco transformado, possibilitando a produção de tabela de quantização apropriadas para o propósito mencionado anteriormente.

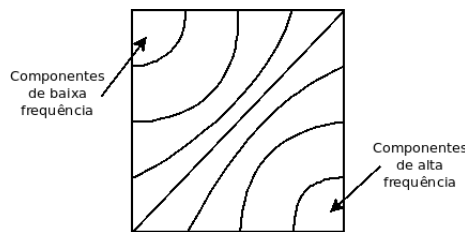


Figura 3.1: Distribuição da frequência espacial no domínio da DCT.

O padrão H.261 [17] utiliza o mesmo método de compressão para as imagens intra e inter codificadas. Atualmente vários métodos que levam em consideração as características do HVS têm sido propostos para melhorar a qualidade do vídeo reconstruído.

Em [11], um algoritmo de quantização inter adaptativo é proposto. Este algoritmo baseia-se no modelo apresentado por D. H. Kelly [2], o qual afirma que o sistema visual humano é mais sensível às variações de contraste nas frequências espaciais intermediárias, ao passo que é menos sensível em frequências baixas e altas (Fig. 3.2). Tal algoritmo gera matrizes com base nas limitações do HVS em relação às respostas espaciais e temporais ao propor uma equação que representa a resposta do HVS à frequência temporal em função da velocidade do macrobloco e da frequência espacial

$$G(\alpha, v) = [6.1 + 7.3 |\log(v/3)|^3] v \alpha^2 \exp [-2\alpha(v + 2)/45.9] \quad (3.9)$$

em que v velocidade do macrobloco, medida em graus por segundo, e α é a frequência espacial medida em ciclos por segundo. Ambas são definidas por

$$v = (v_H^2 + v_T^2)^{1/2} \quad \text{e} \quad \alpha_{ij} = \alpha_i + \alpha_j \quad (3.10)$$

separar em duas equações

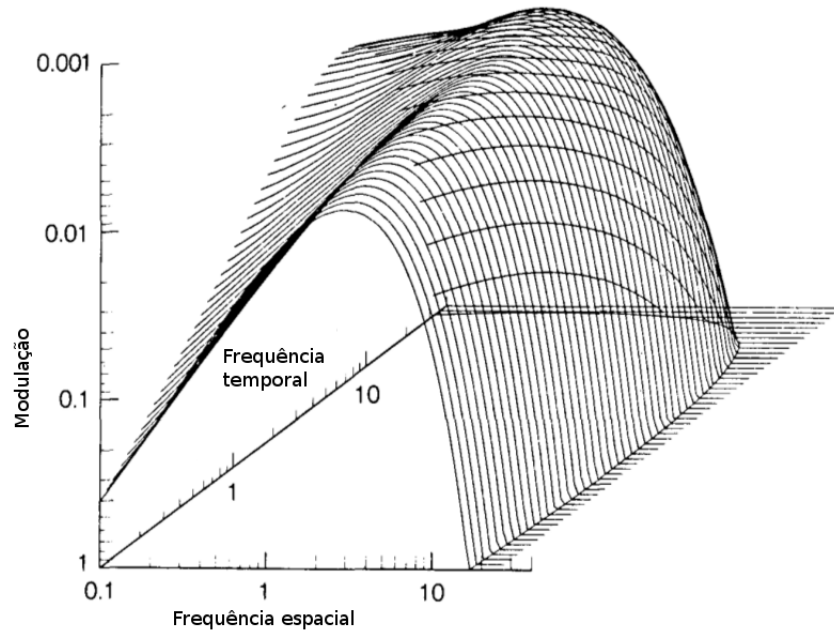


Figura 3.2: Visão perspectiva da superfície de limiar espacial e temporal [2]. Cada curva representa a resposta a frequência espacial em uma frequência espacial fixa.

em que v_H e v_T são as componentes horizontal e vertical da velocidade do macrobloco e α_i e α_j são as componentes vertical e horizontal da frequência espacial. Ambas são definidas por

$$v_H = \frac{m_H}{w} \times f \quad \text{e} \quad v_T = \frac{m_T}{h} \times f \quad (3.11)$$

$$\alpha_i = \frac{m_H}{m} \times \frac{w}{m} \times c_i \quad \text{e} \quad \alpha_j = \frac{m_T}{m} \times \frac{h}{m} \times c_j \quad (3.12)$$

$$c_i = 0.5i \quad \text{e} \quad c_j = 0.5j \quad i, j = 1, 0 \dots 7 \quad (3.13)$$

em que m_H e m_T são os valores absolutos das componentes do vetor de deslocamento do macrobloco¹, w e h são as dimensões da imagem, e m é o tamanho do macrobloco.

