

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM

INSTITUTO DE COMPUTAÇÃO - ICOMP

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA - PPGI

PERCEPTUAL ERROR LOGARITHM (PEL): UM EFICIENTE E EFICAZ MÉTODO ANALÍTICO PARA AVALIAÇÃO DA QUALIDADE DE IMAGEM COM REFERÊNCIA COMPLETA

Sergio Augusto Coelho Bezerra

Manaus - AM
Abril de 2026

Sergio Augusto Coelho Bezerra

PERCEPTUAL ERROR LOGARITHM (PEL): UM EFICIENTE E EFICAZ MÉTODO ANALÍTICO PARA AVALIAÇÃO DA QUALIDADE DE IMAGEM COM REFERÊNCIA COMPLETA

Tese apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Amazonas como requisito para a obtenção do título de Doutor em Informática.

Orientador: José Reginaldo Hughes Carvalho, Dr.

Universidade Federal do Amazonas - UFAM
Instituto de Computação - IComp

Manaus - AM
Abril de 2026

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

B574p Bezerra, Sergio Augusto Coelho
Perceptual error logarithm (PEL): um eficiente e eficaz método analítico para avaliação da qualidade de imagem com referência completa / Sergio Augusto Coelho Bezerra. - 2026.
130 f. : il., color. ; 31 cm.

Orientador(a): José Reginaldo Hughes Carvalho.
Tese (doutorado) - Universidade Federal do Amazonas, Programa de Pós-Graduação em Informática, Manaus, 2026.

1. método analítico com referência completa. 2. avaliação da qualidade de imagem. 3. similaridade de superpíxeis. 4. sistema visual humano. 5. magnitude do gradiente. I. Carvalho, José Reginaldo Hughes. II. Universidade Federal do Amazonas. Programa de Pós-Graduação em Informática. III. Título

---

Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"PERCEPTUAL ERROR LOGARITHM (PEL): UM EFICIENTE E EFICAZ MÉTODO ANALÍTICO PARA AVALIAÇÃO DA QUALIDADE DE IMAGEM COM REFERÊNCIA COMPLETA"

SERGIO AUGUSTO COELHO BEZERRA

Tese de Doutorado defendida e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. José Reginaldo Hughes Carvalho - **Presidente**

Prof. Dr. João Marcos Bastos Cavalcanti - **Membro Externo**

Prof. Dr. José Luiz de Souza Pio - **Membro Externo**

Prof.ª Dr.ª Keiko Verônica Ono Fonseca - **Membro Externo**

gov.br Documento assinado digitalmente
ALEXANDRE DE ALMEIDA PRADO POHL
Data: 07/05/2026 08:59:44-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Alexandre de Almeida Prado Pohl - **Membro Externo**

Manaus, 17 de Abril de 2026.

Documento assinado eletronicamente por **João Marcos Bastos Cavalcanti**, **Professor do Magistério Superior**, em 22/04/2026, às 11:32, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015 .

Documento assinado eletronicamente por **Jose Luiz de Souza Pio**, **Professor do Magistério Superior**, em 22/04/2026, às 12:51, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015 .

Documento assinado eletronicamente por **Keiko Verônica Ono Fonseca**, **Usuário Externo**, em 26/04/2026, às 08:50, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **José Reginaldo Hughes Carvalho**, **Professor do Magistério Superior**, em 29/04/2026, às 15:28, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015 .

A autenticidade deste documento pode ser conferida no site https://sei.ufam.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3117396** e o código CRC **37A361BB**.

Avenida General Rodrigo Octávio, 6200 - Bairro Coroado I Campus Universitário Senador Arthur Virgílio Filho, Setor Norte - Telefone: (92) 3305-1181 / Ramal 1193
CEP 69080-900, Manaus/AM, coordenadorppgi@icomp.ufam.edu.br

Referência: Processo nº 23105.015194/2026-12 SEI nº 3117396

Com amor, eu dedico esta tese à minha esposa Erika Bezerra, minha mãe Graça Coelho, meu pai Sebastião Bezerra de Oliveira (em memória), minha filha Beatriz, meus filhos Sérgio Júnior e Guilherme, minhas avós Venina Coelho (em memória) e Maria de Oliveira (em memória), minha sogra Valdomira do Amaral, meus irmãos e irmãs, bem como aos demais parentes e amigos.

Agradecimentos

A Deus pela sua infinita misericórdia.

À Universidade Federal do Amazonas (UFAM), ao Instituto de Computação (Icomp), ao Instituto Federal de Educação, Ciência e Tecnologia do Amazonas (IFAM), ao Governo Federal (MEC - CAPES), bem como à Fundação de Amparo à Pesquisa do Amazonas (FAPEAM) pelo fundamental apoio.

Ao professor José Reginaldo Hughes Carvalho pela sabedoria, inteligência e visão quanto à orientação deste trabalho.

À professora Dra. Keiko Verônica Ono Fonseca e ao professor Dr. Alexandre de Almeida Prado Pohl do CPGEI da UTFPR, bem como aos professores Dr. José Luiz de Souza Pio e Dr. João Marcos Bastos Cavalcanti do Icomp pelas contribuições científicas.

Aos professores e servidores administrativos do Icomp da UFAM.

A todos os professores e administrativos do IFAM pelo apoio.

A todos os colegas do laboratório pelo apoio incondicional.

A minha esposa, Erika Bezerra, pelas orações e responsabilidades assumidas durante o meu doutoramento.

E, finalmente, quero render meus profundos agradecimentos à sociedade pelo sacrifício quanto aos investimentos empregados para que esta pesquisa fosse realizada.

PERCEPTUAL ERROR LOGARITHM (PEL): UM EFICIENTE E EFICAZ MÉTODO ANALÍTICO PARA AVALIAÇÃO DA QUALIDADE DE IMAGEM COM REFERÊNCIA COMPLETA

Autor: Sergio Augusto Coelho Bezerra

Orientador: José Reginaldo Hughes Carvalho, Dr.

Resumo

A crescente popularidade de aplicações multimídia tem fomentado investimentos em tecnologias e serviços para garantir o processamento, a transmissão e o armazenamento de grandes volumes de dados, notadamente imagens e vídeos de alta qualidade. Essa garantia exigida por tais aplicações multimídia é comumente subsidiada por técnicas, tais como super-resolução (*super-resolution*, SR), compressão, restauração, recuperação e detecção de objetos em imagem e vídeo. Embora tais técnicas tenham recebido atenção considerável das comunidades acadêmica e industrial nos últimos anos, seus processos frequentemente introduzem artefatos que degradam significativamente a qualidade de imagens e vídeos. Para mitigar tais degradações, os pesquisadores vêm utilizando métodos perceptuais para avaliação da qualidade de imagem (*image quality assessment*, IQA). Contudo, é importante ressaltar que um dos grandes desafios existentes na atualidade é o desenvolvimento de métodos perceptuais que sejam simultaneamente eficientes e eficazes. Neste sentido, esta tese propõe um método analítico perceptual eficiente e eficaz para avaliação da qualidade de imagens com referência completa (*full-reference image quality assessment*, FR-IQA), denominado Logaritmo do Erro Perceptual (*Perceptual Error Logarithm*, PEL). O método PEL cria mapas de luminância e crominância, que são gerados após a redução e a conversão espacial de cores aplicadas às imagens de entrada. Por meio do uso de mapas de crominância e luminância, são gerados mapas simples de diferença absoluta, bem como mapas de características especiais. Os mapas especiais são derivados de extrações de características perceptuais aplicadas a mapas de magnitude de gradiente. Esses mapas são gerados imediatamente após a aplicação da filtragem por desvio padrão local aos mapas de luminância. Por fim, para calcular a pontuação de PEL, uma função logarítmica é aplicada ao mapa de erro perceptual, que combina os mapas simples e especiais, além de ajustar essa combinação com base no peso dos mapas de similaridade de superpíxeis e de energia local. Resultados experimentais e discussões em oito bases de dados de imagens comprovam que a eficácia do método proposto, conforme a aplicação de métricas de avaliação estatísticas como Spearman Rank-Order Correlation Coefficient (SROCC), Kendall Rank-Order Correlation Coefficient (KROCC), Pearson Linear Correlation Coefficient (PLCC) e Root Mean Square Error (RMSE), é muito mais consistente com avaliações subjetivas do que todos os dezenove métodos de IQA da literatura comparados. Além disso, o método proposto mostra-se

eficiente em razão do seu baixo custo computacional necessário para realizar o processamento médio sobre todas as imagens. Sendo assim, PEL pode ser utilizado como método para FR-IQA, ou mesmo como função de perda (*loss function*), a fim de subsidiar diversas técnicas analíticas utilizadas em aplicações multimídia ou modelos baseados em aprendizado, como redes neurais convolucionais (*Convolutional Neural Networks, CNNs*) e redes adversariais generativas (Generative Adversarial Networks, GANs), voltados à imagem e vídeo.

Palavras-chave: método analítico com referência completa, avaliação da qualidade de imagem, similaridade de superpíxeis, sistema visual humano, magnitude do gradiente, modelo baseado em aprendizado.

PERCEPTUAL ERROR LOGARITHM (PEL): UM EFICIENTE E EFICAZ MÉTODO ANALÍTICO PARA AVALIAÇÃO DA QUALIDADE DE IMAGEM COM REFERÊNCIA COMPLETA

Autor: Sergio Augusto Coelho Bezerra

Orientador: José Reginaldo Hughes Carvalho, Dr.

Abstract

The growing popularity of multimedia applications has fostered investments in technologies and services to ensure the processing, transmission, and storage of large volumes of data, notably high-quality images and videos. This guarantee required by such multimedia applications is commonly supported by techniques such as super-resolution (SR), compression, restoration, recovery, and object detection in images and videos. Although these techniques have received considerable attention from the academic and industrial communities recently, their processes often introduce artifacts that significantly degrade the quality of images and videos. To mitigate such degradations, researchers have been using perceptual methods for image quality assessment (IQA). However, it is important to emphasize that one of the great challenges currently existing is the development of perceptual methods that are simultaneously efficient and effective. In this sense, this thesis proposes an efficient and effective perceptual analytical method for evaluating the quality of full-reference images (FR-IQA), called the Perceptual Error Logarithm (PEL). Initially, the PEL method creates luminance and chrominance maps, which are generated after the reduction and spatial conversion of colors applied to the input images. The PEL method, through the use of chrominance and luminance maps, generates simple absolute difference maps, as well as special feature maps. The special maps are derived from perceptual feature extractions applied to gradient magnitude maps. These maps are generated immediately after the application of local standard deviation filtering to the luminance maps. Finally, to calculate the PEL score, a logarithmic function is applied to the perceptual error map, which combines the simple and special maps, in addition to adjusting this combination based on the weight of the superpixel similarity and local energy maps. Experimental results and discussions in eight image databases prove that the effectiveness of the proposed method, according to the application of statistical evaluation metrics such as Spearman Rank-Order Correlation Coefficient (SROCC), Kendall Rank-Order Correlation Coefficient (KROCC), Pearson Linear Correlation Coefficient (PLCC), and Root Mean Square Error (RMSE), is much more consistent with subjective evaluations than all nineteen IQA methods in the literature compared. Furthermore, the proposed method turns out to be efficient due to its low computational cost required to perform the average processing on all images. Thereby, PEL can be used as a method for FR-IQA or even as a loss function to support various analytical

techniques used in multimedia applications or machine learning-based models, such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), focused on image and video.

Keywords: full-reference analytical method, image quality assessment, superpixel similarity, human visual system, gradient magnitude, learning-based model.

LISTA DE ILUSTRAÇÕES

Figura 1 – Seção transversal do olho humano 29
Figura 2 – Espectro eletromagnético visível ao olho humano 30
Figura 3 – Ilusões visuais devido à influência do contraste. 31
Figura 4 – Exemplos de artefatos em imagem. 33
Figura 5 – Exemplo de ambiente para avaliação subjetiva. 35
Figura 6 – Processo computacional do método Structural SIMilarity (SSIM). 39
Figura 7 – Processo computacional do método non-Negative Matrix Factorization (NMF). 40
Figura 8 – Processo computacional do método Perceptual Feature Similarity Index (PFSI). 41
Figura 9 – Processo computacional do método Deep Image QuAlity Measure para FR IQA (DIQaM-FR). 42
Figura 10 – Processo computacional do método Dynamic Receptive Field generation based Image Quality Assessor (DRF-IQA). 43
Figura 11 – Método com Referência Completa. 45
Figura 12 – Método Sem referência. 46
Figura 13 – Método com Referência Reduzida. 46
Figura 14 – Exemplos de Região de Interesse (RoI). 47
Figura 15 – Processo computacional do método Feature-SIMmilarity ($FSIM_C$). 49
Figura 16 – Processo computacional do método Gradient SIMilarity (GSM). 50
Figura 17 – Processo computacional do método Gradient Magnitude Similarity Deviation (GMSD). 50
Figura 18 – Processo computacional do método Multiscale Contrast Similarity Deviation (MCSD) 50
Figura 19 – rocesso computacional do método Contrast and Visual Saliency Similarity-induced index (CVSS). 51
Figura 20 – Processo computacional do método Visual Saliency-based Index (VSI). 52
Figura 21 – Processo computacional do método Perceptual SIMilarity (PSIM). 52
Figura 22 – Processo computacional do método Structural Contrast-Quality Index (SC-QI). 53
Figura 23 – Processo computacional do método Riesz transform and Visual contrast sensitivity-based feature SIMilarity (RVSIM). 54

Figura 24 – Processo computacional do método Sparseness Significance Ranking Measure (SSRM). 54
Figura 25 – Processo computacional do método Edge-Feature-based image segmentation (EFS). . 56
Figura 26 – Processo computacional do método Superpixel-based Similarity Index (SPSIM). . . . 57
Figura 27 – Processo computacional do método Structural Singular Value Decomposition (SSVD). 57
Figura 28 – Processo computacional do método Visual saliency with Color appearance, Gradient Similarity (VCGS). 58
Figura 29 – Processo computacional do método Diffusion Speed Structury Similarity (DSSIM). . . 59
Figura 30 – Processo computacional do método Synergizing Superpixels and Visual Saliency (SSVS). 60
Figura 31 – Processo computacional do método Spectral Understanding of Multi-scale and Multi-channel Error Representations (SUMMER). 61
Figura 32 – Processo Metodológico. 63
Figura 33 – Imagens originais da BDI IVC. 66
Figura 34 – Imagens originais da BDI LIVE. 66
Figura 35 – Imagens originais da BDI CSIQ. 67
Figura 36 – Imagens originais da BDI MICT. 68
Figura 37 – Imagens originais da BDI VCL. 69
Figura 38 – Imagens originais da BDI TID2013. 69
Figura 39 – Imagens originais da BDI CIDIQ. 70
Figura 40 – Amostra das imagens originais da BDI KADID-10K. 71
Figura 41 – Exemplo de formato de gráfico de dispersão (*scatter*) voltado à visão específica quanto ao entendimento da relação entre *scores* objetivos do método SSIM e *scores* subjetivos da BDI LIVE. 73
Figura 42 – Exemplo de *box plot* (gráfico de caixa). 74
Figura 43 – Exemplo de formato de gráfico de dispersão (*scatter*) utilizado em análise específica de Análise de Desempenho versus Custo Computacional. 76
Figura 44 – Fluxograma do processo computacional para o método analítico perceptual proposto voltado à FR-IQA: Perceptual Error Logarithm (PEL). 78
Figura 45 – Exemplo da aplicação da etapa LSD no processo do método PEL. 81
Figura 46 – Exemplos de Diferença Absoluta (AD) aplicada a mapas de características de cromi-nância, luminância e gradiente em imagens da BDI LIVE 85
Figura 47 – Exemplos de mapas resultantes das características que contribuem para a formação do mapa final de Erro Perceptual (EP) do método proposto. 86
Figura 48 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens IVC. 94
Figura 49 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens LIVE. 95

Figura 50 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens CSIQ. 96
Figura 51 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens MICT. 97
Figura 52 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens VCL. 98
Figura 53 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens TID13. 99
Figura 54 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens CIDIQ. 100
Figura 55 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens KADID. 101
Figura 56 – Box plots (Gráficos de caixa) de DMOS subjetivo versus pontuações obtidas por métodos FR-IQA para todas as distorções nas oito bases de dados de imagens (BDIs). . . . 102
Figura 57 – Comparação de Desempenho sobre os Tipos Individuais de Distorções da BDI TID2013 109
Figura 58 – Comparações dos 5 principais métodos FR-IQA em termos de desempenho (SROCC e KROCC geral) e custo computacional (fps). 112

LISTA DE TABELAS

Tabela 1 – Escala Subjetiva (Recomendação da CCIR 500-3). 36
Tabela 2 – Seleção de Métodos Subjetivos. 37
Tabela 3 – Comparação das oito bases de dados de imagens para IQA utilizadas. 64
Tabela 4 – Tipos de Distorções em cada BDI 65
Tabela 5 – Tipos de Distorções em cada BDI. 80
Tabela 6 – Comparações do custo médio de computação em milissegundos/imagem e velocidade de execução (*Running speed*) em quadros por segundo (*Frames Per Second*, fps) para vinte métodos FR-IQA em oito bancos de imagens. 92
Tabela 7 – Comparação de métricas de avaliação de desempenho para métodos IQA em oito BDI. 106
Tabela 8 – Comparação de avaliação de desempenho conforme a métrica PLCC para os métodos IQA Top-5 na BDI TID2013. 108
Tabela 9 – Estatística $F_{\%}$ usada para a significância da diferença entre $RMSE_{MAX}/RMSE_{MIN}$ dos IQMs sobre as oito bases de dados de imagens (BDIs) utilizadas. 111

NOMENCLATURA

BDI	Base de Dados de Imagens
CNN	*Convolutional Neural Network*
DL	*Deep Learning*
$DMOS$	*Difference Mean Opinion Score*
DNN	*Deep Neural Network*
FR	*Full-Referenc*
GAN	*Generative Adversarial Networks*
GM	*Gradient Magnitude*
$HDTV$	*High Definition Television*
HR	*High Resolution*
HVS	*Human Visual System*
IP	*Internet Protocol*
IQA	*Image Quality Assessment*
IQM	*Image Quality Method*
$ITU-T$	*International Telecommunication Union - Telecommunication Standardization Sector*
JND	*Just Noticeable Difference*
$JPEG$	*Joint Photographic Experts Group*
$KROCC$	*Kendall Rank-Order Correlation Coefficient*
MOS	*Mean Opinion Score*
$MPEG$	*Moving Picture Experts Group*

$MS-SSIM$	*Multi-Scale SSIM*
MSE	*Mean Squared Error*
NR	*No-Referenc,* ou *blind*
PC	*Phase Congruency*
PEL	*Perceptual Error Logarithm*
$PLCC$	*Pearson Linear Correlation Coefficient*
$PSNR$	*Peak Signal-to-Noise Ratio*
$RMSE$	*Root Mean Square Error*
RoI	*Region of Interest*
RR	*Reduced-Reference*
SR	*Super-Resolution*
$SROCC$	*Spearman Rank-Order Correlation Coefficient*
$SSIM$	*Structural SIMilarity*
VIF	*Visual Information Fidelity*
VQA	*Video Quality Assessment*
VQM	*Video Quality Method*
VS	*Visual Saliency*

SUMÁRIO

1 INTRODUÇÃO 20
1.1 Contextualização e Justificativa 20
1.2 Definição do Problema 24
1.3 Objetivos 24
1.3.1 Objetivo Geral 24
1.3.2 Objetivos Específicos 24
1.4 Processo Metodológico 25
1.5 Contribuições 25
1.6 Publicação Científica 26
1.7 Estrutura da Tese 26

2 FUNDAMENTAÇÃO TEÓRICA 28
2.1 Sistema Visual Humano 28
2.1.1 Propriedades da Luz 29
2.1.2 Medida da Intensidade e Sensibilidade de Contraste 30
2.2 Avaliação da Qualidade de Imagem (IQA) 32
2.2.1 Artefatos 32
2.2.1.1 *Blockiness* (Blocagem ou distorção de bloco) 32
2.2.1.2 *Blurring* (Borramento) 33
2.2.1.3 *Ringing* 34
2.2.1.4 *Staircase Effect* (Efeito Escada) 34
2.2.2 Métodos Subjetivos 34
2.2.2.1 Seleção dos Materiais e Métodos 34
2.2.2.2 Planejamento dos Procedimentos da Avaliação 36
2.2.2.3 Análises Estatísticas e Discussões dos Resultados 37
2.2.3 Métodos Objetivos 38
2.2.3.1 Métodos Objetivos Analíticos baseados em Píxeis 38
2.2.3.2 Métodos Objetivos Perceptuais Analíticos 39
2.2.3.3 Métodos Objetivos Perceptuais baseados em Aprendizagem 40
2.2.4 Métricas Estatísticas de Predição da Avaliação 43

2.2.5 Classificação de Métodos quanto a sua Referência . . . 45
2.2.5.1 Método *Full-Reference* (FR) . . . 45
2.2.5.2 Método *No-Reference* (NR) . . . 45
2.2.5.3 Método *Reduced-Reference* (RR) . . . 46
2.3 Região de Interesse (RoI) . . . 47