Desta forma, espera-se melhorar a qualidade subjetiva do vídeo decodificado ao quantizar individualmente cada macrobloco respeitando as características do HVS segundo o modelo proposto em [11]

$$Q_{HVS}(i, j) = (m_H + m_T)/p \left(1 - \frac{G(\alpha_{ij}, v)}{G_{max}} \right) \quad (3.14)$$

em que $Q_{HVS}(i, j)$ é uma componente da matriz de quantização, p é um parâmetro de ajuste e G_{max} é o valor máximo de Q_{HVS} .

¹Logo, este método de quantização perceptual só é aplicado as imagens preditas (P, B), cuja qualidade depende diretamente do fator de qualidade das imagens I.

Por fim, o modelo proposto pode ser obtido somando a matriz de quantização baseada no HVS a matriz de quantização plana.

$$Q = Q_{flat} + Q_{HVS} \quad (3.15)$$

*Acho que terminou
um tanto repentinamente.*

Capítulo 4

Procedimento experimental e resultados

Implementação do sistema

Para a realização dos experimentos práticos, ^{mm} ~~o~~ ^{definir} ~~codec, referentes às informações visuais,~~
~~de padrão H.261~~ ^{foi} ~~fora~~ implementado em Python [26], bem como o método de quantização perceptual interadaptativa (seção 3.5).

Conjunto de teste

A análise do método de quantização perceptual [11] utilizado neste trabalho consiste em compará-lo com o utilizado no codificador H.261 padrão com diferentes condições. Para este procedimento quatro vídeos com diferentes características, obtidos em [27], serão utilizados ^T (tabela 4.1), de forma a analisar a compressão e a qualidade visual alcançada.

Vídeos	Atividade Temporal	Duração (seg)	Resolução	FPS
Akiyo_cif	Baixa	12	352 x 288	25
Ice_cif	Baixa	8	352 x 288	30
Stefan_cif	Média	3	352 x 288	25
Foreman_cif	Alta	12	352 x 288	25

Tabela 4.1: Descrição dos vídeos.

Procedimento experimental

Os experimentos realizados têm dois objetivos

Este trabalho tem dois objetivos: avaliar a viabilidade da utilização do método de quantização inter adaptativa e avaliar a utilização de uma tabela plana em detrimento de uma ponderada para a quantização macroblocos preditos.

Logo, o procedimento experimental é dividido em duas etapas, as quais foram submetidas a um processo de avaliação perceptual objetiva.

1. Para analisar em quais situações a quantização perceptual inter adaptativa seria vantajosa, o codificador H.261 com e sem a mesma foi submetido a diferentes fatores de qualidade (0%, 10%, 20%,..., 90%, 100%), de forma a gerar uma variação de taxa em ambos os casos.

Neste trabalho utiliza-se o padrão ~~H.261~~ e não o ~~H.263~~, cujas magnitudes das componentes dos vetores de deslocamento são $[-15; 15]$ e $[-31, 5; 31, 5]$, respectivamente. Portanto, o modelo de quantização perceptual utiliza deslocamento de busca de 15 pixels, a partir da origem, na horizontal e na vertical, $p = 1$, os coeficientes de Q_{flat} iguais a 16 e as taxas utilizadas estão na Tabela 4.1.

(bits/pixel ou bits/amostr)

?? Quais taxas?

2. Para estudar a viabilidade da utilização da tabela de quantização plana, a tabela ponderada utilizada nas imagens I será utilizada nas imagens P e B. ~~Por~~ O resultado proveniente desta configuração será comparado com os obtidos na análise anterior.

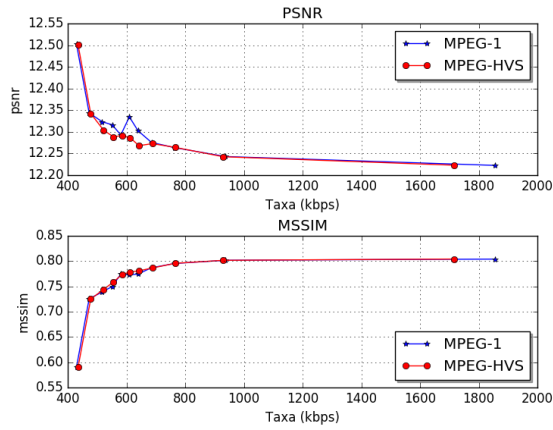
Resultados

Analisando os resultados obtidos, na primeira do procedimento, observa-se que para vídeos com pouca ~~informação de~~ movimento, figuras 4.1(a) e 4.1(b), o codec perceptual apresentou valores de resultados visuais, MSSIM, ~~similares~~ *codecs convencional* aos obtidos pelo padrão. Enquanto que para baixas taxas o codec perceptual apresentou um ganho de qualidade quando submetido a vídeos mais dinâmicos, figuras 4.1(c) e 4.1(d), ~~confirmando~~ *E* a afirmação em [11] de que o método de quantização inter adaptativa beneficia-se do aumento da atividade temporal.

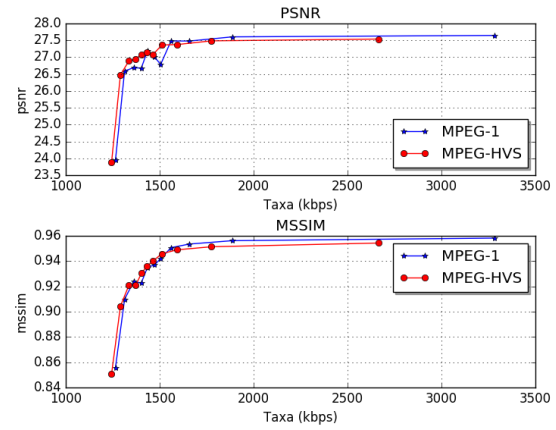
Observa-se que o vídeo Akiyo_cif apresentou um comportamento atípico ~~ao se relacionar~~ *para ao ser comparado com o* os gráficos de PSNR e de MSSIM. Isto é um indício de que o PSNR não é capaz de avaliar de maneira precisa a qualidade visual, sendo para isso necessário a utilização de métodos adicionais capazes enriquecer tal análise.) *→ reformular.*

A melhoria visual obtida pelo codec perceptual deve-se a atribuição de um maior número de bits para as baixas frequências, resultando em uma redução do efeito de blocagem mais acentuado em baixas taxas, como pode ser confirmado através da figura 4.2.

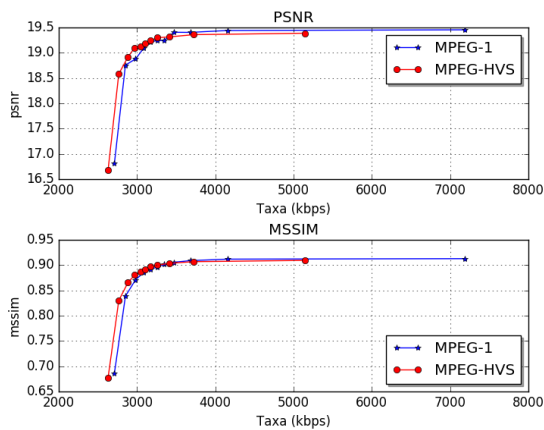
discutível!



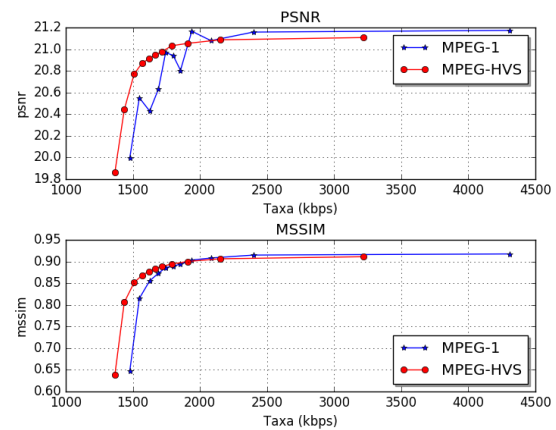
(a) Akiyo.



(b) Ice.



(c) Stefan.



(d) Foreman.