3 TRABALHOS RELACIONADOS . . . 48
3.1 Introdução . . . 48
3.2 Visão Geral dos Métodos Comparados . . . 48

4 PROCESSO METODOLÓGICO . . . 62
4.1 Introdução . . . 62
4.2 Materiais . . . 64
4.2.1 Bases de Dados de Imagens (BDIs) . . . 64
4.2.1.1 BDI: IVC . . . 65
4.2.1.2 BDI: LIVE . . . 66
4.2.1.3 BDI: CSIQ . . . 67
4.2.1.4 BDI: MICT . . . 68
4.2.1.5 BDI: VCL . . . 68
4.2.1.6 BDI: TID2013 . . . 69
4.2.1.7 BDI: CIDIQ . . . 70
4.2.1.8 BDI: KADID . . . 70
4.2.2 Métodos Objetivos Comparados . . . 71
4.3 Verificação das Habilidades Perceptuais dos Métodos . . . 71
4.3.1 Custo Computacional . . . 72
4.3.2 Análise de Desempenho . . . 72
4.3.2.1 Valores DMOS/MOS e Mapeamento para a Escala Subjetiva . . . 72
4.3.2.2 Métricas Estatísticas de Predição da Avaliação . . . 74
4.3.2.3 Significância Estatística . . . 75
4.3.3 Custo Computacional versus Análise de Desempenho . . . 76

5 MÉTODO ANALÍTICO PERCEPTUAL PROPOSTO . . . 77
5.1 Visão Geral . . . 77
5.2 Processo Computacional . . . 78
5.2.1 Pré-processamento . . . 78
5.2.2 Desvio Padrão Local . . . 79
5.2.3 Magnitude do Gradiente . . . 79
5.2.4 Diferença Absoluta (Crominância, Luminância e Gradiente) . . . 83

5.2.5 Energia Local e Constrate 84
5.2.6 Similaridade (Gradiente e Crominância) 87
5.2.7 Similaridade de Super-píxeis 88
5.2.8 Cálculo da Avaliação da Qualidade 89

6 EXPERIMENTOS, ANÁLISES E DISCUSSÕES 91
6.1 Visão Geral 91
6.2 Custo Computacional 91
6.3 Análise de Desempenho 93
6.3.1 Comparação entre Valores Subjetivos e Objetivos Normalizados 93
6.3.1.1 Visão Específica 93
6.3.1.2 Visão Geral 102
6.3.2 Predição da Avaliação Estatística dos Métodos 105
6.3.2.1 Comparação de Desempenho sobre as BDIs 105
6.3.2.2 Comparação de Desempenho sobre os Tipos Individuais de Distorções da BDI TID2013 107
6.3.3 Significância Estatística dos Métodos 110
6.4 Custo Computacional versus Análise de Desempenho 110

7 CONCLUSÃO E TRABALHOS FUTUROS 113
7.1 Trabalhos Futuros 115

Referências 117

1

INTRODUÇÃO

ESTE capítulo apresenta contextualizações, justificativas, definição do problema e objetivos quanto à necessidade do desenvolvimento de novos métodos analíticos perceptuais para avaliação da qualidade de imagens. Além disso, descreve o processo metodológico concebido para nortear esta pesquisa, a produção científica alcançada, as contribuições desta tese e, por fim, sua estrutura organizacional.

1.1 Contextualização e Justificativa

A crescente popularidade de aplicações multimídia tem fomentado investimentos em tecnologias e serviços para garantir o processamento, a transmissão e o armazenamento de grandes volumes de dados, notadamente imagens e vídeos de alta qualidade, ou seja, de alta resolução (*high resolution*, HR[1]). Essa garantia exigida por tais aplicações multimídia é comumente subsidiada por técnicas, tais como super-resolução (*super-resolution*, SR), compressão, restauração, recuperação e detecção de objetos em imagem e vídeo [2].

Embora tais técnicas tenham recebido atenção considerável das comunidades científicas acadêmica e industrial nos últimos anos [3], [4], seus processos frequentemente introduzem artefatos [5], como, por exemplo, efeitos de blocagem, ruído, *ringing*, bordas falsas e borramento, que degradam significativamente a qualidade de imagens e vídeos. Para mitigar tais distorções, os pesquisadores vêm utilizando métodos de avaliação da qualidade de imagem (image quality assessment - IQA) [6].

Em geral, a IQA é classificada em métodos subjetivos e objetivos [7], [8]. O método subjetivo usa um número mínimo de participantes em seu procedimento para avaliar a qualidade da imagem e do vídeo [9]. O Mean Opinion Score (MOS) ou Difference Mean Opinion Score (DMOS) representam uma medida estatística baseada nas opiniões dos participantes, oferecendo uma avaliação quantitativa da qualidade subjetiva [8]. Esses resultados de avaliações subjetivas servem como referência para validar

[1] HR significa que a densidade de píxeis em uma imagem, ou mesmo vídeo, é alta e, portanto, seu conteúdo pode oferecer mais detalhes [1].

métodos objetivos de avaliação da qualidade [10], [11]. Embora valioso, devido ao alto custo associado à manutenção de um grupo ativo de observadores humanos para realizar os testes, o uso de métodos subjetivos para corroborar como a garantia de tais aplicações multimídia não é prático [12], [13]. Portanto, métodos objetivos[2], conhecidos como métodos de qualidade de imagem (*Image Quality Methods*, IQMs) ou métodos de qualidade de vídeo (*Video Quality Methods*, VQMs), foram criados com o propósito de avaliar e relatar automaticamente a qualidade visual percebida sem o envolvimento humano [14], [15]. Neste contexto, dois pontos são considerados importantes nesta pesquisa quanto ao desenvolvimento de métodos objetivos: *habilidade perceptual* e *referência de comparação*.

Quanto à habilidade perceptual, que é o primeiro ponto, um método objetivo pode ser medido com base em sua *eficiência* e *eficácia* quanto à realização de uma avaliação. A eficiência se refere ao custo computacional, geralmente medido pelo tempo médio de execução de cada imagem, ou ainda, pela complexidade do algoritmo, sendo a primeira forma mais usual. Já a eficácia de um método está diretamente associada ao seu nível de desempenho[3], que é avaliado com relação à sua capacidade de prever a avaliação subjetiva da qualidade de uma imagem [11]. Portanto, um método eficaz deve ter uma predição de alta qualidade (ou seja, alta correlação com a pontuação subjetiva humana) [16].

O segundo ponto trata-se da referência de comparação, que dependendo da fonte de imagem ou vídeo disponível, os métodos voltados à avaliação de qualidade podem ser classificados como: referência completa (*Full-Reference*, FR), quando a fonte de referência está disponível; referência reduzida (*Reduced-Reference*, RR), quando informações parciais da fonte de referência estão disponíveis; e sem referência (*No-Reference*, NR, ou *blind*), quando a fonte de referência não está disponível [13]. A existência ou não da informação de referência é muito importante, pois isso afeta diretamente a complexidade do método a ser desenvolvido, bem como os cenários nos quais ele pode ser aplicado.

Os métodos para avaliação da qualidade de imagens podem ser divididos em dois grupos: *métodos baseados em píxeis*; e *métodos perceptuais* [17]. Os métodos baseados em píxeis, devido ao seu baixo custo computacional, são matematicamente fáceis de implementar e bastante eficientes. Embora simples e amplamente utilizados, os IQMs baseados em píxeis para a avaliação da qualidade de imagens com referência completa (*full-reference image quality assessment*, FR-IQA), como erro médio quadrático (*Mean Squared Error*, MSE), relação sinal-ruído de pico (*Peak Signal-to-Noise Ratio*, PSNR) e raiz do erro quadrático médio (*Root Mean Squared Error*, RMSE) [18], são criticados por sua baixa eficácia, [19], [20]. Os métodos perceptuais, por outro lado, consideram as características do sistema visual humano (*Human Visual System*, HVS) em uma tentativa de incorporar medidas de qualidade perceptual. Nesse sentido, vários métodos foram propostos, os quais podem ser divididos em duas categorias: *métodos analíticos*; e *métodos baseados em aprendizagem*.

Os métodos perceptuais que não usam técnicas de aprendizagem são considerados aqui como

[2] Método objetivo tem como sinônimo na literatura o termo *métrica objetiva*. Tal termo é muito comum quando se refere a processos matematicamente simples para avaliar a qualidade de imagem, como em MSE e PSNR. Por esta razão, o termo métrica é bastante utilizado na área de modelos baseados em aprendizagem.

[3] Na Seção 4.3.2.2 são descritas as métricas estatísticas de desempenho utilizadas nesta tese.

métodos analíticos perceptuais. Nesta categoria, *Structural SIMilarity* (SSIM) [21] pode ser considerada o marco da maioria dos métodos perceptuais voltados à FR-IQA. Com base na hipótese de que o HVS é altamente adaptado para perceber informações estruturadas, o SSIM pode prever a qualidade da imagem, capturando mudanças baseadas em contraste, luminância e informações estruturais entre as imagens de referência e as degradadas (ou imagens de testes). Enquanto isso, sua versão estendida, ou seja, a *Multi-scale SSIM* (MS-SSIM) [22], produz melhores previsões do que sua contraparte de escala única. O método *Visual Information Fidelity* (VIF) [23], e sua versão estendida, usam coeficientes wavelet para quantificar informações compartilhadas entre imagens de referência e distorcidas. SSIM, MS-SSIM e VIF podem oferecer desempenho estatisticamente muito melhor, entretanto tais métodos não são consistentes com a intuição de que diferentes regiões em uma imagem podem ter contribuições muito diferentes em relação à percepção da imagem pelo HVS [24]. Além disso, as versões SSIM e VIF são menos eficientes devido ao seu maior custo computacional quando comparadas ao método SSIM, bem como métodos combinados ou baseados em fusão para FR-IQA como em [25–29]. É importante destacar que os métodos baseados em píxeis também são analíticos, mas não são considerados perceptuais. Vários outros métodos analíticos FR-IQA usados neste trabalho são apresentados na Seção 3.1.

Os métodos IQA baseados em aprendizagem são uma das novas tendências para IQA perceptual. Eles visam usar técnicas de aprendizagem para deduzir uma função matemática para o método a partir de uma relação entre diferentes efeitos de distorção e a qualidade da imagem [30]. Na última década, várias abordagens para implementar técnicas de aprendizagem em métodos para IQA, visando prever a qualidade da imagem com a mesma acurácia do julgamento humano, foram desenvolvidas. As primeiras abordagens utilizaram técnicas de aprendizagem de máquina[4,5] como em [33–35]. Narwaria e Lin [35] propuseram um método baseado em decomposição de valor singular (*Singular Value Decomposition*, SVD) usando support vector regression (*support vector regression*, SVR). Com base em *Extreme Learning Machine* (ELM), Wang *et al.* [33] propuseram um novo método FR que aplica fatoração de matriz não negativa (*non-Negative Matrix Factorization*, NMF) para medir degradações de imagem. Vários outros métodos FR-IQA baseados em aprendizagem podem ser encontrados na literatura, como em [34], baseados em *Random Forest* (RF).

Mais recentemente, métodos de IQA baseados em aprendizagem têm sido desenvolvidos por meio de redes neurais profundas (*Deep Learning*, DL) como em [36–40]. DL é um subconjunto do aprendizado de máquina que usa redes neurais de várias camadas para simular o complexo poder de tomada de decisão do cérebro humano [41]. Segundo Tekalp [42], com os recentes avanços em DL é possível implementar facilmente algoritmos de processamento de sinal não linear (neural) que são otimizados para qualquer critério de avaliação perceptual desejado. Nesse sentido, em nível de IQA, Bosse *et al.* [36] apresentaram uma abordagem baseada em DL para NR e FR-IQA. Varga [43] apresenta uma nova

[4] Aprendizado de máquina é a programação de computadores para otimizar os parâmetros de um modelo usando os dados de treinamento ou a experiência passada. O modelo pode ser preditivo para fazer previsões no futuro, ou descritivo para obter conhecimento a partir dos dados, ou ambos [31].

[5] O uso de técnicas de aprendizado de máquina e inteligência artificial se expandiu rapidamente em diferentes áreas da vida, como finanças, educação, medicina, indústria manufatureira, segurança cibernética [32], inclusive, na área de IQA.

arquitetura NR-IQA baseada em aprendizado profundo que depende da fusão de decisão de pontuações de qualidade de múltiplas imagens, provenientes de diferentes tipos de redes neurais convolucionais (*Convolutional Neural Networks*, CNNs). Outros métodos FR-IQA podem ser encontrados na literatura, tendo como arquitetura uma CNN ou uma rede neural profunda (*Deep Neural Network*, DNN)[6] [37], [38], [46–51].

É importante destacar que recentes métodos baseados em aprendizagem aplicados a outros domínios têm recebido considerável atenção da comunidade científica, tais como super-resolução (*super-resolution*, SR) [6], [39], [52–54], subaquático (*underwater*) [40], [55], [56] e médico [57], [58]. Contudo, apesar desse promissor emprego de modelos baseados em DL, o grande desafio é encontrar medidas de IQA e avaliação da qualidade de vídeo (*video quality assessment*, VQA) que correlacionem bem com o julgamento humano [42]. Nesse sentido, para mitigar tal desafio, diversos métodos IQA baseados em aprendizagem foram propostos na última década, como os citados anteriormente. Contudo, Yuzhen *et al.* [59] destacam que a consistência entre tais métodos e suas avaliações subjetivas de qualidade em aplicações multimídia é questionável. Além disso, Pinson [60] pontua, ainda, que há uma grande lacuna em termos de acurácia e reprodutibilidade na indústria desses métodos, principalmente em relação aos métodos NR.

Diante do exposto, pode-se inferir que tal lacuna possa ser a razão para que métodos analíticos tradicionais em IQA, como PSNR e SSIM, ou ainda alguma combinação de métodos analíticos, continuem sendo amplamente utilizados em razão da eficiência que possuem, mesmo esses métodos apresentando baixa eficácia devido à sua pouca relação com o julgamento humano [4], [52], [61]. Nesse sentido, é importante ressaltar que, naturalmente, os processos de técnicas, modelos e arquiteturas serão consequentemente limitados conforme à eficiência e à eficácia dos métodos objetivos, bem como quaisquer investigações científicas que utilizem IQA como base de cálculo, como nas funções de perdas em redes neurais voltadas para imagens e vídeos.

Considerando essa grande lacuna quanto à necessidade de métodos objetivos mais eficientes e eficazes à IQA, esta tese propõe um método analítico perceptual eficiente e eficaz para avaliação da qualidade de imagens com referência completa, denominado Logaritmo do Erro Perceptual (*Perceptual Error Logarithm*, PEL). Desse modo, espera-se que esse novo método possa beneficiar diversas tecnologias e serviços que necessitem garantir, por meio da IQA, o processamento, a transmissão e o armazenamento de grandes volumes de dados.

[6] Uma Deep Neural Network (DNN) é um tipo de rede neural artificial que possui múltiplas camadas ocultas entre as camadas de entrada e de saída, permitindo que o modelo aprenda representações de dados em vários níveis de abstração. Ao contrário de redes neurais tradicionais com poucas camadas, as DNNs conseguem extrair características hierárquicas complexas a partir de dados brutos, sendo fundamentais para avanços em visão computacional, reconhecimento de fala e processamento de linguagem natural [44], [45].

1.2 Definição do Problema

Além das limitações anteriormente destacadas, diversas pesquisas, como em [62–70], fazem um chamamento à comunidade científica quanto à necessidade de desenvolver novos métodos objetivos para IQA. Recentemente, propostas podem ser encontradas tanto para métodos analíticos perceptuais quanto para métodos baseados em aprendizagem, mas, segundo Xue *et al.* [71] e Jia *et al.* [72], o grande desafio é desenvolver métodos objetivos para IQA que sejam mais eficientes e eficazes do que os métodos descritos na literatura estudada. Nesse sentido, esta tese investigou a seguinte problemática:

- É possível construir métodos analíticos perceptuais competitivos em relação ao estado da arte, que sejam mais eficientes e eficazes para avaliar a qualidade de imagens com referência completa?

A hipótese central desta tese sustenta que a combinação harmoniosa de características de métodos analíticos tradicionais e perceptuais viabiliza a criação de novos modelos. Tais métodos propõem-se a ser mais eficientes, eficazes e competitivos em relação ao estado da arte.

1.3 Objetivos

Os objetivos geral e específicos, que conduziram à construção do conhecimento no âmbito deste trabalho, são descritos a seguir.

1.3.1 Objetivo Geral

- Criar um novo método analítico perceptual para a avaliação da qualidade de imagens com referência completa, que seja mais eficiente e eficaz, bem como competitivo em relação ao estado da arte.

1.3.2 Objetivos Específicos

Os seguintes objetivos específicos foram considerados para que o objetivo geral deste trabalho fosse alcançado:

- Explorar características de métodos analíticos que possam ser combinadas harmoniosamente a fim de representar melhor a percepção subjetiva;
- Investigar na literatura características psicofísicas do sistema visual humano que contribuam com a modelagem estrutural do método;
- Subsidiar a criação do novo método com tais características, a fim que o mesmo possa correlacionar bem suas avaliações de qualidade de imagens com o julgamento humano.

1.4 Processo Metodológico

Os procedimentos adotados para nortear os trabalhos foram planejados a fim de subsidiar a criação de novos métodos perceptuais voltados à FR-IQA. Este trabalho baseia-se em uma pesquisa com abordagem metodológica quantitativa. A coleta de artigos não apresenta nenhum protocolo de revisão sistemática, sendo realizada de maneira exploratória.

Os passos que compõem o processo metodológico para criação do novo método analítico perceptual, proposto nesta tese, são abordados em detalhes no Capítulo 4, mas, em geral, incluem:

i) Investigação de base de dados de imagens (BDI) e métodos existentes na literatura voltados à FR-IQA;

ii) Descrição do Processo Computacional do método proposto; e

iii) Avaliação do novo método para verificar sua eficiência e eficácia competitiva em relação aos métodos do estado da arte sob BDI voltadas à FR-IQA.

1.5 Contribuições

A principal contribuição desta tese é a produção do método PEL, que apresenta um custo-benefício superior a dezenove métodos da literatura, bem como sua publicação em uma revista de extrato superior exigida pelo PPGI/UFAM, como descrito na próxima seção. Contudo, vale ressaltar que o método PEL alcançou tal capacidade perceptual devido a diversas contribuições parciais, resumidas abaixo:

- Um processo metodológico voltado para o desenvolvimento de novos métodos para IQA;
- Destaque de bordas fortes e regiões de interesse (*Region of Interest*, RoI). Em geral, os métodos de última geração [24], [73–77], que utilizam a magnitude do gradiente, geram seus mapas após a etapa inicial de pré-processamento, o que, por sua vez, acaba destacando tanto bordas fracas quanto fortes. No entanto, o HVS tende a prestar mais atenção às bordas fortes na imagem [34]. Ao adicionar uma etapa de desvio local entre o pré-processamento e a etapa de magnitude do gradiente, o método PEL foi capaz de destacar naturalmente tanto bordas fortes quanto regiões de interesse, conforme demonstrado nas Seções 5.2.3, 5.2.4 e 5.2.8;
- Importância do controle dinâmico da intensidade dos mapas de energia local formados a partir de mapas de magnitude do gradiente, bem como seu uso como peso para as etapas seguintes, conforme demonstrado nas Seções 5.2.5, 5.2.7 e 5.2.8;
- Mudança assertiva da similaridade de superpixels, conforme originalmente proposto por [75] e usado por [78], com a remoção do mapa de características de luminância, bem como a adição da

média do mapa de energia local proposto aqui como um ajuste dinâmico do mapa de similaridade de gradiente, conforme demonstrado na Seção 5.2.7;

- Criação de fatores α e β, que são pesos formados pela similaridade de superpixels e mapas de características de energia local, para ajustar dinamicamente o mapa de características gerado pelo método PEL, ou seja, o Erro Perceptual (EP), conforme demonstrado na Seção 5.2.8;
- Comparação detalhada do método proposto com dezenove métodos de última geração em oito bancos de dados de imagens usados nos experimentos, que são *benchmarks* para avaliar o desempenho de métodos de IQA;
- Apresentação dos resultados produzidos pelo método PEL em uma escala de valores de 0 a 5, correlacionados com os valores da escala DMOS;
- Apresentação simplificada dos resultados das métricas de avaliação por meio de uma classificação com números subscritos, conforme mostrado na Tabela 7; e
- Além disso, esta tese apresenta uma nova maneira de comparar os métodos em termos de significância estatística, conforme mostrado na Tabela 9.

1.6 Publicação Científica

Esta seção é destinada a registros de submissões e publicações científicas subsidiadas com base no conhecimento gerado desta pesquisa.

Como fruto deste trabalho, são especificados a seguir os dados do artigo aceito pela revista, de extrato superior, IEEE Access (ISSN: 2169-3536) [79]:

- Sergio A. C. Bezerra, Sérgio A. C. Bezerra Júnior, José L. de S. Pio, José R. H. Carvalho, and Keiko V. O. Fonseca. Perceptual error logarithm: An efficient and effective analytical method for full-reference image quality assessment. IEEE Access, vol. 13, pp. 68587-68606, 2025, doi: 10.1109/ACCESS.2025.3560918.

1.7 Estrutura da Tese

Além deste capítulo introdutório, a seguir são apresentados os demais capítulos desta tese. O Capítulo 2 apresenta os fundamentos para o leitor que ainda não está familiarizado com os temas abordados nesta pesquisa. O Capítulo 3 descreve uma visão geral dos métodos analíticos perceptuais comparados em relação ao método PEL. Os materiais e métodos utilizados nesta pesquisa são descritos em detalhes no Capítulo 4. O Capítulo 5 apresenta o novo método analítico perceptual proposto nesta pesquisa para

avaliação da qualidade de imagem. O Capítulo 6 relata os experimentos, análises e discussões do método proposto. Por fim, são apresentadas as conclusões, bem como os trabalhos futuros no Capítulo 7.

2

FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os fundamentos para compreensão dos assuntos a serem abordados no processo computacional, experimentos, análises e discussões nesta tese, a saber: Sistema Visual Humano (Human System Visual, HVS), Avaliação da Qualidade de Imagens (IQA) e Região de Interesse (RoI).

2.1 Sistema Visual Humano

Dentre as principais capacidades sensoriais dos seres humanos que permitem adequada percepção dos objetos no ambiente, a visão é uma das mais importantes [80]. Segundo Bear *et al.* [81], com base na luz que atinge a retina após ser refletida pelos objetos, os animais, dentre eles os humanos, identificam o mundo complexo ao seu redor. A luz é energia eletromagnética emitida na forma de ondas. Essas ondas colidem com os objetos, sendo absorvidas, espalhadas, refletidas ou desviadas. Devido à natureza das ondas eletromagnéticas e suas interações com o ambiente, é possível ao sistema visual extrair informações acerca do mundo. A importância da visão pode ser devidamente certificada pelo fato de que cerca de metade do córtex cerebral humano está envolvido com a análise do mundo visual. O sistema visual dos mamíferos começa no olho. No fundo do olho, como pode ser visto na seção transversal do olho humano na Fig. 1 em Kolb *et al.* [82], localiza-se a retina, que contém fotorreceptores especializados na conversão de energia luminosa em atividade neural. Como uma câmera, o olho ajusta-se automaticamente às diferenças de iluminação e focaliza automaticamente objetos de interesse. A visão, na verdade, envolve a percepção de numerosas propriedades distintas dos objetos [83], [84].

Nesse contexto, para Lent [85], a percepção, entretanto, é o aspecto mais apurado e sofisticado da modalidade visual. Como as demais modalidades sensoriais, a visão também se subdivide em submodalidades diferentes, a saber: a localização espacial, que permite identificar em que posição do campo de visão aparece um determinado objeto; a medida da intensidade da luz, que possibilita estimar a luminância de cada objeto localizado neste campo de visão; a sensibilidade de contraste, medida das

variações relativas de luminância; a discriminação de formas, que permite diferenciar e reconhecer os objetos segundo os seus contornos; a detecção de movimento, utilizada para perceber objetos que se movem, diferenciando naturalmente daqueles que estão parados; e, por fim, a visão de cores.

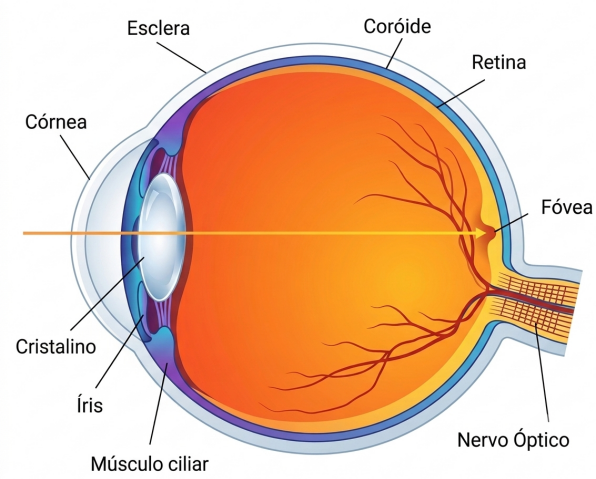


Figura 1 – Seção transversal do olho humano.
Fonte: Adaptado de Kolb *et al.* [82].

Devido à natureza desse processo, notadamente complexo, justifica-se a grande dificuldade do desenvolvimento de sistemas visuais automatizados. No entanto, para que desenvolvedores, engenheiros e pesquisadores possam construir máquinas e sistemas cada vez mais eficientes e eficazes, é imprescindível que esses compreendam o funcionamento do próprio sistema visual [80], claro, conforme as submodalidades da visão necessárias ao foco da investigação científica em questão. Sendo assim, nesta seção são descritos alguns temas relevantes que influenciam no funcionamento do HVS, bem como na seção 5.2 serão apresentadas mais características perceptuais do HVS utilizadas nesta tese. Mais detalhes sobre o HVS podem ser encontrados em [81], [83–87].

2.1.1 Propriedades da Luz

Entre as mais notáveis descobertas científicas do século XIX, está a de que a luz é uma onda eletromagnética [88]. A radiação eletromagnética é originada de diversas fontes, incluindo antenas de rádio e televisão, celulares, máquinas de raios-X, copiadoras e o sol. A luz é a radiação eletromagnética percebida pelo sistema visual humano. A radiação eletromagnética pode ser descrita como uma onda de energia e, como tal, possui comprimento de onda, frequência e amplitude. O comprimento de onda é a distância entre os sucessivos "picos" ou "vales" da onda. A frequência é o número de comprimentos de onda por segundo. A amplitude é a diferença de altura entre o topo do pico e o fundo do vale da onda [81].

Sabe-se que existe uma variação ampla e contínua nos comprimentos de onda e frequência das ondas eletromagnéticas. A luz é a parte do espectro eletromagnético entre o comprimento de onda de 400

nm e 700 nm, conforme Fig. 2 [85]. Segundo Ramalho Júnior *et al.* [88], o maior comprimento de onda da luz visível provoca a sensação de vermelho. À medida que o comprimento de onda diminui, a sensação de cor muda para alaranjado, amarelo, verde, azul, anil até atingir o violeta, que tem menor comprimento de onda. Bear *et al.* [81] ressaltam, no entanto, que o conteúdo energético da radiação eletromagnética é proporcional a sua frequência. Com isso, radiações de alta frequência (comprimento de onda curta) têm maior conteúdo energético, enquanto radiações emitidas em baixas frequências (comprimento de onda longa) possuem menor energia. Outro ponto importante é que a onda eletromagnética viaja em linha reta e, assim, pode ser descrita como um raio ou radiação. O estudo dos raios de luz e suas interações é denominado de óptica. Essas interações incluem a reflexão, a absorção e a refração. O HVS, por meio de refração, forma imagens na retina de cada olho.

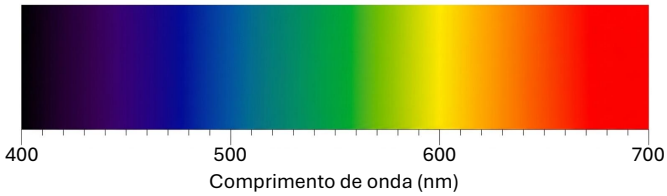


Figura 2 – Espectro eletromagnético visível ao olho humano.
Fonte: Adaptado de Lent [85].

2.1.2 Medida da Intensidade e Sensibilidade de Contraste

Além das descrições na Seção 2.1.1, é importante salientar, conforme Gonzáles e Woods [89], que a caracterização da luz é essencial para a ciência das cores. Se a luz for acromática (sem cores), seu único atributo será sua intensidade. A luz acromática é aquela que pode ser notada em um aparelho de televisão branco e preto. Neste caso, o termo nível de cinza refere-se a uma medida escalar de intensidade que varia do preto ao branco, com valores intermediários de cinza. Já a luz cromática engloba o espectro de energia eletromagnética desde aproximadamente 400 até 700 nm, como mostrado na Fig. 2. Três valores básicos são usados para descrever a qualidade de uma fonte de luz cromática, a saber: radiância, luminância e brilho. A radiância é a quantidade total de energia que flui de uma fonte de luz, sendo geralmente medida em watt (W). A luminância, medida em lúmen (lm), dá uma medida da quantidade de energia que um observador percebe de uma fonte de luz. E, por fim, o brilho é um descritor subjetivo, que é praticamente impossível de ser medido. Ele incorpora a noção acromática de intensidade, sendo um dos fatores fundamentais na descrição da sensação de cores.

O contraste é uma medida das variações relativas de luminância [90]. Esta medida, segundo Gazzaniga e Heatherton [86], também é conhecida como limiar de diferença – que é a diferença apenas perceptível (DAP), do inglês *just noticeable difference* (JND), entre dois estímulos. Diversas equações para descrever a relação entre mudanças na intensidade do estímulo e as correspondentes mudanças na

experiência perceptual desses estímulos foram propostas. O psicofísico alemão Ernst Weber descobriu que o limiar de diferença para um estímulo é uma proporção constante de sua intensidade de luminância. Essa medida ficou conhecida como Lei de Weber [91], conforme a equação a seguir:

$$C_W = \frac{\Delta L}{L}, \tag{2.1}$$

em que C_W é uma constante, conhecida como a fração ou contraste de Weber; L é a intensidade do estímulo padrão (luminância); e ΔL é a DAP para o estímulo. Gustav Theodor Fechner (1801-1887), um físico alemão, estabeleceu uma relação mais ampla entre a sensação e a intensidade de um estímulo. Ele combinou a lei de Weber com a ousada suposição de que cada DAP representa uma etapa igual na magnitude psicológica de uma sensação. Essa relação também é denominada de Lei de Weber-Fechner. Para Pedrini e Schwartz [80], esta lei é apenas uma aproximação da percepção sensorial real, embora muitas medidas de contraste utilizadas em visão computacional sejam baseadas nesse conceito.

Apesar dessas considerações, existem muitas ilusões visuais envolvendo a percepção do nível de luminosidade. Por exemplo, na Fig. 3, embora os dois quadrados centrais sejam do mesmo tom de cinza, o quadro da esquerda parece mais escuro do que o da direita [81].

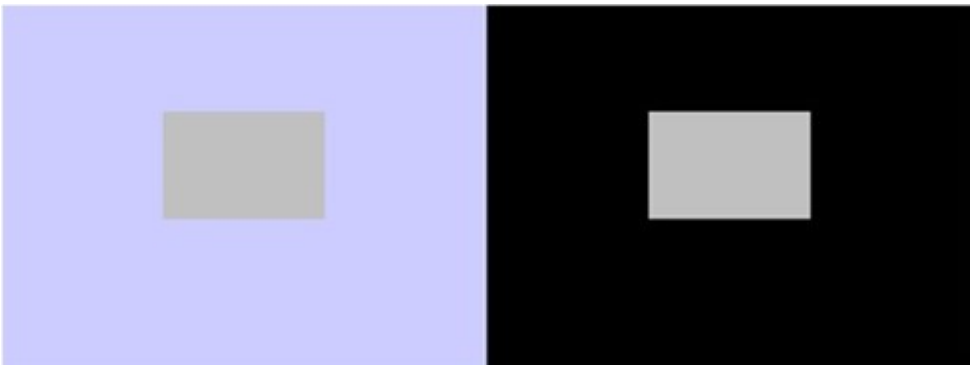


Figura 3 – Ilusões visuais devido à influência do contraste. Os quadrados centrais possuem a mesma tonalidade de cinza, contudo, devido ao fundo mais claro no quadro da esquerda, esse quadrado central parece mais escuro. Fonte: Adaptado de Bear *et al.* [81].

A percepção da figura exata, para Winkler [90], depende em grande parte das características dos estímulos. As funções de sensibilidade de contraste (Contraste Sensitivity Functions, CSF) são geralmente usadas para quantificar essas dependências. A sensibilidade de contraste é definida como a inversa do limiar de diferença (ou de contraste). Em medidas do CSF, o contraste de estímulo periódico (frequentemente senoidal) com frequências variando de L_{min} a L_{max} é definido como a medida de contraste de Michelson, proposto por Michelson [92], representada pela equação a seguir:

$$C_M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}, \tag{2.2}$$

onde L_{max} e L_{min} correspondem à luminância mínima e máxima do estímulo. Para Pedrini e Schwartz [80], as medidas de contraste C_W e C_M, embora sejam bons preditores de contraste para estímulos

simples, falham quando os estímulos tornam-se mais complexos e cobrem um intervalo de frequência maior.

2.2 Avaliação da Qualidade de Imagem (IQA)

Como descrito na Seção 1.1, a garantia exigida por aplicações multimídia é comumente subsidiada por técnicas, tais como super-resolução, compressão, restauração, recuperação e detecção de objetos em imagem e vídeo [2]. Apesar de tais técnicas estarem recebendo bastante atenção das comunidades acadêmica e industrial nos últimos anos [3], [4], seus processos frequentemente introduzem artefatos que degradam significativamente a qualidade de imagens e vídeos [5], como, por exemplo, efeitos de blocagem, ruído, *ringing*, bordas falsas e borramento. Na Seção 2.2.1 são abordados alguns artefatos encontrados em imagem e vídeo. Nesse contexto, vale ressaltar que, para mitigar tais distorções, os pesquisadores vêm utilizando métodos para avaliação da qualidade de imagem (*image quality assessment* - IQA), que podem ser agrupados em duas categorias, denominadas métodos subjetivos (Seção 2.2.2) e métodos objetivos, detalhadas, respectivamente, nas seções 2.2.2 e 2.2.3.

2.2.1 Artefatos

Nas subseções a seguir, são descritos alguns artefatos para fins de entendimento deste trabalho. Entretanto, outros artefatos podem ser encontrados na literatura sobre qualidade. A ITU-T P.930 (1996) [93] é uma recomendação que pode ser utilizada para simular artefatos resultantes da compressão. Essa recomendação descreve cinco artefatos (distorções de blocos, borramento, *edge busyness*, ruído e *jerkiness*). Yuen [94] apresenta um excelente material sobre avaliação e classificação de artefatos. Outras referências também fazem menção a este tema, como em Winkler [7] e Harte [95]. Nas subseções seguintes, são descritos alguns artefatos relevantes para o contexto desta pesquisa.

2.2.1.1 *Blockiness* (Blocagem ou distorção de bloco)

Distorção de bloco, também conhecida como blocagem (*blocking* ou *blockiness*), é frequentemente causada quantização grosseira dos coeficientes da transformada em cada bloco, resultando em descontinuidades de intensidade nos limites dos blocos. Artefato característico de esquemas de codificação baseados em blocos (como JPEG, MPEG[7], H.261[8]) [93], [97]. O algoritmo clássico de compressão JPEG, por exemplo, em altas taxas de compressão, produz blocagem [98]. Para um dado nível de quantização, a blocagem

[7] MPEG: Moving Picture Experts Group (Grupo de Especialistas em Imagens em Movimento).

[8] H.261: é o primeiro padrão internacional de compressão de vídeo digital para fins práticos. Utiliza técnicas de previsão de quadros (conhecidas como I-frames e P-frames) para eliminar redundâncias espaciais (dentro da mesma imagem) e temporais (entre imagens sucessivas). O H.261 estabeleceu a base para praticamente todos os codecs de vídeo modernos que usamos atualmente, como o H.264 (AVC), H.265 (HEVC) e H.266 (VVC) [96].

é geralmente mais visível em áreas de gradientes suaves ou homogêneas da imagem, enquanto que em regiões com grande número de bordas ou texturas ásperas tende a ser menos perceptível pelo HVS. Uma maneira de simular blocagem é por meio de mudanças nos valores de luminância em um bloco de píxeis, de modo a fazer com que este bloco se diferencie dos seus vizinhos. Uma observação importante a considerar é o tamanho do bloco de 8x8 píxeis, pois muitos codificadores trabalham com a Transformada Discreta de Cosseno (Discrete Cosine Transform, DCT) [93]. Na Fig. 4(b), efeitos de blocagem podem ser notados, claramente, na parte da base do farol e no fundo da imagem.



(a) (b) (c)

Figura 4 – (a) Imagem original; (b) Ilustração de artefatos típicos de compressão causados por métodos baseados em DCT; (c) métodos baseados em Wavelet.
Fonte: Winkler (2005, p. 43).

2.2.1.2 *Blurring* (Borramento)

O borramento (*blurring*) manifesta-se como uma perda moderada de detalhe espacial para regiões de imagens com alta atividade espacial, como áreas com texturas ásperas ou ao redor das bordas dos objetos da cena [94], reduzindo dessa forma a nitidez das bordas. Essas degradações são resultados da limitação da resolução espacial (subamostragem), desfoque óptico (lente fora de foco), movimento da câmera ou do objeto, ou filtragem passa-baixa [93]. O HVS é muito sensível à perda de bordas, cruciais para a

identificação de formas e estruturas, sendo um artefato considerado intrusivo em imagens médicas [99]. Na Fig. 4(c), por exemplo, é bem evidente o efeito do borramento, principalmente na base do farol, na casa e nas laterais do farol.

2.2.1.3 *Ringing*

O efeito *ringing* manifesta-se como oscilações artificiais (ecos) nas proximidades de transições de alto contraste. É uma distorção do resultado direto de quantização devido às altas frequências irregulares na reconstrução de vídeo ou mesmo, *upscale* em super-resolução de imagem [94], [99], [100]. Este artefato é mais perceptível ao longo de bordas com contraste alto (bordas fortes) em áreas de texturas suaves. O *ringing* afeta as componentes de luminância e crominância da imagem e do vídeo [90]. O HVS interpreta esses artefatos como "bordas fantasmagóricas", o que pode causar interpretações diagnósticas errôneas em imagens médicas [99]. Na Fig. 4(b), percebe-se o efeito do *ringing* ao redor dos contornos e bordas da torre.

2.2.1.4 *Staircase Effect* (Efeito Escada)

O artefato *Staircase Effect* manifesta-se como degraus indesejados em linhas curvas ou diagonais [93]. Esse artefato ocorre porque as imagens que são bases da DCT são mais adequadas para a representação de linhas horizontais e verticais, enquanto linhas com outras orientações requerem coeficientes DCT com frequências mais altas para reconstrução acurada [90], [101]. A distorção causada pelo *Staircase Effect* é facilmente percebida como uma quebra na continuidade das linhas, sendo um artefato que degrada significativamente a estética e a precisão geométrica [99]. Na Fig. 4(b), pode-se ver esse efeito ao redor da borda branca inclinada do farol.

2.2.2 Métodos Subjetivos

Além do que já foi mencionado na Seção 1.1, o método subjetivo é o tipo de avaliação padrão (*benchmark*) para se medir a qualidade de imagens ou vídeos. Para isso, em geral, existem três etapas fundamentais para se realizar essa avaliação. São elas: Seleção dos materiais e métodos; Planejamento dos procedimentos da avaliação; e Análises estatísticas e discussões dos resultados.

2.2.2.1 Seleção dos Materiais e Métodos

Como o objetivo principal é avaliar a qualidade de imagens ou vídeos, primeiramente, devem ser selecionados os materiais e métodos a serem utilizados.

Os materiais utilizados incluem as fontes de dados (imagens e/ou vídeos) e o próprio ambiente de testes com seus recursos. A seleção das fontes de dados é um ponto importante que deve ser considerado, pois as características do conteúdo dessas fontes podem influenciar nos resultados da avaliação. Corriveau [9] relata que desenvolvedores de algoritmos muitas vezes esquecem que, além da compressão, transmissão e manipulação de imagens, o maior fator que pode ser considerado é o conteúdo.

Detalhes das características e número de sequências de testes são definidos com o projeto experimental. A fim de evitar cansaço por parte dos observadores e alcançar um mínimo de confiança nos resultados, geralmente quatro tipos diferentes de cenas devem ser escolhidos para as sequências. Vale ressaltar também que, quando se quer comparar resultados de diferentes laboratórios, é necessário usar um conjunto comum de sequências de fontes de dados [102], [103].

Geralmente, em um ambiente de teste, existem monitores (como CRT, LCD, Plasma, entre outros), filmadoras, dispositivos de manipulação de fonte de dados, ferramentas para registro das avaliações, etc. No entanto, em um contexto mais geral, os três maiores fatores que devem ser considerados em um ambiente, segundo Corriveau [9], são a iluminação, ruído ambiente e a qualidade e calibração do monitor. De qualquer forma, antes de começar a realizar algum teste subjetivo, deve-se consultar as informações apontadas em recomendações da ITU em [102], [104] e [105].



Figura 5 – Exemplo de ambiente para avaliação subjetiva. Um ambiente de testes para formatos de HDTV.
Fonte: Hoffmann *et al.* [106].