Figura 4.1: Comparativo dos valores de MSSIM e PSNR pelos codecs padrão e perceptual.



(a)



(b)



(c)

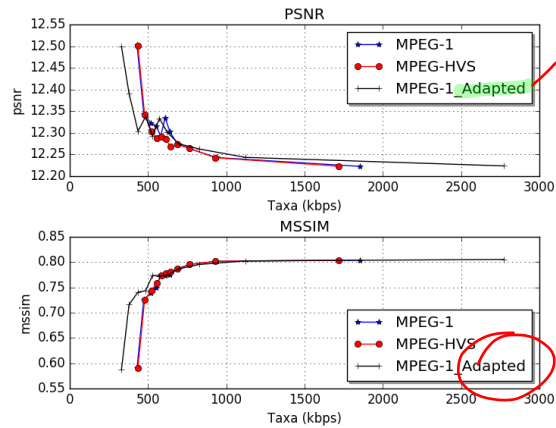


(d)

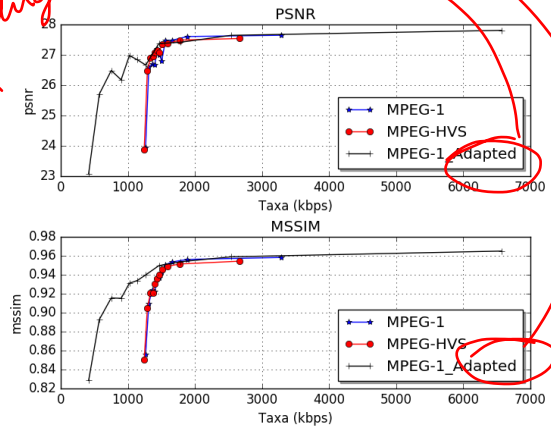
Figura 4.2: Comparativo entre os codecs padrão, (a) e (b), e perceptual, (c) e (d).

Utilizando a tabela ponderada de quantização na codificação das imagens P e B, os resultados obtidos são ainda mais interessantes. Diferentemente do que foi apontado em [23], os resultados obtidos mostram que esta adaptação gera taxas muito menores quando comparadas com os codificadores padrão e perceptual.

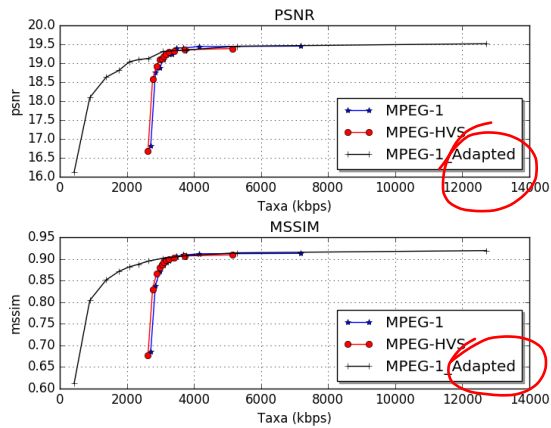
*Comentários
adicionais devem
ser feitos.*



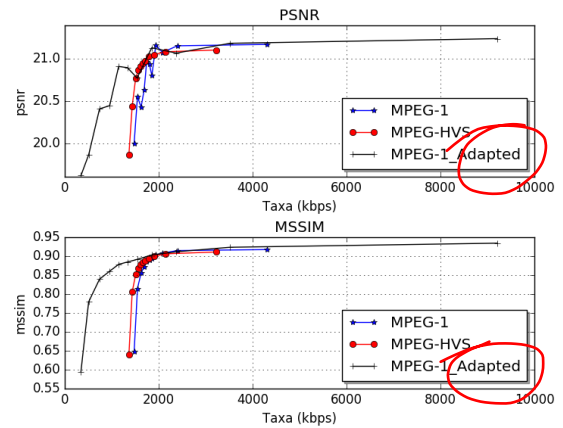
(a) Akiyo.



(b) Ice.



(c) Stefan.



(d) Foreman.

Figura 4.3: Comparativo dos valores de MSSIM e PSNR pelos codecs padrão, padrão adaptado e perceptual.

Capítulo 5

Considerações finais

Conclusões

Trace um retrospecto do texto: o que foi abordado e depois para para os comentários sobre resultados

Apesar de necessário, o processo de compressão de informações visuais é responsável pelo surgimento de artefatos na informação codificada. Portanto, o desafio é encontrar um meio termo entre taxa de compressão e qualidade perceptual.