Os métodos subjetivos, segundo Cuenca *et al.* [107], têm sido amplamente utilizados para avaliar novas tecnologias em telefonia e vídeo digital, como, por exemplo, novos sistemas de codificação e

compressão. No caso da tecnologia de imagem e vídeo digital, um grupo de pessoas está envolvido na avaliação subjetiva da qualidade, como ilustrado na Fig. 5, onde um grupo avalia vídeos em formato HDTV[9] [106]. Cada indivíduo é solicitado a opinar sobre a qualidade da imagem ou vídeo com uma pontuação S_i. Cada sequência, uma pontuação média, conhecida como MOS (*Mean Opinion Score*), é obtida conforme a equação a seguir:

$$MOS = \frac{1}{k}\sum_{i=1}^{k} S_i, \tag{2.3}$$

onde k é o número de participantes individuais e S_i é a pontuação fornecida pelo i-ésimo indivíduo. A escala recomendada pelo CCIR[10] é mostrada na Tabela 2, que define uma escala para a qualidade de imagem/vídeo, uma para a degradação sofrida pela imagem/vídeo e uma correspondente pontuação entre qualidade e degradação.

Tabela 1 – Escala Subjetiva* (Recomendação da CCIR 500-3).

Pontuação	Qualidade	Degradação
5	Excelente	Imperceptível
4	Bom	Perceptível, não irritante
3	Razoável	Ligeiramente irritante
2	Ruim	Irritante
1	Péssimo	Muito irritante

* Fonte: Cuenca *et al.* [107] .

Segundo a Recommendation BT.500-11 [104], existem duas classes ou modos de se realizar uma avaliação subjetiva: aquelas voltadas para quantificar a qualidade e aquelas destinadas a quantificar um tipo de degradação (erro, defeito ou *impairment*). Outra diferença fundamental, quanto aos testes de avaliação, é se o método usa referências explícitas ou implícitas, ou se o método não usa nenhuma referência.

Independente do modo ou referência, de qualquer forma, o método deve ter o propósito de avaliar um problema em particular. A Tabela 2 apresenta uma relação entre um problema típico de avaliação e o método a ser usado para resolvê-lo. Nesse sentido, diversos métodos foram propostos, onde os principais são listados na Tabela 2 [102], [104].

2.2.2.2 Planejamento dos Procedimentos da Avaliação

Para se conduzir testes sobre avaliações de qualidade em imagem ou vídeo por meio de métodos subjetivos, deve-se realizar um planejamento detalhado dos procedimentos a serem seguidos. Esses procedimentos envolvem, conforme [102]: as condições de visualização; a forma de exibição das fontes de dados; a seleção dos participantes como avaliadores; as instruções e treinamentos desses participantes. Os

[9] TV digital de alta definição (High Definition Television, HDTV)

[10] Comitê Consultivo Internacional de Rádio (International Radio Consultative Committee, CCIR). Vale ressaltar, que a Recomendação CCIR 500-3 (atualmente designada como Recomendação ITU-R BT.500) padroniza as "Metodologias para a avaliação subjetiva da qualidade da imagem de televisão". Ela é a norma de referência global para avaliar como o olho humano percebe defeitos, distorções e compressões em imagens e vídeos [105].

Tabela 2 – Seleção de Métodos Subjetivos.*

Problema a ser avaliado	**Método**
Medir a qualidade de sistemas em relação a uma referência.	Escala Contínua de Qualidade com Duplo Estímulo (Double-Stimulus Continuous Quality-Scale, DSCQS)
Medir a robustez de sistemas (ou seja, características de falhas).	Escala de imparidades com duplo estímulo (Double-Stimulus Impairment Scale, DSIS)
Medir a fidelidade entre duas sequências de imagens/vídeos degradadas.	Avaliação Contínua com Duplo Estímulo Simultâneo (Simultaneous Double Stimulus for Continuous Evaluation, SDSCE)
Medidas de qualidade de sistemas sem referência.	Avaliação Contínua de Qualidade com estímulo Simples (Single Stimulus Continuous Quality Evaluation, SSCQE)

* Fonte: adaptado da Recomendação ITU-R BT.500-11 [104].

procedimentos sobre as condições de visualização estão relacionados diretamente com as configurações dos parâmetros dos monitores (brilho, contraste, etc), dos fatores que devem ser considerados sobre o ambiente de testes, bem como das condições específicas dos participantes (avaliadores) quanto à distância e ângulo visual. Todas as condições, evidentemente, devem obedecer aos interesses do que se quer avaliar. Contudo, é importante considerar procedimentos que foram realizados por outros laboratórios, a fim de fazer comparações, refutações ou validações para um propósito específico.

A forma de exibição das fontes pode ser de duas formas. A primeira é realizada pela transmissão ou reprodução da imagem/vídeo em tempo real através dos sistemas de ensaio, enquanto avaliadores estão assistindo e respondendo. A segunda é realizada de forma *off-line* por transformação dos registros das pontuações, referentes aos testes, em novos dados, nos quais serão gravados para gerar um novo conjunto de registros. Quanto à seleção dos participantes, eles podem ser especialistas ou não especialistas. Todos os indivíduos devem ser submetidos a um teste de acuidade, daltonismo e outras deficiências visuais. Para produzir resultados confiáveis, muitos participantes deverão ser utilizados. É geralmente aceito que entre 16 e 24 participantes possam fornecer resultados estatisticamente válidos, entretanto, quanto mais participantes, melhor [9].

Antes de iniciar o experimento, um cenário da aplicação planejada do sistema em teste deve ser dado aos avaliadores. Além disso, uma descrição do tipo de avaliação, a escala de opinião, bem como a apresentação dos estímulos, é dada em forma escrita. A faixa e tipos de degradação devem ser apresentados em ensaios preliminares, que podem conter sequências de imagens ou vídeos diferentes dos utilizados nos testes reais. Perguntas sobre o procedimento ou sobre o significado das instruções devem ser respondidas com cuidado para evitar predisposição e somente antes do início da sessão [102].

2.2.2.3 Análises Estatísticas e Discussões dos Resultados

Uma vez que a pessoa responsável pela avaliação tenha realizado a experiência subjetiva, e tendo em mãos todos os requisitos, ela tem uma coleção de dados subjetivos que devem ser analisados estatisticamente. Neste contexto, o responsável tem de determinar até que ponto os dados podem ser confiáveis. Para isso, determina-se o intervalo de confiança. Sendo assim, o cálculo de dados de cada participante tem de

ser investigado e uma análise estatística dos resultados pode ser aplicada com métodos normalizados [108]. Análises estatísticas de dados, referentes a métodos subjetivos, podem ser encontradas em [10], [11], [102], [104]. Em Hoffmann *et al.* [106] pode ser encontrado um estudo, envolvendo o formato HDTV com resolução 1920x1080 píxeis, que serve também como um bom exemplo para uma avaliação subjetiva, inclusive em relação a todas as etapas citadas.

2.2.3 Métodos Objetivos

Os métodos objetivos têm a finalidade de predizer a opinião de pessoas sobre a qualidade de imagens ou vídeos, sem que sejam realizados testes subjetivos.

Os métodos de FR-IQA podem ser divididos em dois grupos: *métodos baseados em píxeis* (Seção 2.2.3.1); e *métodos perceptuais* [17]. Os métodos baseados em píxeis, devido ao seu baixo custo computacional, são matematicamente fáceis de implementar e são muito eficientes. Os métodos perceptuais, por outro lado, consideram as características do HVS em uma tentativa de incorporar medidas de qualidade perceptual. Nesse sentido, vários métodos foram propostos, os quais podem ser divididos em duas categorias: *métodos analíticos* (Seção 2.2.3.2); e *métodos baseados em aprendizagem* (Seção 2.2.3.3).

2.2.3.1 Métodos Objetivos Analíticos baseados em Píxeis

Embora simples e amplamente utilizados, os IQMs baseados em píxeis, como erro médio quadrático (*Mean Squared Error*, MSE), relação sinal-ruído de Pico (*Peak Signal-to-Noise Ratio*, PSNR) e raiz do erro quadrático médio (*Root Mean Squared Error*, RMSE) [18], são criticados por sua baixa eficácia, [19], [20]. Vale ressaltar, no entanto, que tais métodos vêm sendo amplamente utilizados como função de perdas (*loss function*) em modelos baseados em aprendizagem.

Os métodos MSE e PSNR são calculados, respectivamente, conforme as Eq. 2.4 e 2.5. Pela Eq. 2.4, tem-se x como um píxel da imagem de referência e y como um píxel da imagem de teste (ou a imagem degradada). O resultado dessa equação é um valor, conhecido como *score* ou índice, que indicará a qualidade da imagem de teste, onde zero indica que a imagem não tem nenhuma degradação. O MSE é a soma do quadrado das diferenças entre os valores dos níveis de cinza entre essas duas imagens, sendo $x = x_i|i = 1, 2, 3, ..., N$ e $y = yi|i = 1, 2, 3, ..., N$, onde N é o número de píxeis e, x_i e y_i são as intensidades dos píxeis das imagens x e y, respectivamente.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2, \tag{2.4}$$

O método PSNR é utilizado para avaliar a diferença global entre duas imagens, semelhantemente em vídeos, definida como:

$$PSNR = 10log_{10}\frac{(2^n - 1)}{MSE}, \tag{2.5}$$

onde n é o número de bits por píxel do quadro (exemplo, $2^n = 256$ para quadros com píxeis de 8 bits). O método PSNR é expresso em decibéis (dB). Valores típicos de PSNR variam entre 20 e 40. Quanto maior for o valor desse método, mais a imagem/vídeo de teste se aproximará em qualidade da imagem/vídeo original [80], [97].

2.2.3.2 Métodos Objetivos Perceptuais Analíticos

Como descrito na Seção 1.1, os métodos perceptuais que não usam técnicas de aprendizagem são considerados aqui como métodos analíticos perceptuais. Nesta categoria, a Structural SIMilarity (SSIM) [21] pode ser considerada o marco da maioria dos métodos de IQA perceptuais. Com base na hipótese de que o HVS é altamente adaptado para perceber informações estruturadas, o SSIM pode prever a qualidade da imagem, capturando mudanças baseadas em contraste, luminância e informações estruturais entre as imagens de referência e as distorcidas (ou imagens de testes). A Fig. 6 apresenta o diagrama do método SSIM. O SSIM supera significativamente métodos de erro tradicional (PSNR/MSE) ao apresentar maior consistência com a percepção subjetiva de qualidade [14]. Comparado a sua versão estendida, ou seja, a *Multi-scale SSIM* (MS-SSIM) [22], que utiliza um esquema multiescala para capturar variações estruturais em diferentes resoluções, o SSIM original é mais simples, mas menos robusto a variações de escala.

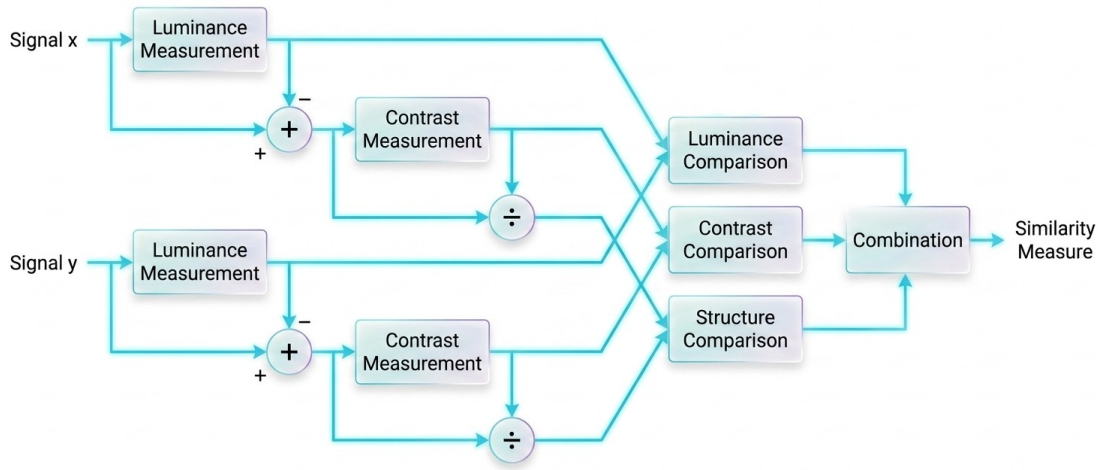


Figura 6 – Processo computacional do método Structural SIMilarity (SSIM).
Fonte: Adaptado de Wang *et al.* [14].

O método *Visual Information Fidelity* (VIF) [23], e sua versão estendida, usam coeficientes wavelet para quantificar informações compartilhadas entre imagens de referência e distorcidas. SSIM, MS-SSIM e VIF podem oferecer desempenho estatisticamente muito melhor, entretanto tais métodos não são consistentes com a intuição de que diferentes regiões em uma imagem podem ter contribuições muito diferentes em relação à percepção da imagem pelo HVS [24]. Além disso, as versões SSIM e VIF são menos eficientes devido ao seu maior custo computacional quando comparadas ao método SSIM, bem

como métodos combinados ou baseados em fusão para FR-IQA como em [25–29]. É importante destacar que os métodos baseados em píxeis também são analíticos, mas não são considerados perceptuais. Vários outros métodos analíticos FR-IQA usados neste trabalho são apresentados na Seção 3.1.

2.2.3.3 Métodos Objetivos Perceptuais baseados em Aprendizagem

Os métodos IQA baseados em aprendizagem são uma das novas tendências para IQA perceptual. Eles visam usar técnicas de aprendizagem para deduzir uma função matemática para o método a partir de uma relação entre diferentes efeitos de distorção e a qualidade da imagem [30]. Na Seção 1.1, esses métodos já foram tratados de forma geral. Vale ressaltar, que apesar de não serem objetos de estudo nesta tese, a seguir, são compilados diagramas de alguns processos computacionais, sem entrar em detalhes de seu funcionamento, a fim de mostrar os recursos computacionais necessários e complexos para se construir alguns desses métodos baseados em aprendizagem.

Na última década, várias abordagens para implementar técnicas de aprendizagem em métodos para IQA, visando prever a qualidade da imagem com a mesma acurácia do julgamento humano, foram desenvolvidas. As primeiras abordagens utilizaram técnicas de aprendizagem de máquina[11,12] como em [33–35]. Narwaria e Lin [35] propuseram um método baseado em decomposição de valor singular (*Singular Value Decomposition*, SVD) usando support vector regression (*support vector regression*, SVR).

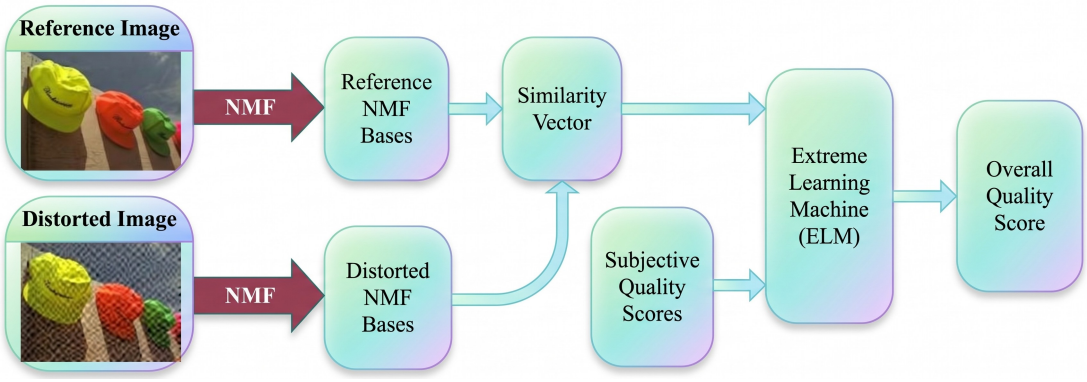


Figura 7 – Processo computacional do método non-Negative Matrix Factorization (NMF). Fonte: Adaptado de Wang *et al.* [33].

Wang *et al.* [33] propuseram um novo método FR que aplica fatoração de matriz não negativa (*non-Negative Matrix Factorization*, NMF) para medir degradações de imagem. O método FR-IQA proposto, conforme Fig. 7, emprega a fatoração de NMF para descrição de distorções e a máquina de aprendizado extremo (*Extreme Learning Machine*, ELM) para o *pooling* de efeitos. A NMF é utilizada para decompor as imagens original e distorcida em bases e coeficientes, aproveitando sua representação baseada em

[11] Aprendizado de máquina é a programação de computadores para otimizar os parâmetros de um modelo usando os dados de treinamento ou a experiência passada. O modelo pode ser preditivo para fazer previsões no futuro, ou descritivo para obter conhecimento a partir dos dados, ou ambos [31].

[12] O uso de técnicas de aprendizado de máquina e inteligência artificial se expandiu rapidamente em diferentes áreas da vida, como finanças, educação, medicina, indústria manufatureira, segurança cibernética [32], inclusive, na área de IQA.

partes para capturar degradações de forma consistente com a percepção humana. O modelo foca o aprendizado da relação entre o vetor de similaridade (obtido via similaridade de cosseno das bases NMF) e as pontuações subjetivas (MOS). Em termos de eficiência, a ELM proporciona uma velocidade de aprendizado significativamente mais rápida que redes neurais ou SVR, além de maior precisão. O tempo de processamento do método é competitivo, sendo mais rápido que métricas como MAD, VIF e IW-SSIM, demonstrando alta eficácia ao superar algoritmos do estado da arte em termos de precisão e robustez em múltiplos BDIs.

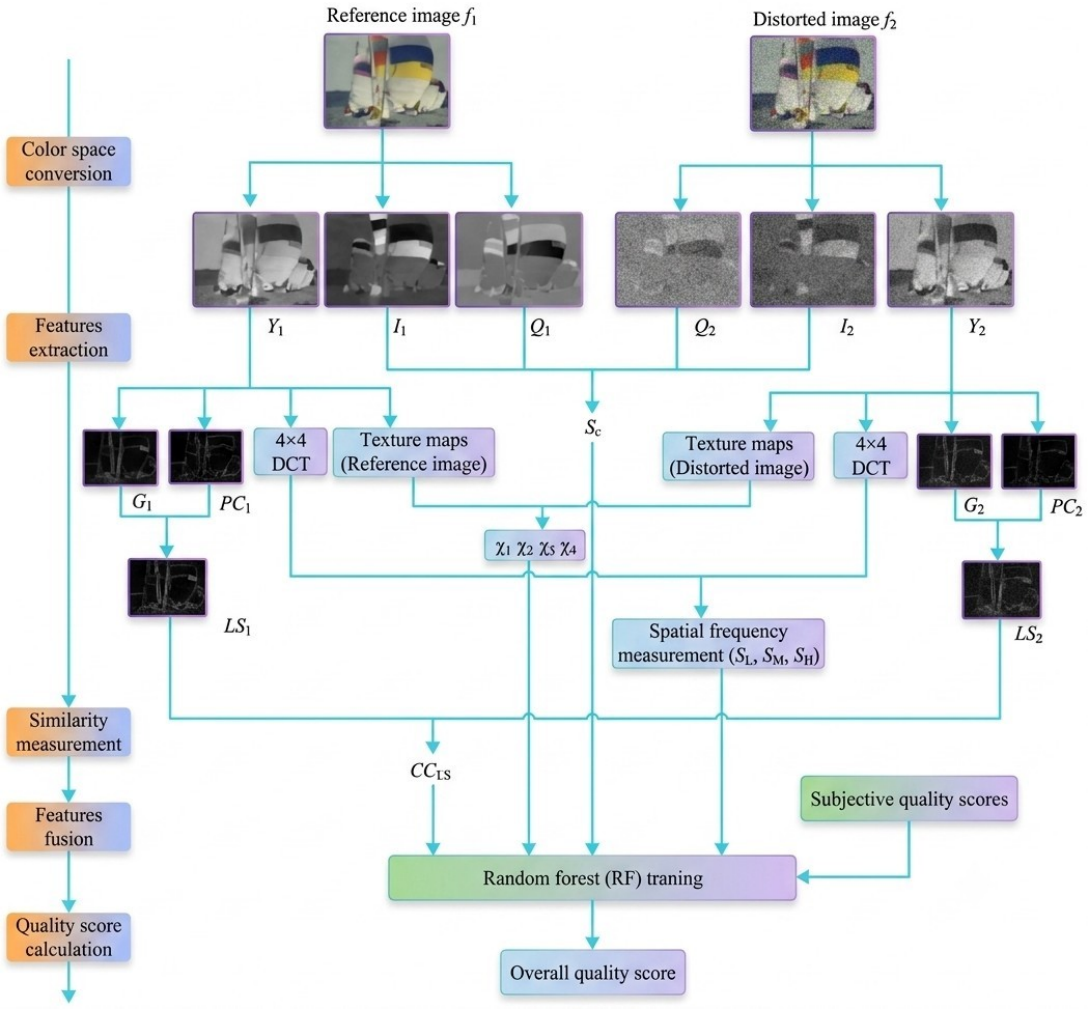


Figura 8 – Processo computacional do método Perceptual Feature Similarity Index (PFSI). Fonte: Adaptado de Tang *et al.* [34].