Através da análise desenvolvida neste trabalho pode-se concluir que a utilização do método de quantização inter adaptativa gera resultados vantajosos para vídeos com alta atividade temporal, enquanto ~~que~~ apresenta resultados similares ao codec padrão quando submetido a vídeos com baixa atividade temporal. Dessa forma, os resultados da avaliação objetiva obtidos neste trabalho então condizentes com os resultados subjetivos apresentados em [11].


Portanto, a utilização do método de quantização inter adaptativa é, se não vantajosa, no mínimo equivalente ao padrão ~~H.261~~.

O mesmo não pode ser dito quando comparamos o codec ~~H.261~~ utilizando apenas a tabela de quantização ponderada em relação aos codificadores analisados na primeira fase do experimento. Apesar de a utilização de uma tabela plana para a quantização dos macroblocos inter codificados possuir embasamento teórico, ~~como em~~ [23], os resultados obtidos utilizando apenas a tabela ponderada foram os melhores alcançados neste trabalho, divergindo da literatura. No entanto, este resultado ~~carece de fundamentações teóricas.~~

requer investigação mais ampla.

Trabalhos futuros

A próxima fase desta pesquisa tem como objetivos: obter o entendimento a respeito do funcionamento de codecs mais atuais; analisar de maneira mais profunda os resultados alcançados na segunda fase experimental e estudar outros métodos de melhoria de qualidade perceptual mais eficazes.



*um pouco
mais de
detalhes.*

Apêndice A

Tabelas de codificação para compressão JPEG

Objetivando otimizar o processo de compressão JPEG, tabelas Huffman pré definidas são utilizadas para a codificação de entropia. Neste apêndice explica-se como utilizar estas tabelas.

Considerando que se deseja codificar um bloco transformado e quantizado,

[-26 -3 1 -3 -2 -6 2 -4 1 -4 1 1 5 0 2 0 0 -1 2 0 0 0 0 0 -1 -1 EOB]

devemos separar este processo em duas partes:

1. Codificação do coeficiente DC: supondo que o coeficiente DC do macrobloco imediatamente anterior é (-17) ,¹ tem-se que, segundo a tabela A.1, a diferença $d = [-26 - (-17)]$ ou $d = -9$ pertence à categoria $k = 4$. Conforme a tabela A.2, o código base utilizado para a categoria 4 é 101 e seu tamanho total é de 7 bits, em que os k bits restantes devem ser os k bits menos significativos da diferença positiva d ou os k bits menos significativos da diferença negativa d menos 1. Para a diferença $d = -9$ os k bits menos significativos são (0111)-1 ou 0110. Logo, a palavra código para o coeficiente DC em questão é 1010110.
2. Codificação do coeficiente AC: sua diferença em relação a codificação de coeficientes DC é que ela leva em consideração o número de zeros que antecedem os coeficientes AC não nulos. Logo o primeiro coeficiente AC não nulo (-3) é codificado como 0100, pois segundo a tabela A.1 ele pertence à categoria 2 sem zeros antecedentes, dando origem ao código base 01, e os dois bits restantes são obtidos de maneira similar a utilizada na codificação de coeficientes DC.

¹Se for o primeiro macrobloco de uma linha utiliza-se o valor 0.

Faixa	Categoria da diferença DC	Categoria AC
0	0	N/A
-1, 1	1	1
-3, -2, 2, 3	2	2
-7,..., -4, 4,..., 7	3	3
-15,..., -8, 8,..., 15	4	4
-31,..., -16, 16,..., 31	5	5
-63,..., -32, 32,..., 63	6	6
-127,..., -64, 64,..., 127	7	7
-255,..., -128, 128,..., 255	8	8
-511,..., -256, 256,..., 511	9	9
-1023,..., -512, 512,..., 1023	A	A
-2047,..., -1024, 1024,..., 2047	B	B
-4095,..., -2048, 2048,..., 4095	C	C
-8191,..., -4096, 4096,..., 8191	D	D
-16383,..., -8192, 8192,..., 16383	E	E
-32767,..., -16384, 16384,..., 32767	F	N/A

Tabela A.1: Categorias de coeficientes de codificação JPEG.