Vários outros métodos FR-IQA baseados em aprendizagem podem ser encontrados na literatura, como em [34], baseados em *Random Forest* (RF). Tang *et al.*[34] propuseram o método Perceptual Feature Similarity Index (PFSI), Fig. 8, que avalia a qualidade de imagens coloridas combinando características dos domínios espacial e de frequência. PFSI extrai o gradiente (GM) no domínio espacial, depois utiliza filtros log-Gabor (textura) e DCT (frequência) para capturar propriedades do HVS. Uma inovação é o uso de um mapa de estrutura localizada (LS), que combina GM e congruência de fase (PC), semelhante

a FSIM [24], para melhorar a percepção de distorções estruturais, superando limitações individuais de cada mapa. O modelo utiliza RF como ferramenta de regressão para fundir vetores de similaridade e pontuações subjetivas, superando métodos de *pooling* tradicionais. Embora o tempo médio de processamento (0,1348s) seja superior a métodos simples como GMSD (0,0083s), ele é comparável ao FSIM (0,0983s). O PFSI demonstra eficácia superior em predição de qualidade, com alta precisão e robustez em diversos BDIs, superando amplamente métricas do estado da arte.

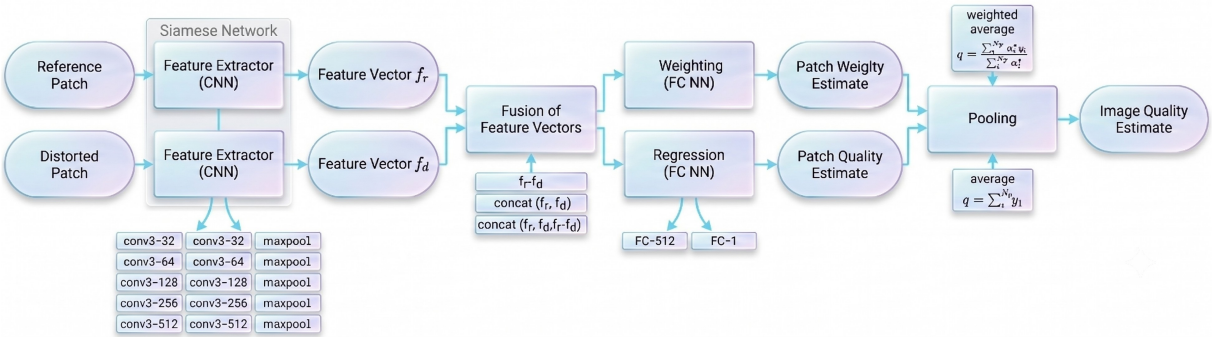


Figura 9 – Processo computacional do método Deep Image QuAlity Measure para FR IQA (DIQaM-FR). Fonte: Adaptado de Bosse *et al.* [36].

Mais recentemente, métodos de IQA baseados em aprendizagem têm sido desenvolvidos por meio de redes neurais profundas (*Deep Learning*, DL) como em [36–40]. Nesse sentido, em nível de IQA, Bosse *et al.* [36] apresentaram uma abordagem baseada em DL para NR e FR-IQA, onde a Fig 9 ilustra o método Deep Image QuAlity Measure para FR IQA (DIQaM-FR). Varga [43] apresenta uma nova arquitetura NR-IQA baseada em aprendizado profundo que depende da fusão de decisão de pontuações de qualidade de múltiplas imagens, provenientes de diferentes tipos de redes neurais convolucionais (*Convolutional Neural Networks*, CNNs).

Outros métodos FR-IQA podem ser encontrados na literatura, tendo como arquitetura uma CNN ou uma DNN [37], [38], [46–51], [109]. Nesse contexto, Kim *et al.* [37] propuseram o modelo Dynamic Receptive Field generation based Image Quality Assessor (DRF-IQA), ilustrado na Fig. 10, que possui uma estrutura de avaliação de qualidade de imagem baseada em campos receptivos dinâmicos que se ajustam ao tipo de distorção. O modelo divide o processamento em dois fluxos: representação dinâmica de erro (*stream 1*) e *pooling* de qualidade sensível à sensibilidade visual (*stream 2*). O *stream 1* foca na representação dinâmica de erro, onde uma CNN gera perfis de campos receptivos dinâmicos a partir da imagem distorcida. Esses perfis são convoluídos com as imagens de referência e distorcidas, resultando em mapas de erro espacial. Já o *stream 2* envolve a geração do mapa de sensibilidade visual. Este mapa é utilizado como um mecanismo de ponderação que atua sobre o mapa de erro espacial gerado no primeiro fluxo, refinando a estimativa da qualidade final ao priorizar áreas com maior sensibilidade percebida. Diferente de métodos holísticos, ele gera campos receptivos via uma CNN treinada para produzir mapas de erro espaciais, os quais são posteriormente ponderados por um mapa de sensibilidade visual. O

modelo se destaca por sua alta eficácia e precisão de predição superior em diversas BDIs.

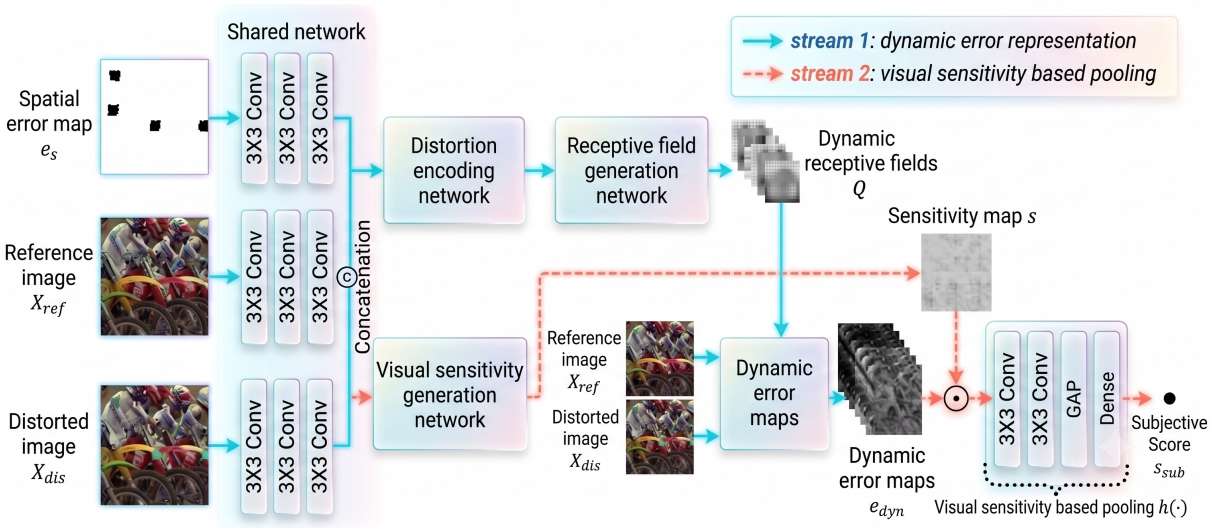


Figura 10 – Processo computacional do método Dynamic Receptive Field generation based Image Quality Assessor (DRF-IQA).
Fonte: Adaptado de Kim *et al.* [37].

2.2.4 Métricas Estatísticas de Predição da Avaliação

A validação de modelos de Avaliação de Qualidade de Imagem (IQA) depende da capacidade do algoritmo em replicar o julgamento humano. Para medir essa eficácia, tanto na literatura [16], [71], [77], [110–116], quanto em grupos de especialistas como o VQEG [11] definem métricas que avaliam a acurácia, a monotonicidade e a consistência das predições em relação aos scores subjetivos (Mean Opinion Score - MOS).

Antes da aplicação de algumas métricas, é comum realizar um mapeamento não linear (regressão logística) para compensar variações na escala de avaliação. As métricas mais utilizadas são descritas a seguir, conforme [10], [11].

O Coeficiente de Correlação Linear de Pearson (*Pearson Linear Correlation Coefficient*, PLCC) mede a precisão da predição, avaliando a correlação linear entre os valores objetivos previstos pelo modelo (s_i) e as notas dadas por humanos (o_i).

A fórmula do PLCC é expressa por:

$$PLCC = \frac{\sum_{i=1}^{N}(s_i - \bar{s})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{N}(s_i - \bar{s})^2 \sum_{i=1}^{N}(o_i - \bar{o})^2}},$$

onde N é o número de amostras, e $\bar{s}$ e $\bar{o}$ representam as médias das pontuações objetivas e subjetivas, respectivamente [77]. Valores próximos de 1 ou -1 indicam alta precisão linear.

O Coeficiente de Correlação de Postos de Spearman (*Spearman Rank-Order Correlation Coefficient*, SROCC) é utilizado para avaliar a monotonicidade da predição. Diferente do PLCC, esta métrica não depende dos valores exatos, mas sim da ordem (postos) das imagens, sendo ideal para verificar se o

modelo classifica corretamente qual imagem tem melhor qualidade que outra.

A fórmula do SROCC é dada por:

$$SROCC = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N(N^2 - 1)},$$

sendo d_i a diferença entre a posição (rank) da imagem i na avaliação objetiva e na avaliação subjetiva [111].

Similar ao SROCC, o Coeficiente de Correlação de Postos de Kendall (*Kendall Rank-Order Correlation Coefficient*, KROCC) [117] também mede a monotonicidade, mas foca na concordância de pares. Ele verifica se a ordem de qualidade entre dois pares de imagens é a mesma para o humano e para o algoritmo.

A fórmula é definida como:

$$KROCC = \frac{N_c - N_d}{0.5N(N-1)},$$

onde N_c é o número de pares concordantes e N_d o número de pares discordantes [114].

O Erro Quadrático Médio (*Root Mean Square Error*, RMSE) é a métrica utilizada para quantificar o erro absoluto de predição de um modelo de avaliação de qualidade. De acordo com a metodologia estabelecida pelo VQEG [11], o RMSE deve ser calculado a partir do erro absoluto de predição, denotado como P_{error}, após a aplicação de uma função de mapeamento para alinhar os scores objetivos aos subjetivos.

A fórmula para o cálculo do RMSE é definida como:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (P_{error}(i))^2}{N - d}},$$

onde N denota o número total de sequências de imagens consideradas na análise. $P_{error}(i)$ é o erro de predição para a amostra i, resultante da diferença entre o valor previsto pelo modelo (após o mapeamento) e o valor subjetivo real (MOS ou $DMOS$). d representa o número de graus de liberdade da função de mapeamento utilizada.

Conforme as diretrizes do VQEG, o valor de d depende da complexidade da função de ajuste. Por exemplo, no caso de um mapeamento que utilize uma função polinomial monotônica de 3ª ordem, utiliza-se $d = 4$, uma vez que existem quatro coeficientes na função de ajuste. Esta métrica é fundamental para a análise da consistência da predição, pois, ao contrário das correlações de postos, o RMSE penaliza variações significativas entre a qualidade prevista e a percebida, oferecendo uma visão clara da magnitude do erro residual do modelo [11].

Em resumo, enquanto o PLCC e o RMSE avaliam a proximidade numérica dos valores previstos em relação aos julgamentos humanos, o SROCC e o KROCC focam na capacidade do modelo em ordenar corretamente diferentes níveis de distorção. Um modelo de IQA robusto deve apresentar altos valores de PLCC e SROCC/KROCC, aliados a baixos valores de RMSE [113].

2.2.5 Classificação de Métodos quanto a sua Referência

Os métodos podem ser classificados, quanto a sua referência, em três categorias, a saber: referência completa (*Full-Reference*, FR), quando a fonte de referência está disponível; referência reduzida (*Reduced-Reference*, RR), quando informações parciais da fonte de referência estão disponíveis; e sem referência (*No-Reference*, NR, ou *blind*), quando a fonte de referência não está disponível [13]. Os exemplos a seguir foram ilustrados para vídeos, mas os preceitos quanto à referência da imagem para classificação de um método aplicado às imagens são os mesmos.

2.2.5.1 Método *Full-Reference* (FR)

Os métodos FR são muito utilizados em algoritmos de compressão para orientação e ajustes de parâmetros no encoder[13], bem como teste e monitoramento da qualidade no decoder. Os métodos FR realizam cálculos de comparação quadro a quadro entre o vídeo de referência e o vídeo de teste; essa categoria exige que o vídeo de referência, usualmente em uma forma não comprimida, encontre-se à disposição no mesmo lugar do vídeo de teste. Esse método é ilustrado na Fig. 11 para um cenário em uma rede IP (*Internet Protocol*).

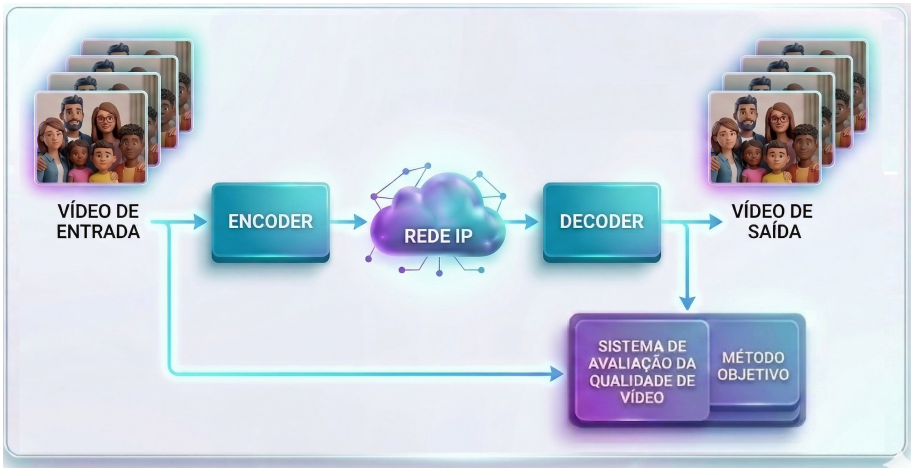


Figura 11 – Referência Completa (*Full-Reference*, FR).
Fonte: Autoria própria.

2.2.5.2 Método *No-Reference* (NR)

Métodos NR realizam os cálculos considerando apenas o vídeo de teste. A Fig. 12 ilustra um cenário em uma rede IP para avaliações de qualidade de vídeo feitas com métodos NR, que são os mais complexos,

[13] A compressão é o processo de compactar dados em um número menor de bits. A compressão envolve um par complementar de sistemas: um codificador (encoder) e um decodificador (decoder). O codificador converte os dados de origem em uma forma comprimida (ocupando um número reduzido de bits) antes da transmissão ou armazenamento, e o decodificador converte a forma comprimida de volta em uma representação dos dados de vídeo originais. O par codificador/decodificador é frequentemente descrito como um CODEC (codificador/decodificador), como por exemplo, MPEG-4 e H.264 [97].

pois não usam a referência (fonte dos dados) em suas avaliações. A construção desses métodos, devido à ausência da referência, torna-se um desafio atualmente.

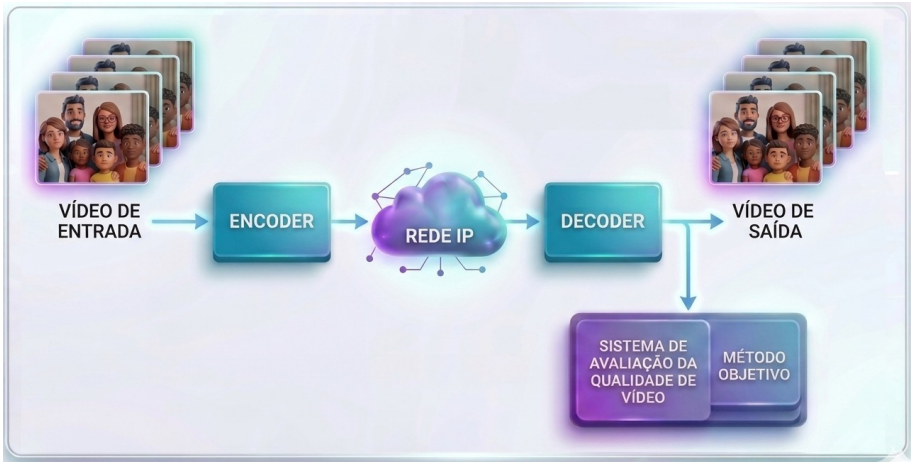


Figura 12 – Sem referência (*No-Reference*, NR).
Fonte: Autoria própria.

2.2.5.3 Método *Reduced-Reference* (RR)

Métodos RR realizam os cálculos considerando algumas características extraídas do vídeo de referência e as comparam com as características do vídeo de teste. A Fig. 13 ilustra um cenário em uma rede IP, onde, além do vídeo de entrada codificado, transmitido e decodificado, algumas características do vídeo são extraídas, como bordas, por exemplo, e enviadas também para o destino. Com o vídeo decodificado, bem como as características do vídeo, evidentemente sincronizadas espaço-temporalmente, um sistema, por meio de métodos RR, pode avaliar a qualidade deste vídeo no destino.

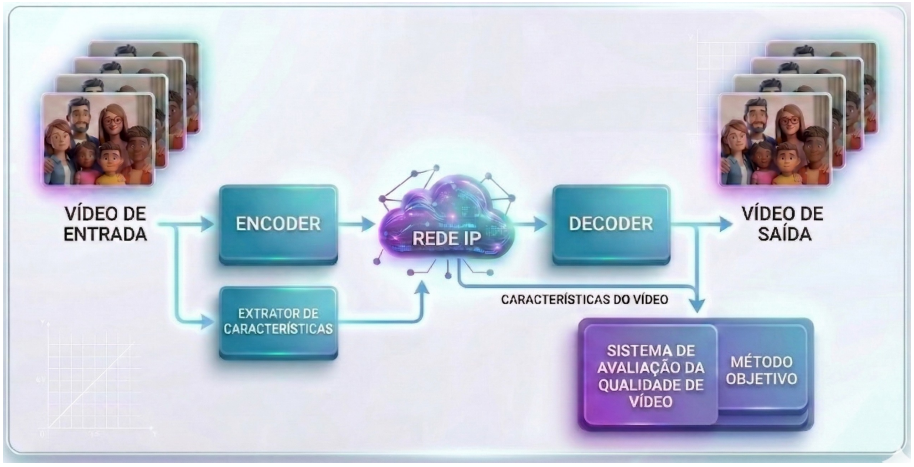


Figura 13 – Referência Reduzida (*Reduced-Reference*, RR).
Fonte: Autoria própria.

2.3 Região de Interesse (RoI)

A Região de Interesse (*Region of Interest*, RoI) é definida como um subconjunto ou partição específica de uma imagem que contém as informações mais relevantes para uma determinada análise ou para a percepção humana. No contexto do processamento digital, a identificação de uma RoI permite a redução da complexidade computacional, concentrando os recursos de processamento apenas nos píxeis que possuem alto valor informativo, ignorando áreas redundantes ou de fundo [114], [118], [111], [119].

Sob o guarda-chuva da Avaliação de Qualidade de Imagem (IQA), o conceito de RoI está intrinsecamente ligado à Saliência Visual (VS). Estudos da literatura, como em [120], [77], [114], [121], demonstram que o HVS não analisa uma imagem de forma uniforme; em vez disso, ele foca a atenção em regiões específicas que atraem o olhar devido a características como contraste, cor, brilho, bordas ou movimento [118], [119]. Portanto, modelos modernos de IQA utilizam mapas de RoI para ponderar a qualidade local, atribuindo pesos maiores às distorções que ocorrem dentro das áreas de interesse, uma vez que estas têm um impacto significativamente maior na opinião subjetiva final (MOS) do que falhas em áreas periféricas [120].

A Fig. 45 apresenta exemplos de RoI extraídas da BDI denominada VAIQ (Visual Attention for Image) [122]. A BDI VAIQ[14] foi criada com base nos resultados de um experimento de rastreamento ocular envolvendo 15 observadores humanos do campus de Campbelltown da Universidade de Western Sydney. As idades variaram de 20 a 60 anos, com uma média de 42 anos. Essa BDI facilita a incorporação de modelos de atenção visual (*Visual Attention*, VA) em métricas de qualidade de imagem projetadas com base nas BDIs IVC [123], LIVE [12] e MICT [124].



(a) (b) (c) (d)

(e) (f) (g) (h)

Figura 14 – (a-d) são imagens de referência, originais da BDI LIVE [12]. (e-h) são mapas com região de interesse (Region of Interest, RoI) na BDI VAIQ em [122] correspondentes, respectivamente, às imagens de referência de (a-d).
Fonte: Autoria própria.

[14] A BDI VAIQ está disponível gratuitamente para a comunidade científica [122].

3

TRABALHOS RELACIONADOS

Este capítulo apresenta os trabalhos relacionados a esta tese, que contém uma breve revisão das características perceptuais presentes nos métodos analíticos comparados, bem como são abordadas suas habilidades perceptuais (eficácia e eficiência). Vale ressaltar que os métodos mencionados na Seção 2.2.3.2 não são novamente abordados neste capítulo.

3.1 Introdução

A ideia central das seções a seguir é subsidiar o leitor quanto às características perceptuais presentes nos métodos analíticos comparados nesta tese. Para isso, tais características são descritas de forma geral, bem como são ilustrados os processos computacionais de cada método. Além disso, para suscitar discussões mais assertivas sobre as habilidades perceptuais, algumas observações são consideradas sobre a eficácia e eficiência desses métodos. Vale ressaltar, contudo, que para uma investigação mais aprofundada desses métodos, as fontes primárias de seus artigos devem ser consultadas.

3.2 Visão Geral dos Métodos Comparados

Além das características dos métodos analíticos mencionados na Seção 2.2.3.2, existem outros métodos que se baseiam na magnitude do gradiente (*gradient magnitude*, GM), ou GM combinada com outras características, como, por exemplo, congruência de fase (*phase congruency*, PC) ou saliência visual (*visual saliency*, VS). Neste contexto, a similaridade de características (Feature SIMilarity, FSIM/FSIM$_C$) [24], Fig. 15, emprega duas características (PC e GM) para caracterizar a qualidade visual percebida. A PC é usada como a característica primária no cálculo da FSIM. Enquanto isso, considerando que a PC é invariante ao contraste, a GM da imagem é computada como a característica secundária para codificar a informação de contraste. Zhang *et al.* [24] garantem que PC e GM são complementares e refletem diferentes aspectos do HVS em relação à avaliação da qualidade da imagem local.

4

PROCESSO METODOLÓGICO

ESTE capítulo apresenta, em detalhes, o processo metodológico de pesquisa utilizado para a criação do método analítico perceptual proposto à FR-IQA. Nesse contexto, são apresentadas oito bases de dados de imagens (BDI), bem como os procedimentos experimentais e análises estatísticas que subsidiam a comprovação das habilidades perceptuais (eficiência e eficácia) de PEL em relação aos dezenove métodos FR-IQA do estado da arte comparados.

4.1 Introdução

Como mencionado na Seção 1.4, esta tese é baseada em uma pesquisa com abordagem quantitativa. A coleta de artigos foi realizada de maneira exploratória.

Os passos que compõem o processo metodológico para criação do novo método analítico perceptual proposto nesta tese são listados a seguir:

i) Investigação de BDI e métodos existentes na literatura voltados à FR-IQA, que constituem os materiais necessários às devidas explorações científicas;

ii) Descrição do Processo Computacional do método proposto; e

iii) Avaliação do método proposto para verificar suas habilidades perceptuais competitivas, sob BDIs voltadas à FR-IQA, em relação aos métodos do estado da arte.

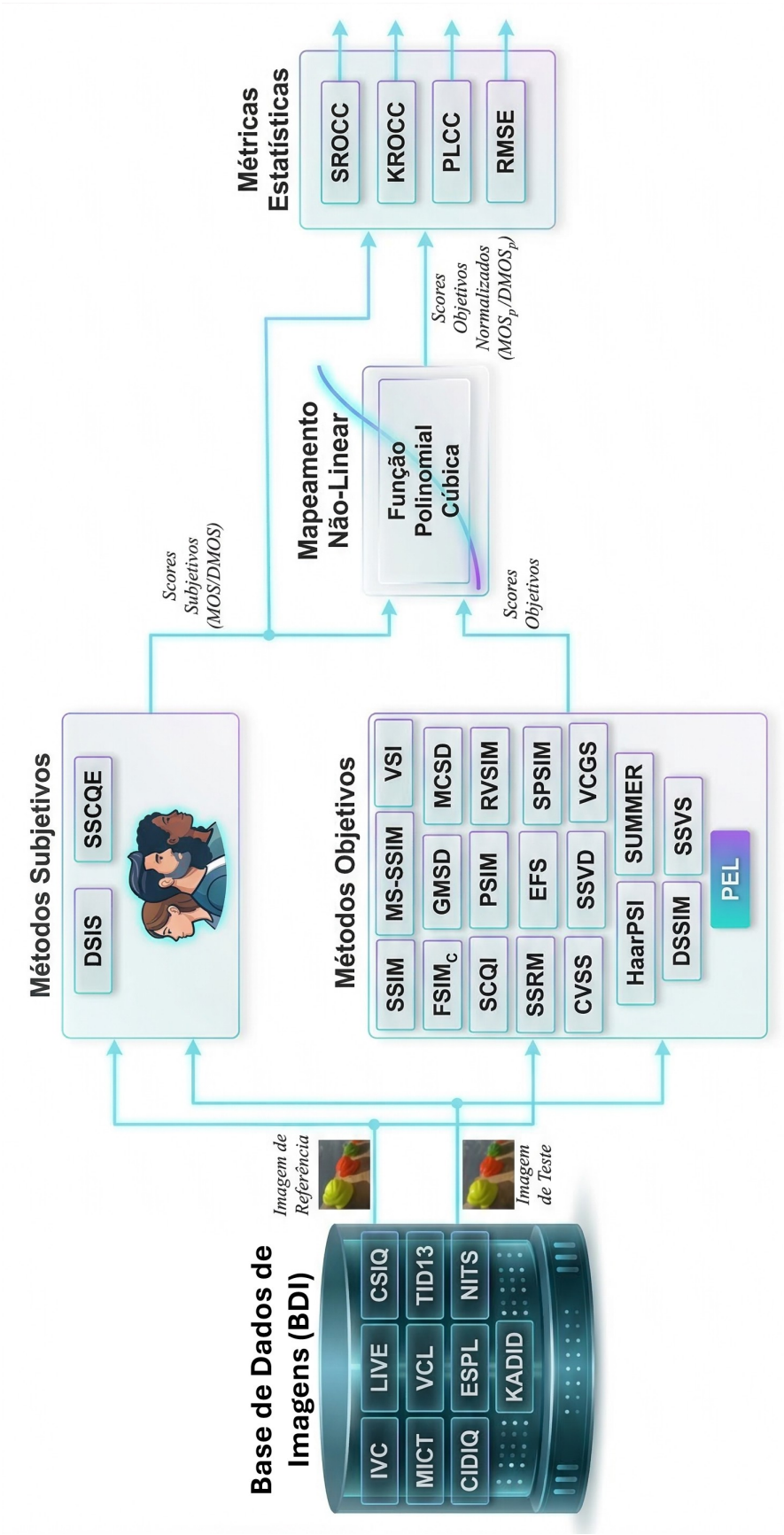


Figura 32 – Processo Metodológico.
Fonte: Autoria própria.

Os passos i e iii têm seus planejamentos apresentados nas próximas seções, conforme procedimentos recomendados nos relatórios finais produzidos pelo VQEG (Video Quality Experts Group) em 2003 [10] e 2008 [11], bem como suas implementações no Capítulo 6. O passo ii, refere-se ao desenvolvimento de métodos a serem propostos, neste sentido, o Capítulo 5 detalha o processo computacional do método PEL proposto. A Fig. 32 apresenta uma visão geral e sistemática do Processo Metodológico constituído nesta tese, descrito em mais detalhes nas seções seguintes.

4.2 Materiais

Os materiais utilizados nesta tese, descritos nas subseções a seguir, são constituídos por oito bases de dados de imagens, bem como dezenove métodos objetivos voltados à FR-IQA do estado da arte, além do próprio método PEL proposto.

Tabela 3 – Comparação das oito bases de dados de imagens para IQA utilizadas.

Base de Dados de Imagens (BDI)	Ano	Avaliadores	Imagens de Referência	Imagens Distorcidas	Tipos de Distorções	Resolução (pixel)
IVC [123]	2005	15	10	185[a]	4	512x512
LIVE [12]	2005	20 - 29	29	779	5	618x453 - 768x512[b]
CSIQ [134]	2010	25	30	866	6	512x512
MICT [124]	2011	27	14	182	2	768x512
VCL [135]	2012	118	23	552	4	683x512 - 768x512[c]
TID13 [136]	2013	971	25	3.000	24	512x384
CIDIQ [137]	2014	17	23	690	6	800x800
KADID [138]	2019	2.209	81	10.125	25	512x384

[a] Somente imagens coloridas foram utilizadas.
[b] 640x512, 610x488, 610x488, 627x482, 634x505, 480x720, 634x438 e 632x505.
[c] 512x684, 512x771, 512x768 e 771x512.

4.2.1 Bases de Dados de Imagens (BDIs)

As BDIs IVC [123], LIVE [12], CSIQ [134], MICT [124], VCL [135], TID13 [136], CIDIQ [137] e KADID [138], amplamente utilizadas na literatura, foram usadas nos experimentos para avaliar o desempenho dos métodos. Cada BDI contém tanto imagens de referência quanto imagens distorcidas, bem como suas pontuações subjetivas correspondentes, sendo tais pontuações medidas em Mean Opinion Score (MOS) ou Differential Mean Opinion Score (DMOS).

Informações resumidas sobre essas oito bases de dados de imagens são descritas nas subseções a seguir, mas a Tabela 3 apresenta uma visão geral sobre tais informações, onde no total há 235 imagens de referência (ou originais) e 16.379 distorcidas (ou de testes). A Tabela 5 apresenta os tipos de distorção em cada BDI. Vale destacar, no entanto, como em [126] e [75], que a comparação de desempenho em tipos

Tabela 4 – Tipos de Distorções em cada BDI*

IV	LI	CS	MI	VC	TI	CI	ES	KA	NI	Distortion Types
-	+	+	-	+	+	-	-	+	-	Additive white Gaussian noise
-	-	-	-	-	+	-	-	+	-	Noise in color components
-	-	-	-	-	+	-	-	-	-	Spatially correlated noise
-	-	-	-	-	+	-	-	+	+	Masked noise or Color quantiz.
-	-	-	-	-	+	-	+	-	-	High frequency noise
-	-	-	-	-	+	-	-	+	-	Impulse noise
-	-	-	-	-	+	-	-	-	-	Quantization noise
+	+	+	-	+	+	+	+	+	+	Gaussian blur
-	-	-	-	-	+	-	-	-	-	Image denoising
+	+	+	+	+	+	+	+	+	+	JPEG compression
+	+	+	+	+	+	+	-	+	+	JPEG2000 compression
-	-	-	-	-	+	-	-	-	+	JPEG transmission errors
-	-	-	-	-	+	-	-	-	-	JPEG2000 transmission errors
-	-	-	-	-	+	-	-	+	-	Non eccentricity pattern noise
-	-	-	-	-	+	-	-	+	-	Local block-wise distortions
-	-	-	-	-	+	-	-	+	-	Mean shift
-	-	+	-	-	+	+	-	+	+	Contrast change
-	-	-	-	-	+	+	-	+	-	Change of color saturation
-	-	-	-	-	+	-	-	+	+	Multiplicative Gaussian noise
-	-	-	-	-	+	-	-	-	-	Comfort noise
-	-	-	-	-	+	-	-	-	-	Lossy compression of noisy images
-	-	-	-	-	+	-	-	-	-	Image color quantization
-	-	-	-	-	+	-	-	-	-	Chromatic aberrations
-	-	-	-	-	+	-	-	-	-	Sparse sampling and reconstr.
+	-	-	-	-	-	-	-	-	-	LAR coding
-	+	-	-	-	-	-	-	-	-	Fast fading rayleigh
-	-	+	-	-	-	-	-	-	-	Additive pink Gaussian noise
-	-	-	-	-	-	+	-	-	-	Poisson noise
-	-	-	-	-	-	-	-	+	-	Lens blur - circular kernel
-	-	-	-	-	-	-	-	+	+	Motion blur - line kernel
-	-	-	-	-	-	-	-	+	-	Color shift - randomly translate
-	-	-	-	-	-	-	-	+	-	Color saturation 1 with HSV
-	-	-	-	-	-	-	-	+	-	Denoise - Gaussian white noise
-	-	-	-	-	-	-	-	+	-	Brighten - non-linearly adjust
-	-	-	-	-	-	-	-	+	-	Darken - similar to brighten
-	-	-	-	-	-	-	-	+	-	Jitter- bicubic interpolation
-	-	-	-	-	-	-	-	+	+	Pixelate - downsize and upsize
-	-	-	-	-	-	-	-	+	-	Quantization - Otsus method
-	-	-	-	-	-	-	-	+	-	High sharpen - unsharp masking
-	-	-	-	-	-	-	+	-	-	Interpolation - down and upsampling
-	-	-	-	-	-	-	+	-	-	Ringing

* IV: IVC, LI: LIVE, CS: CSIQ, MI: MICT, VC: VCL, TI: TID13, CI: CIDIQ, ES: ESPL, KA: KADID, NI: NITS.

individuais de distorções não é abordada no primeiro momento desta tese, ficando apenas para a versão final.

4.2.1.1 BDI: IVC

IVC [123] é uma BDI desenvolvida na Universidade de Nantes, França, em 2005. Ela consiste em 10 imagens de referência no formato 'bmp', 185 imagens distorcidas, quatro tipos de distorção e sua avaliação subjetiva, realizada por meio do método DSIS, é avaliada por 15 indivíduos. Todas as imagens possuem uma resolução de 512x512 píxeis. O MOS é calculado a partir da pontuação de qualidade individual. A escala de qualidade mostrada ao observador humano para a classificação varia de 1 a 5. A Fig. 33 apresenta as imagens de referência de IVC.



Figura 33 – Imagens originais da BDI IVC [123].



Figura 34 – Imagens originais da BDI LIVE [12].

4.2.1.2 BDI: LIVE

LIVE [12] é uma BDI com imagens coloridas, como podem ser vistas suas imagens de referência na Fig. 34, desenvolvida pela Universidade do Texas, Austin, em 2005. Ela consiste em 29 imagens de referência

no formato 'bmp', 779 imagens distorcidas e cinco tipos de distorção, e sua avaliação subjetiva, realizada por meio do método SSCQE, é avaliada por 20 a 29 indivíduos. Os indivíduos foram solicitados a fornecer sua percepção de qualidade em uma escala linear contínua, dividida em cinco regiões iguais marcadas com os adjetivos "Péssima", "Ruim", "Razoável", "Bom"e "Excelente". O DMOS também é calculado a partir da pontuação de qualidade individual.



Figura 35 – Imagens originais da BDI CSIQ [134].

4.2.1.3 BDI: CSIQ

CSIQ [134] é uma BDI com imagens coloridas, como podem ser vistas suas imagens de referência na Fig. 35, desenvolvida pela Oklahoma State University, EUA, em 2009. Ela consiste em 30 imagens de referência no formato 'png', 866 imagens distorcidas em 6 tipos diferentes de distorção, em quatro ou cinco níveis distintos. Os experimentos foram realizados, por meio do método SSCQ, sobre 5.000 avaliações subjetivas

a partir de 35 observadores diferentes. As imagens da CSIQ foram avaliadas subjetivamente com base na exibição sequencial em quatro monitores LCD calibrados, dispostos lado a lado e à mesma distância de observação em relação ao observador. Todas as imagens possuem uma resolução de 512x512 píxeis. O DMOS foi calculado a partir da pontuação de qualidade individual.



Figura 36 – Imagens originais da BDI MICT [124].

4.2.1.4 BDI: MICT

MICT [124] é uma BDI com imagens coloridas, como podem ser vistas suas imagens de referência na Fig. 36, desenvolvida pela Universidade de Toyama, Japão, em 2008. Ela consiste em 14 imagens de referência no formato 'bmp', 182 imagens distorcidas, sendo os dois tipos de distorção JPEG ou JPEG2000, e sua avaliação subjetiva, realizada por meio do método DSIS, é avaliada por 14 indivíduos. Todas as imagens possuem uma resolução de 768x512 píxeis. O MOS é calculado a partir da pontuação de qualidade individual. A escala de qualidade mostrada ao observador humano para a classificação varia de 1 a 5.

4.2.1.5 BDI: VCL

A BDI VCL@FER [135], chamada apenas de VCL nesta tese, consiste em 575 imagens no formato 'bmp', 23 das quais são originais, com 4 tipos diferentes de degradação por imagem e 6 graus de degradação. Os tipos de degradação são: JPEG2000, JPEG, Ruído gaussiano e Ruído branco. O teste foi conduzido em 118 indivíduos que tiveram que classificar 11.307 imagens no total (cerca de 96 imagens por observador). Cada imagem foi avaliada entre 16 e 36 vezes por meio do método SSCQE. A pontuação média de opinião (MOS) foi calculada conforme a Rec. ITU-R BT.500-11. A Fig. 37 apresenta as imagens de referência de VCL.

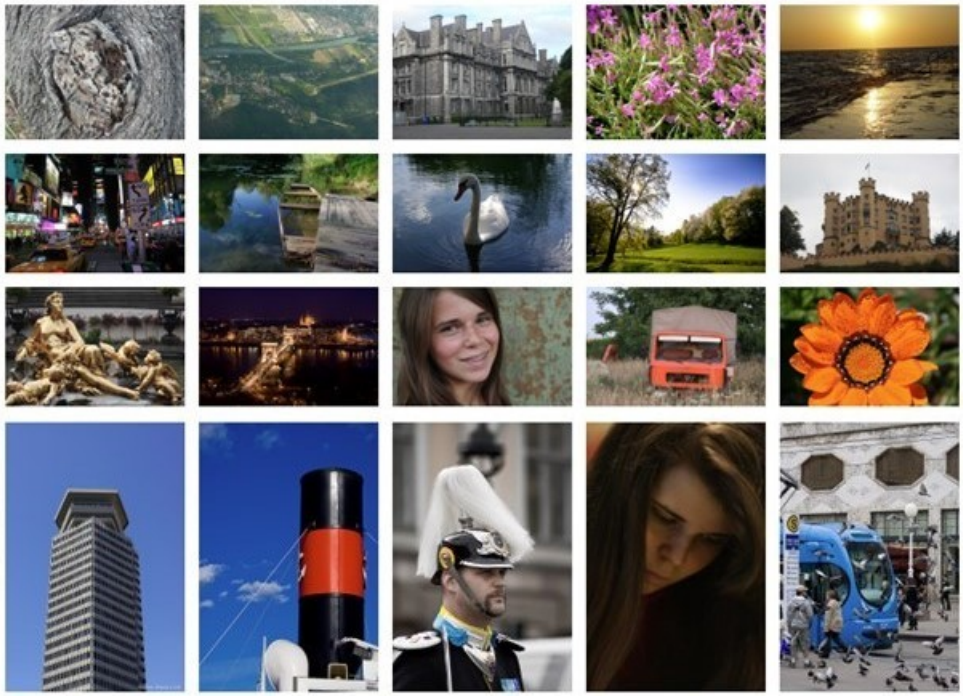


Figura 37 – Imagens originais da BDI VCL [135].



Figura 38 – Imagens originais da BDI TID2013 [136].

4.2.1.6 BDI: TID2013

TID2013 [136] é uma BDI com imagens coloridas, cujas imagens de referência podem ser vistas na Fig. 38, foi desenvolvida pela Universidade de Tecnologia de Tampere, Finlândia, em 2013. Ela consiste em 25

imagens de referência no formato 'bmp', 3.000 imagens distorcidas e 24 tipos de distorção, e sua avaliação subjetiva, realizada por meio do método DSIS, é avaliada por 971 indivíduos. O MOS é então calculado a partir da pontuação de qualidade individual. Todas as imagens possuem uma resolução de 512x384 píxeis. A escala de qualidade mostrada ao observador humano para a classificação varia de 0 a 9.

4.2.1.7 BDI: CIDIQ

A BDI Colourlab Image Database: Image Quality (CID:IQ) [137], simplificado aqui como CIDIQ, contém 23 imagens pictóricas selecionadas como imagens de referência com seis distorções diferentes em 5 níveis, totalizando 690 imagens distorcidas, sendo ambas no formato 'bmp'. As distorções são classificadas como compressão JPEG, compressão JPEG2000, ruído de Poisson, desfoque e dois algoritmos de mapeamento de gama. A resolução de todas as imagens é de 800 por 800 píxeis. O CIDIQ contém as pontuações subjetivas de 17 indivíduos. A Fig. 39 apresenta as imagens de referência de CIDIQ.



Figura 39 – Imagens originais da BDI CIDIQ [137].

4.2.1.8 BDI: KADID

KADID-10k [138] é uma BDI com suas imagens distorcidas artificialmente, desenvolvida pela Universidade de Konstanz, Alemanha, em 2019. Ela consiste em 81 imagens de referência no formato 'png', 10.125 imagens distorcidas e 25 tipos de distorção, onde tais imagens foram avaliadas subjetivamente por 2209 indivíduos, realizada por meio do método SSCQE. Com base na avaliação, o DMOS é calculado. A escala de qualidade usada varia de 1 a 5. A Fig. 40 apresenta as imagens de referência de KADID.