Categoria	Código Base	Tamanho	Categoria	Código Base	Tamanho
0	010	3	6	1110	10
1	011	4	7	11110	12
2	100	5	8	111110	14
3	00	5	9	1111110	16
4	101	7	A	11111110	18
5	110	8	B	111111110	20

Tabela A.2: Códigos JPEG DC.

Zeros/ Categoria	Código base	Tamanho	Zeros/ Categoria	Código base	Tamanho
0/0	1010	4	8/1	111111000	9
0/1	00	2	8/2	111111111000000	15
0/2	01	2	8/3	1111111110110110	16
0/3	100	3	8/4	1111111110110111	16
0/4	1011	4	8/5	1111111110111000	16
0/5	11010	5	8/6	1111111110111001	16
0/6	1111000	7	8/7	1111111110111010	16
0/7	11111000	8	8/8	1111111110111011	16
0/8	1111110110	10	8/9	1111111110111100	16
0/9	1111111110000010	16	8/A	1111111110111101	16
0/A	1111111110000011	16	9/1	111111001	9
1/1	1100	4	9/2	1111111110111110	16
1/2	11011	5	9/3	1111111110111111	16
1/3	1111001	7	9/4	1111111111000000	16
1/4	111110110	9	9/5	1111111111000001	16
1/5	11111110110	11	9/6	1111111111000010	16
1/6	1111111110000100	16	9/7	1111111111000011	16
1/7	1111111110000101	16	9/8	1111111111000100	16
1/8	1111111110000110	16	9/9	1111111111000101	16
1/9	1111111110000111	16	9/A	1111111111000110	16
1/A	1111111110001000	16	A/1	111111010	9
2/1	11100	5	A/2	1111111111000111	16
2/2	11111001	8	A/3	1111111111001000	16
2/3	1111110111	10	A/4	1111111111001001	16
2/4	111111110100	12	A/5	1111111111001010	16
2/5	1111111110001001	16	A/6	1111111111001011	16
2/6	1111111110001010	16	A/7	1111111111001100	16
2/7	1111111110001011	16	A/8	1111111111001101	16

Tabela A.3: Códigos JPEG AC.

Zeros/ Categoria	Código base	Tamanho	Zeros/ Categoria	Código base	Tamanho
2/8	1111111110001100	16	A/9	1111111111001110	16
2/9	1111111110001101	16	A/A	1111111111001111	16
2/A	1111111110001110	16	B/1	1111111001	10
3/1	111010	6	B/2	1111111111010000	16
3/2	111110111	9	B/3	1111111111010001	16
3/3	111111110101	12	B/4	1111111111010010	16
3/4	1111111110001111	16	B/5	1111111111010011	16
3/5	1111111110010000	16	B/6	1111111111010100	16
3/6	1111111110010001	16	B/7	1111111111010101	16
3/7	1111111110010010	16	B/8	1111111111010110	16
3/8	1111111110010011	16	B/9	1111111111010111	16
3/9	1111111110010100	16	B/A	1111111111011000	16
3/A	1111111110010101	16	C/1	1111111010	10
4/1	111011	6	C/2	1111111111011001	16
4/2	1111111000	10	C/3	1111111111011010	16
4/3	1111111110010110	16	C/4	1111111111011011	16
4/4	1111111110010111	16	C/5	1111111111011100	16
4/5	1111111110011000	16	C/6	1111111111011101	16
4/6	1111111110011001	16	C/7	1111111111011110	16
4/7	1111111110011010	16	C/8	1111111111011111	16
4/8	1111111110011011	16	C/9	1111111111100000	16
4/9	1111111110011100	16	C/A	1111111111100001	16
4/A	1111111110011101	16	D/1	11111111000	10
5/1	1111010	7	D/2	1111111111100010	16
5/2	11111110111	11	D/3	1111111111100011	16
5/3	1111111110011110	16	D/4	1111111111100100	16
5/4	1111111110011111	16	D/5	1111111111100101	16
5/5	1111111110100000	16	D/6	1111111111100110	16
5/6	1111111110100001	16	D/7	1111111111100111	16

Tabela A.4: Códigos JPEG AC (continuação.)