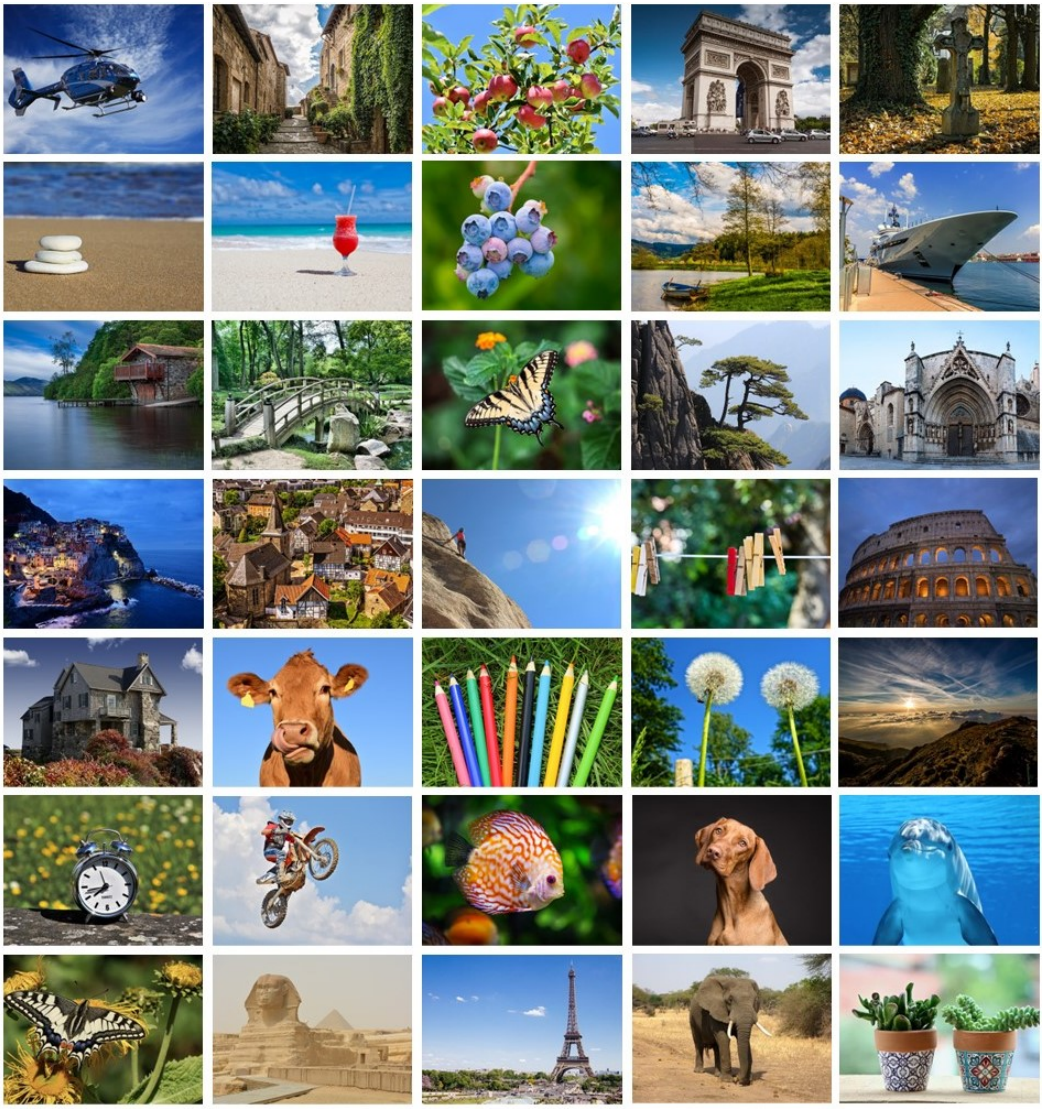


Figura 40 – Amostra das imagens originais da BDI KADID-10K [138].

4.2.2 Métodos Objetivos Comparados

Os resultados do método proposto nesta tese são comparados, sob as oito BDI descritas na seção anterior, com dezenove métodos FR-IQA do estado da arte, incluindo SSIM [21], MS-SSIM [22], FSIM$_C$ [24], GMSD [71], VSI [120], MCSD [16], SCQI [126], PSIM [125], RVSIM [73], SSRM [127], HaarPSI [128], EFS [74], SPSIM [75], CVSS [72], SSVD [76], SUMMER [133], VCGS [77], DSSIM [129] e SSVS [78].

4.3 Verificação das Habilidades Perceptuais dos Métodos

Para se verificar as habilidades perceptuais do método proposto, foram realizadas avaliações de sua eficiência (custo computacional, Seção 4.3.1) e de sua eficácia (Análise de Desempenho, Seção 4.3.2) em relação aos dezenove (19) métodos do estado da arte sob oito (8) bases de dados de imagens voltadas à FR-IQA. Além disso, um ponto bastante relevante quanto à escolha por um método é seu custo-benefício,

que muitas vezes é ignorado por diversas pesquisas. Nesta tese, o custo-benefício foi estabelecido na Seção 4.3.3, onde é verificada a correlação da eficiência do método em relação a sua eficácia.

4.3.1 Custo Computacional

Como mencionado na Seção 1.1, a eficiência se refere ao custo computacional, geralmente medido pelo tempo médio de execução de cada imagem, ou pela complexidade do algoritmo, sendo a primeira forma mais usual.

Para facilitar a leitura dos resultados quanto à classificação ascendente do custo computacional para os dezenove (19) métodos do estado da arte, bem como para o método proposto, foram criados índices subscritos de 1 (melhor) a 20 (pior), idealizados nesta tese. Vale ressaltar que, na literatura, geralmente não é evidenciada essa ordem nas tabelas, deixando a cargo do próprio leitor classificar os resultados de cada método.

4.3.2 Análise de Desempenho

Com base nos procedimentos recomendados pelo Video Quality Experts Group (VQEG) [10], [11], três etapas foram aplicadas para analisar o desempenho dos vinte métodos. A primeira etapa, descrita na Seção 4.3.2.1, consiste em mapear as previsões dos métodos em uma escala subjetiva. A segunda etapa, descrita na Seção 4.3.2.2, utiliza métricas estatísticas para avaliar o desempenho de cada método. Por fim, a significância estatística dos resultados é analisada na Seção 4.3.2.3.

4.3.2.1 Valores DMOS/MOS e Mapeamento para a Escala Subjetiva

O mapeamento de dados requer o uso de uma etapa de mapeamento não-linear. Portanto, para remover qualquer não linearidade decorrente do processo de classificação subjetiva, bem como para facilitar a comparação de métodos em um domínio comum, a relação entre cada previsão do método e a classificação subjetiva correspondente deve ser estimada. Para isso, uma regressão não-linear entre o conjunto de classificações de qualidade de imagem (image quality ratings, IQRs) e o DMOS correspondente (ou MOS, conforme a BDI) deve ser realizada [11].

A função de regressão logística foi empregada para a regressão não-linear por todos os métodos comparados nesta tese. Embora válida e amplamente utilizada, a função logística é uma recomendação obsoleta do VQEG publicada em agosto de 2003 [10]. Em nosso trabalho, a função polinomial cúbica recomendada pelo VQEG no Relatório Final de setembro de 2008 [11] foi aplicada. O polinômio cúbico é definido da seguinte forma:

$$DMOS_p = ax^3 + bx^2 + cx + d \tag{4.1}$$

onde $DMOS_p$ é o $DMOS$ previsto. Os pesos a, b e c e a constante d são obtidos ajustando a função aos dados $[DMOS, IQR]$.

Para fins de entendimento quanto aos resultados dos mapeamentos dos dados, categorizou-se nesta tese uma comparação em nível de *visão específica* ou *geral*.

A *visão específica* se refere a uma comparação micro ou individualizada dos valores normalizados de um determinado método de IQA em relação aos resultados subjetivos de um determinado BDI, representado, respectivamente, pelo *scores* objetivos em relação aos *scores* subjetivos. Tal comparação pode ser realizada por meio de um gráfico 2D, no formato de gráfico de dispersão (*scatter*), onde em um eixo são exibidos os *scores* objetivos, e em outro eixo, perpendicularmente, são exibidos *scores* subjetivos. A Fig. 41 apresenta um exemplo do formato de gráfico de dispersão dos scores objetivos do método SSIM em relação aos scores subjetivos da BDI LIVE.

A *visão geral* se refere a uma comparação macro ou ampliada da qualidade da distribuição de

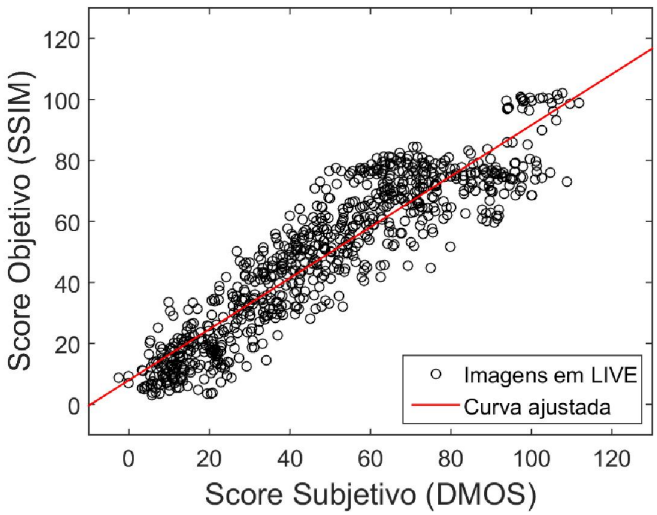


Figura 41 – Exemplo de formato de gráfico de dispersão (*scatter*) voltado à visão específica quanto ao entendimento da relação entre *scores* objetivos do método SSIM e *scores* subjetivos da BDI LIVE.
Fonte: Autoria própria.

dados entre os valores normalizados de um conjunto de métodos de IQA e os resultados subjetivos de uma determinada BDI. Nesse sentido, enquanto a visão específica opta por um gráfico no formato *scatter*, a visão geral usa um gráfico no formato *box plot*. Embora o uso de diagramas de dispersão tenha se tornado prática padrão em diversos estudos [24], [73], [74–77], obter uma análise de comparação de desempenho mais detalhada em relação à qualidade de distribuição dos dados e, ao mesmo tempo, abrangente quanto ao conjunto de métodos envolvidos é viável por meio da utilização de um gráfico *box plot*. Este formato permite que haja uma consolidação dos dados de todos os métodos em um único gráfico para cada BDI, permitindo assim uma análise comparativa ampla do desempenho desses métodos. A Fig. 42 apresenta, à esquerda na imagem, um exemplo do relacionamento DMOSxDMOS$_p$ subjetivo versus pontuações obtidas por cinco métodos FR-IQA (PSIM, HaarPSI, SPSIM, SSVS e PEL) para todas as

imagens de testes da BDI CSIQ em [79].

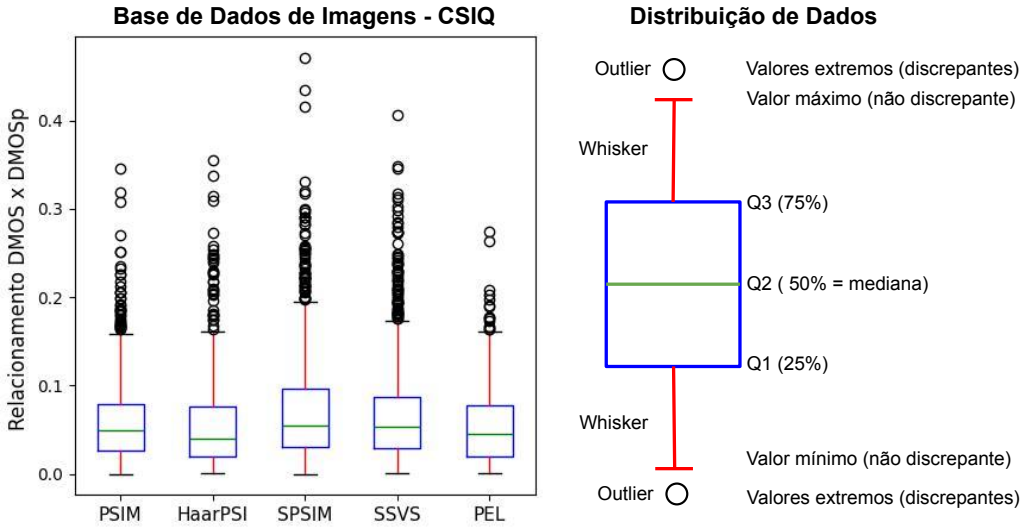


Figura 42 – Exemplo de *box plot* (gráfico de caixa), à esquerda, Relacionamento de DMOSxDMOS$_p$ subjetivo versus pontuações obtidas por cinco métodos FR-IQA para todas as distorções da BDI CSIQ em [79] e, à direita, uma imagem genérica de um *box plot* com todos os seus componentes.
Fonte: Autoria própria.

O *box plot*, também conhecido como diagrama de caixa e bigode, tornou-se a técnica padrão para apresentar cinco valores, consistindo dos quartis mínimo, máximo, mediana e inferior/superior de qualquer base de dados [139], [140]. Este conjunto de valores, como mostrado na Fig. 42, é uma maneira rápida de resumir a distribuição de dados, onde, conforme a imagem à direita, a parte inferior (quartil inferior, Q1) e a parte superior (quartil superior, Q2) de cada caixa azul são os percentis 25 e 75, respectivamente. O interior desta caixa indica o intervalo interquartil (IQR), que é a área entre os quartis superior e inferior (ou seja, IQR = Q3-Q1) e consiste em 50% da distribuição. A linha verde central dentro de cada caixa é a mediana (Q3). As linhas vermelhas que se estendem da caixa central até os pontos de dados mais extremos com hachuras colocadas no final são chamadas de *whiskers*, que são os valores mínimo e máximo na BDI. Finalmente, os *outliers* são considerados valores discrepantes, representados pelo símbolo preto "○" [140–143].

4.3.2.2 Métricas Estatísticas de Predição da Avaliação

O desempenho dos métodos de IQA, conforme os relatórios técnicos do VQEG [10], é avaliado em relação à sua capacidade de prever (ou predizer) a avaliação subjetiva da qualidade de uma imagem com base em três habilidades perceptuais: acurácia, monotonicidade e consistência [113].

- *Acurácia*: representa a capacidade de prever as classificações subjetivas de qualidade com baixo

erro;

- *Monotonicidade*: representa o grau em que as previsões do método concordam com as magnitudes relativas das classificações subjetivas de qualidade; em outras palavras, as pontuações objetivas de qualidade devem ser monotônicas em sua relação com as pontuações subjetivas; e

- *Consistência*: representa o grau em que um IQM mantém a precisão em toda a gama de imagens de teste. Vale a pena destacar que um método é considerado *robusto* quando sua consistência permanece diante de uma variedade de distorções, especialmente quando tal variedade pertença a uma quantidade de banco de imagens mais diversificada possível.

Quatro métricas estatísticas de predição da avaliação são utilizadas para comparar o desempenho dos métodos de IQA nesta pesquisa: o Coeficiente de Correlação RankOrder de Spearman (SROCC), o Coeficiente de Correlação RankOrder de Kendall (KROCC), o Coeficiente de Correlação Linear de Pearson (PLCC) e a Raiz do Erro Quadrático Médio (RMSE). As definições matemáticas dessas métricas estatísticas de avaliação podem ser encontradas na Seção 2.2.4 ou em [11], [71], [111]. Dentre essas métricas, conforme [10], [71], [111], [113], PLCC é adotada para avaliar a acurácia da predição; SROCC e KROCC são empregadas para avaliar a monotonicidade de predição e a RMSE é usada como uma medição da consistência de predição dos IQMs.

Em relação aos resultados apresentados por essas métricas, é importante destacar que um método melhor apresenta valores maiores de PLCC, SROCC e KROCC, porém menores de RMSE. Uma correspondência perfeita entre os escores objetivos mapeados (ou seja, $DMOS_p$ ou MOS_p) e os subjetivos, PLCC = SROCC = KROCC = 1 e RMSE = 0 [110].

4.3.2.3 Significância Estatística

A significância estatística dos resultados é dada pela razão entre os RMSEs dos métodos de IQA, conforme segue:

$$\zeta = \frac{(RMSE_{MAX})^2}{(RMSE_{MIN})^2} \tag{4.2}$$

onde $RMSE_{MAX}$ e $RMSE_{MIN}$ são o maior e o menor RMSE, respectivamente. O parâmetro ζ é avaliado com base na função de distribuição F (*F-distribution function*), que possui um $F_{critical}$ de 5% e garante um nível de significância de 95%. Se ζ for maior que o $F_{critical}$, então há uma diferença significativa entre os valores de RMSE [11]. Da mesma forma, a distribuição F em porcentagem, $F_{\%} = 100 \cdot (\zeta - 1)$, fornece o nível de significância absoluto. Dessa forma, dois métodos de IQA podem ser considerados estatisticamente diferentes para valores de ζ maiores que 1,05, $\zeta > 1,05$ (ou $F_{\%} > 5\%$). Mais detalhes sobre a significância estatística dos resultados podem ser encontrados em [11].

4.3.3 Custo Computacional versus Análise de Desempenho

Nas Seções 4.3.1 e 4.3.2 foram considerados dois pontos importantes quanto a avaliações, a saber, o custo computacional e a análise de desempenho. Apesar da grande importância de tais avaliações serem realizadas de maneira separada, sabe-se que é de suma importância compará-las correlacionadas. Neste contexto, como descrito na Seção 1.1, os métodos, em geral, não conseguem obter uma boa relação entre seu custo computacional e seu desempenho, notadamente conhecido como custo-benefício.

O custo-benefício de um método é verificado pelo seu custo computacional, representado nesta tese pela Velocidade de Execução (*frames* por segundo, fps), pelo seu desempenho, medido por meio de alguma métrica estatística (como SROCC, KROCC, PLCC, dentre outras). Tal comparação pode ser realizada por meio de um gráfico 2D, no formato de gráfico de dispersão (*scatter*), onde em um eixo são exibidos os valores devidamente normalizados de desempenho dos métodos, por meio de alguma métrica estatística (como SROCC, KROCC ou PLCC), e em outro eixo, perpendicularmente, são exibidos os valores dos custos computacionais, via fps, por exemplo. Essa comparação, em geral, é realizada considerando os 3 (Top-3) ou 5 (Top-5) primeiros métodos com melhor desempenho. A Fig. 43 apresenta um exemplo do formato de gráfico de dispersão de Top-5 de desempenho a partir da métrica SROCC pela velocidade de execução para oito BDIs, onde quanto maior for o valor de SROCC pelo número de fps, melhor a relação custo-benefício.

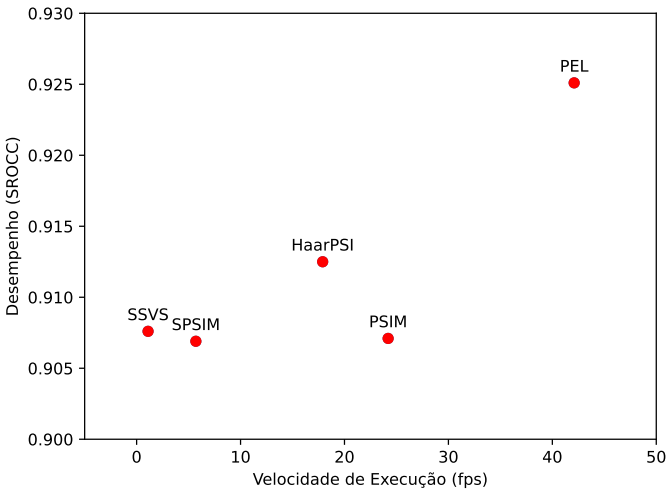


Figura 43 – Exemplo de formato de gráfico de dispersão (*scatter*) utilizado em análise específica de Análise de Desempenho versus Custo Computacional.
Fonte: Adaptado de [79].

5

MÉTODO ANALÍTICO PERCEPTUAL PROPOSTO

Este capítulo apresenta uma visão geral do método analítico perceptual proposto, denominado Logaritmo do Erro Perceptual (Perceptual Error Logarithm, PEL), para avaliação da qualidade de imagem com referência completa, bem como descreve todas as etapas do seu processo computacional.

5.1 Visão Geral

Esta seção apresenta o método analítico proposto para avaliação de qualidade de imagem de referência completa, denominado Logaritmo de Erro Perceptual (Perceptual Error Logarithm, PEL). A Fig. 44 mostra o processo computacional do método PEL. r e t representam as imagens de referência e teste (distorcidas), respectivamente. Inicialmente, mapas de luminância (Y_r/Y_t) e crominância (I_r/I_t e Q_r/Q_t) são gerados após a redução e conversão de cor espacial aplicadas às imagens de entrada. Por meio do uso de mapas de crominância e luminância, mapas de diferença absoluta simples (AD_I, AD_Q e AD_L) são gerados, bem como mapas de características especiais (S_G, LE_G, AD_G, AD_G e C_G). Os mapas especiais são derivados de extrações de características perceptuais aplicadas a mapas de magnitude de gradiente (GM_r e GM_t). Os mapas GM_r/GM_t são gerados a partir dos mapas LSD_r e LSD_t, que são gerados imediatamente após a aplicação da filtragem de desvio padrão local (LSD) aos mapas de luminância. Finalmente, para calcular a pontuação PEL, uma função logarítmica é aplicada ao mapa de erro perceptual (PE), que combina os mapas simples e especiais, bem como ajusta essa combinação com base no peso dos mapas de similaridade de superpixel (W_{SP}) e energia local (W_{LE}). Detalhes sobre cada etapa são fornecidos na seção a seguir.