Zeros/ Categoria	Código base	Tamanho	Zeros/ Categoria	Código base	Tamanho
5/7	1111111110100010	16	D/8	1111111111101000	16
5/8	1111111110100011	16	D/9	1111111111101001	16
5/9	1111111110100100	16	D/A	1111111111101010	16
5/A	1111111110100101	16	E/1	1111111111101011	11
6/1	1111011	7	E/2	1111111111101100	16
6/2	111111110110	12	E/3	1111111111101101	16
6/3	1111111110100110	16	E/4	1111111111101110	16
6/4	1111111110100111	16	E/5	1111111111101111	16
6/5	1111111110101000	16	E/6	1111111111100000	16
6/6	1111111110101001	16	E/7	1111111111100001	16
6/7	1111111110101010	16	E/8	1111111111100010	16
6/8	1111111110101011	16	E/9	1111111111100011	16
6/9	1111111110101100	16	E/A	1111111111101000	16
6/A	1111111110101101	16	F/0	11111111001	11
7/1	11111010	8	F/1	1111111111101010	16
7/2	111111110111	12	F/2	1111111111101010	16
7/3	1111111110101110	16	F/3	1111111111101011	16
7/4	1111111110101111	16	F/4	1111111111110000	16
7/5	1111111110110000	16	F/5	1111111111110001	16
7/6	1111111110110001	16	F/6	1111111111110010	16
7/7	1111111110110010	16	F/7	1111111111110011	16
7/8	1111111110110011	16	F/8	1111111111111000	16
7/9	1111111110110100	16	F/9	1111111111111001	16
7/A	1111111110110101	16	F/A	1111111111111010	16

Tabela A.5: Códigos JPEG AC (continuação).

Apêndice B

Tabelas de codificação para compressão MPEG-1

De maneira similar a codificação Huffman utilizada para codificar as imagens processadas, também há uma tabela prédefinida para a codificação dos vetores de deslocamento a fim de otimizar o processo de compressão.

Para a codificação das informações de movimento dos macroblocos é necessário calcular a diferença entre os vetores de deslocamento dos macroblocos atual e anterior, de forma que se o atual for o primeiro de uma determinada linha o vetor de deslocamento do anterior será considerado nulo $(0, 0)$. Por fim, essa diferença será representada pelo código correspondente na tabela B.1.

MVD	Código	MVD	Código
-16 & 16	00000011001	1	010
-15 & 17	00000011011	2 & -30	0010
-14 & 18	00000011101	3 & -29	00010
-13 & 19	00000011111	4 & -28	0000110
-12 & 20	00000100001	5 & -27	00001010
-11 & 21	00000100011	6 & -26	00001000
-10 & 22	0000010011	7 & -25	00000110
-9 & 23	0000010101	8 & -24	0000010110
-8 & 24	0000010111	9 & -23	0000010100
-7 & 25	00000111	10 & -22	0000010010
-6 & 26	00001001	11 & -21	00000100010
-5 & 27	00001011	12 & -20	00000100000
-4 & 28	0000111	13 & -19	00000011110
-3 & 29	00011	14 & -18	00000011100
-2 & 30	0011	15 & -17	00000011010
-1	011	N/A	N/A
0	1	N/A	N/A

Tabela B.1: Códigos de tamanho variável para os vetores de deslocamento.

Referências Bibliográficas

- [1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- [2] D. H. Kelly, “Motion and vision. ii. stabilized spatio-temporal threshold surface,” *J. Opt. Soc. Am.*, vol. 69, no. 10, pp. 1340–1349, Oct 1979. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=josa-69-10-1340>
- [3] A. Redondi, M. Cesana, and M. Tagliasacchi, “Low bitrate coding schemes for local image descriptors,” 2012, pp. 124–129, cited By 6. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84870605750&partnerID=40&md5=36a3c310d83eb80c36005d94ec72ffaa>
- [4] C. Singh, N. Singh, and R. Tripathi, “Optimization of standards for video compression tools over wireless networks,” 2012, pp. 114–118, cited By 0. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84862111611&partnerID=40&md5=c04186a826548dc0706b41c12308327f>
- [5] H. Song and C.-C. Jay Kuo, “Rate control for low-bit-rate video via variable-encoding frame rates,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 4, pp. 512–521, 2001, cited By 88. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-0035307517&partnerID=40&md5=b2d06ea8d0e2cac993c70f68745c1039>
- [6] B. Vizzotto, B. Zatt, M. Shafique, S. Bampi, and J. Henkel, “A model predictive controller for frame-level rate control in multiview video coding,” 2012, pp. 485–490, cited By 1. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84868110949&partnerID=40&md5=1ef2b60451ea8a334c78401dca4ddff0>

- [7] “Em 5 anos, o planeta terá mais dispositivos móveis ligados à Internet do que habitantes.” [Online]. Available: <http://startupi.com.br/2013/02/em-5-anos-o-planeta-tera-mais-dispositivos-moveis-ligados-a-internet-do-que-pessoas/>
- [8] K. Seshadrinathan and A. C. Bovik, “Motion-based perceptual quality assessment of video,” *Proc. SPIE*, vol. 7240, pp. 72 400X–72 400X–12, 2009. [Online]. Available: <http://dx.doi.org/10.1117/12.811817>
- [9] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, “Objective video quality assessment methods: A classification, review, and performance comparison,” *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 165–182, 2011.
- [10] J. Joskowicz, J.-C. López-Ardao, M. A. González Ortega, and C. L. García, *A Mathematical Model for Evaluating the Perceptual Quality of Video*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 164–175. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02472-6_15
- [11] J. Li, J. Koivusaari, J. Takala, M. Gabbouj, and H. Chen, “Human visual system based adaptive inter quantization,” 2008.
- [12] C. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [13] D. A. Huffman, “A method for the construction of minimum-redundancy codes,” *Proceedings of the Institute of Radio Engineers*, vol. 40, no. 9, pp. 1098–1101, September 1952.
- [14] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, 2nd ed. Norwell, MA, USA: Kluwer Academic Publishers, 1997.
- [15] R. A. Young, “Oh say, can you see? the physiology of vision,” *Proc. SPIE*, vol. 1453, pp. 92–123, 1991. [Online]. Available: <http://dx.doi.org/10.1117/12.44348>
- [16] I. Telegraph and T. C. Committee, *CCITT Recommendation T.81: Terminal Equipment and Protocols for Telematic Services : Information Technology - Digital Compression and Coding of Continuous-tone Still Images - Requirements and Guidelines*. International Telecommunication Union, 1993. [Online]. Available: <https://books.google.com.br/books?id=XCcXHwAACAAJ>

- [17] T. S. S. of ITU., *ITU-T Recommendation H.261: Line Transmission of Non-Telephone Signals : Video Codec for Audiovisual Services at P X 64 Kbits*. International Telecommunication Union, 1993. [Online]. Available: <https://books.google.com.br/books?id=WG3jHgAACAAJ>
- [18] J. Bialkowski, M. Barkowsky, and A. Kaup, "Fast video transcoding from h.263 to h.264/mpeg-4 avc," *Multimedia Tools Appl.*, vol. 35, no. 2, pp. 127–146, Nov. 2007. [Online]. Available: <http://dx.doi.org/10.1007/s11042-007-0126-7>
- [19] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, July 2003.
- [20] K. Cabeen and P. Gent, "Image compression and the discrete cosine transform," *College of the Redwoods*, 1998.
- [21] T. S. S. of ITU., *ITU-T Recommendation H.261: Line Transmission of Non-Telephone Signals : Video Codec for Audiovisual Services at P X 64 Kbits*. International Telecommunication Union, 1993. [Online]. Available: <https://books.google.com.br/books?id=WG3jHgAACAAJ>
- [22] C. Poynton, "Chroma subsampling notation," *Retrieved June*, vol. 19, p. 2004, 2002.
- [23] M. Ghanbari, *Standard codecs: Image compression to advanced video coding*. Iet, 2003, no. 49.
- [24] A. Banitalebi, S. Nader-Esfahani, and A. N. Avanaki, "A perceptual based motion compensation technique for video coding," in *2010 6th Iranian Conference on Machine Vision and Image Processing*, Oct 2010, pp. 1–5.
- [25] Z. Wang and A. Bovik, *Modern Image Quality Assessment*, ser. Synthesis lectures on image. Morgan & Claypool Publishers, 2006. [Online]. Available: <https://books.google.com.br/books?id=F6lYVwyZJz4C>
- [26] "PythonBrasil - PythonBrasil," 00000. [Online]. Available: <http://wiki.python.org.br/>
- [27] "Xiph.org :: Derf's Test Media Collection," 00002. [Online]. Available: <https://media.xiph.org/video/derf/>