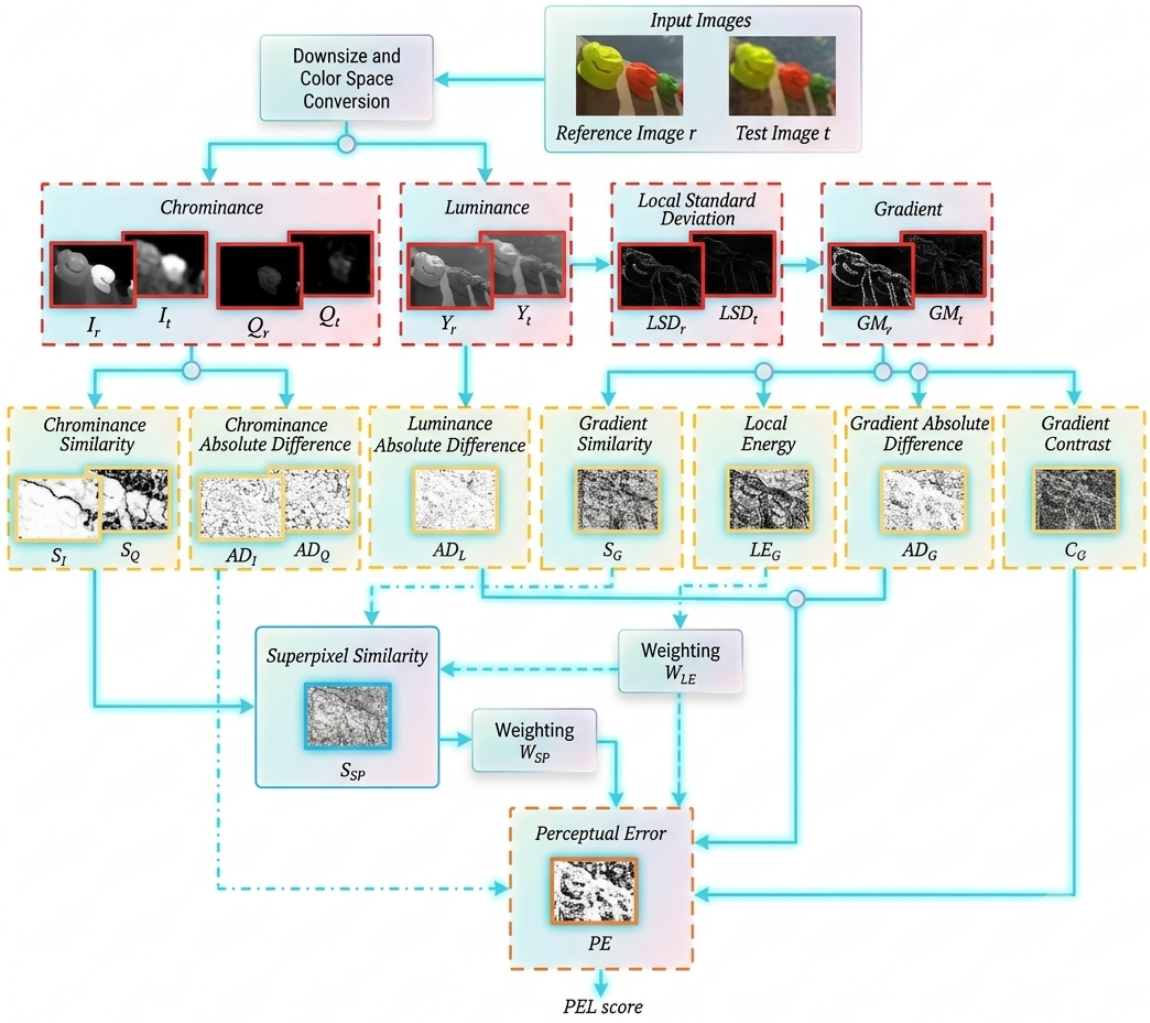


Figura 44 – Fluxograma do processo computacional para o método analítico perceptual proposto voltado à FR-IQA: Perceptual Error Logarithm (PEL).
Fonte: Autoria própria.

5.2 Processo Computacional

5.2.1 Pré-processamento

A redução do tamanho da imagem e a conversão do espaço de cores são duas etapas iniciais importantes no processo computacional de métodos objetivos. A redução do tamanho da imagem adiciona eficiência ao método, o que reduz o tempo necessário para processar as etapas seguintes. Todas as imagens têm suas dimensões reduzidas para 256 segundo a escala empírica proposta por [144]. Em relação à conversão do espaço de cores, as imagens coloridas RGB são transformadas para o espaço YIQ com um mapeamento linear [145] definido como:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{5.1}$$

onde Y é o canal de luminância para avaliar as distorções estruturais das imagens, I e Q são dois canais de crominância usados para caracterizar a degradação da qualidade da imagem causada por distorções de cor [74].

Na Fig. 44, Y_r e Y_t são os mapas de luminância de referência e teste (distorcidos), respectivamente, enquanto $I_r(I_t)$ e $Q_r(Q_t)$ são, respectivamente, os canais cromáticos I and Q, que pertencem à imagem r (t).

5.2.2 Desvio Padrão Local

O Desvio Padrão Local (*Local Standard Deviation*, LSD) é amplamente utilizado em muitas aplicações devido às informações de textura, como em redes de sensores [146], extração de características [147], imagens médicas [148], [149], aplicações de vigilância [150] e avaliação da qualidade de imagem [151], [152]. Nesta tese, uma função foi aplicada aos mapas de luminância, que realizou a filtragem do desvio padrão local da imagem Y e retornou o LSD da imagem filtrada, como segue:

$$LSD(i) = \Big(\frac{1}{W-1}\sum_{i=1}^{W}(Y(i) - \overline{Y}_i)^2\Big)^{1/2}, \tag{5.2}$$

onde o valor de cinza de cada píxel de saída $LSD(i)$ é o desvio padrão local da vizinhança (W) de 3 por 3 em torno do valor de cinza correspondente de cada píxel de entrada $Y(i)$, em que $\overline{Y}_i$ denota o valor médio na vizinhança. O desvio padrão local dos mapas de luminância de referência (Y_r) e de teste (Y_t) são definidos como LSD_r e LSD_t, respectivamente.

5.2.3 Magnitude do Gradiente

Bordas representam variações locais significativas em imagens e servem como características importantes para análise de imagens [153]. Se uma borda for definida como uma mudança de nível de cinza, então a derivada, ou gradiente, é uma base natural para um operador de borda [154]. Existem vários operadores diferentes para calcular o gradiente da imagem, a saber: Sobel, Roberts, Prewitt e Scharr [77].

O gradiente é amplamente utilizado para avaliação da qualidade de imagem, pois captura mudanças de estrutura e contraste [34]. Uma propriedade importante associada aos gradientes é sua magnitude [153]. Nesta tese, o operador de Prewitt foi utilizado para calcular a magnitude do gradiente (*Gradient Magnitude*, GM), dada por:

$$GM(i) = \sqrt{G_h^2(i) + G_v^2(i)} \tag{5.3}$$

onde $G_h(i)$ e $G_v(i)$ são usados para extrair gradientes horizontais e verticais da imagem, respectivamente, e i representa o i-ésimo píxel.

Na literatura, como por exemplo em [24], [73], [74–77], a magnitude do gradiente é aplicada

diretamente aos mapas de luminância, ou seja, logo após a etapa de pré-processamento. Por outro lado, o método proposto nesta tese aplica a magnitude do gradiente aos mapas gerados pela função de desvio padrão local. Com isso, o método PEL obteve resultados mais correlacionados com o julgamento humano.

Para entender a razão dessa melhora, a seguir são apresentados resultados escalares, na Tabela 5, e ilustrativos, na Fig. 45. Na Tabela 5, que registra a utilização de 4 métricas estatísticas e 8 BDIs utilizados nesta tese, totalizando 32 avaliações, são apresentados resultados na primeira coluna, nominada como PEL*, quando não se aplica a etapa LSD no processo do método PEL, e na segunda coluna, nominada como PEL, quando aplica-se tal etapa. Nesse sentido, nota-se a vantagem, quanto à eficácia, de se utilizar a etapa de LSD entre o pré-processamento e a etapa de Gradiente, pois foram alcançadas 27 avaliações pelo que se denota de PEL, enquanto PEL* alcançou os cinco primeiros lugares, totalizando apenas 15% do total de 32 avaliações. Vale ressaltar que a única vantagem de PEL* é, praticamente, em relação ao BDI MICT, que contém apenas imagens distorcidas com JPEG e JPEG2000. Quem sabe em um contexto apenas com distorções dessa natureza fosse vantajoso não usar o LSD como etapa intermediária, mas com diversas outras distorções existentes nas outras BDI não foi o caso.

Tabela 5 – Tipos de Distorções em cada BDI.

BDI	Métrica	PEL*	PEL
IVC 2005	SROCC↑	$\mathbf{0.9125}_2$	$\mathbf{0.9438}_1$
	KROCC↑	$\mathbf{0.7450}_2$	$\mathbf{0.7899}_1$
	PLCC↑	$\mathbf{0.9201}_2$	$\mathbf{0.9513}_1$
	RMSE↓	$\mathbf{0.4824}_2$	$\mathbf{0.3796}_1$
LIVE 2005	SROCC↑	$\mathbf{0.9529}_2$	$\mathbf{0.9657}_1$
	KROCC↑	$\mathbf{0.8092}_2$	$\mathbf{0.8374}_1$
	PLCC↑	$\mathbf{0.9556}_2$	$\mathbf{0.9665}_1$
	RMSE↓	$\mathbf{8.0681}_2$	$\mathbf{7.0276}_1$
CSIQ 2010	SROCC↑	$\mathbf{0.9387}_2$	$\mathbf{0.9627}_1$
	KROCC↑	$\mathbf{0.7782}_2$	$\mathbf{0.8264}_1$
	PLCC↑	$\mathbf{0.9435}_2$	$\mathbf{0.9672}_1$
	RMSE↓	$\mathbf{0.0872}_2$	$\mathbf{0.0669}_1$
MICT 2011	SROCC↑	$\mathbf{0.9637}_1$	$\mathbf{0.9536}_2$
	KROCC↑	$\mathbf{0.8410}_1$	$\mathbf{0.8154}_2$
	PLCC↑	$\mathbf{0.9676}_1$	$\mathbf{0.9572}_2$
	RMSE↓	$\mathbf{0.2893}_1$	$\mathbf{0.3319}_2$
VCL 2012	SROCC↑	$\mathbf{0.9488}_1$	$\mathbf{0.9483}_2$
	KROCC↑	$\mathbf{0.7925}_2$	$\mathbf{0.7935}_1$
	PLCC↑	$\mathbf{0.9502}_2$	$\mathbf{0.9518}_1$
	RMSE↓	$\mathbf{7.6805}_2$	$\mathbf{7.5579}_1$
TID13 2013	SROCC↑	$\mathbf{0.8594}_2$	$\mathbf{0.8796}_1$
	KROCC↑	$\mathbf{0.6694}_2$	$\mathbf{0.6981}_1$
	PLCC↑	$\mathbf{0.8780}_2$	$\mathbf{0.9020}_1$
	RMSE↓	$\mathbf{0.5939}_2$	$\mathbf{0.5357}_1$
CIDIQ 2014	SROCC↑	$\mathbf{0.8610}_2$	$\mathbf{0.8672}_1$
	KROCC↑	$\mathbf{0.6797}_2$	$\mathbf{0.6901}_1$
	PLCC↑	$\mathbf{0.8643}_2$	$\mathbf{0.8700}_1$
	RMSE↓	$\mathbf{0.7553}_2$	$\mathbf{0.7404}_1$
KADID 2019	SROCC↑	$\mathbf{0.8471}_2$	$\mathbf{0.8762}_1$
	KROCC↑	$\mathbf{0.6557}_2$	$\mathbf{0.6908}_1$
	PLCC↑	$\mathbf{0.8463}_2$	$\mathbf{0.8752}_1$
	RMSE↓	$\mathbf{0.5768}_2$	$\mathbf{0.5238}_1$

* PEL sem a etapa de LSD (Local Standard Deviation).
Fonte: Autoria própria.

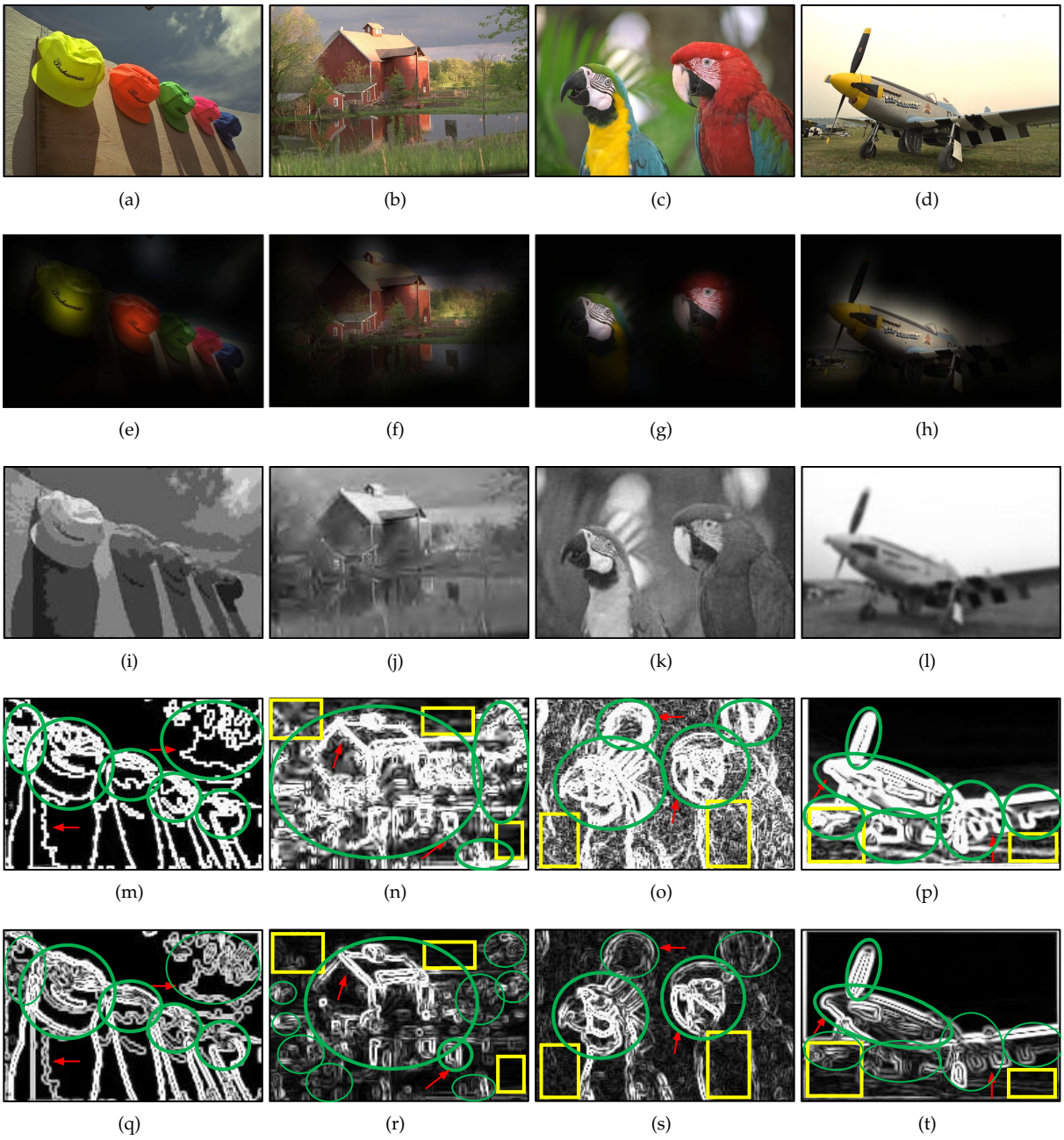


Figura 45 – (a-d) são imagens de referência, originais da BDI LIVE [12]. (e-h) são mapas com região de interesse (Region of Interest, RoI) em [122] correspondentes, respectivamente, às imagens de referência de (a-d). (i-l) são imagens distorcidas a partir de suas originais (caps, house, parrots e plane) da mesma BDI LIVE [12], que foram corrompidas, respectivamente, com JPEG (i), JPEG2000 (j), ruído gaussiano branco (k) e gaussiano bidimensional circular simétrico (l). (m-p) são mapas de magnitude de gradiente das imagens (i-l) normalmente usados após pré-processamento por métodos da literatura. (q-t) são mapas de magnitude de gradiente das imagens (i-l) gerados pelo método PEL após a etapa de desvio padrão local. As setas vermelhas, os retângulos amarelos e os círculos verdes comparam os resultados apresentados pelos mapas (m-p) e (q-t) em relação às distorções, respectivamente, nas bordas, nas regiões de segundo plano e de primeiro plano. Vale ressaltar que os círculos verdes foram coincidentemente, em grande parte, posicionados nas imagens de (m-p) e (q-t) para efeitos de comparação em relação aos mapas de RoI destacados de (e-h). Os círculos verdes densos destacam a presença de bordas fortes, enquanto os menos densos destacam bordas fracas, respectivamente, representando dessa forma maior ou menor percepção da RoI por parte de métodos da literatura ou do proposto nesta tese.
Fonte: Autoria própria.

Além dos dados escalares, a Fig. 45 apresenta alguns resultados ilustrativos advindos da comparação visual dessas duas formas. As Fig. 45(a) a 45(d) são imagens de referência, originais da BDI LIVE [12]. As Fig. 45(e) a 45(h) são mapas com região de interesse (*Region of Interest*, RoI[16])) em [122] correspondentes, respectivamente, às imagens de referência da Fig. 45(a) a 45(d). Os conteúdos da Fig. 45(m) a 45(p) apresentam as imagens distorcidas de suas originais (caps, house, parrots e plane) na BDI LIVE [12], que foram corrompidas, respectivamente, com JPEG (Fig. 45(i)), JPEG2000 (Fig. 45(j)), ruído gaussiano branco (Fig. 45(k)) e gaussiano bidimensional circular simétrico (Fig. 45(l)). Os mapas de magnitude de gradiente, frequentemente gerados por métodos da literatura, como da Fig. 45(m) a 45(p), são resultados obtidos a partir das imagens distorcidas mostradas na Fig. 45(i) a 45(l). Já os mapas de magnitude de gradiente gerados nesta tese, em virtude do uso da camada intermediária LSD, são apresentados nas imagens da Fig. 45(q) a 45(t).

As comparações dessas imagens na Fig. 45 destacadas por setas, retângulos e círculos, nas duas formas, revelam distinções quanto à percepção de bordas distorcidas. Quanto a isso, Tang *et al.* [34] pontuam que o HVS tende a prestar mais atenção às bordas fortes da imagem. Nesse contexto, pode ser observado que as setas vermelhas da Fig. 45(m) a 45(p) destacam tanto as bordas originais quanto sua região ao redor. Por outro lado, as setas vermelhas da Fig. 45(q) a 45(t) destacam quase que exclusivamente as regiões distorcidas em torno das bordas, mas estas, em geral, não são computadas como distorção.

Os retângulos amarelos na Fig. 45 destacam, em segundo plano nas imagens, como as duas formas caracterizam a filtragem das bordas fortes e fracas. Nesse sentido, métodos da literatura, como os citados anteriormente, trabalham com as bordas tanto fracas quanto fortes, como destacadas pelo retângulo amarelo nas imagens dos mapas GM da Fig. 45(m) a 45(p). Isso acontece porque a magnitude do gradiente é aplicada diretamente em mapas de luminância. Por outro lado, o método PEL trabalha com mapas GM, oriundos dos mapas de LSD, que possuem, em grande parte, bordas fortes, como destacadas no retângulo amarelo da Fig. 45(q) a 45(t), enquanto as fracas são, em sua maioria, ignoradas.

Os círculos verdes na Fig. 45 destacam, em primeiro plano nas imagens, a RoI do HVS. Vale ressaltar que os círculos verdes foram coincidentemente, em grande parte, posicionados nas imagens da Fig. 45(m) a 45(p) e da Fig. 45(q) a 45(t) para efeitos de comparação em relação aos mapas de RoI destacados da Fig. 45(e) a 45(h). Os círculos verdes densos destacam a presença de bordas fortes, enquanto os menos densos destacam bordas fracas, respectivamente, representando dessa forma maior ou menor percepção da RoI por parte de métodos da literatura ou do proposto nesta tese.

Comparando os resultados apresentados nas imagens da Fig. 45(m) a 45(p) e da Fig. 45(q) a 45(t), pode-se notar que praticamente não existem diferenças significativas entre a Fig. 45(m) e a Fig. 45(q) quanto às RoI nas duas formas, como destacadas pelos círculos verdes, e corroborado pelo que [122] destaca. Um outro ponto observado é a pequena diferença entre as duas formas quanto à quantidade de bordas finas nas imagens corrompidas com JPEG. Por outro lado, existe uma grande diferença entre as duas formas quanto à quantidade de bordas finas apresentadas entre a Fig. 45(n) e a Fig. 45(r) oriundas

[16] RoI: veja Seção 2.3

de imagens corrompidas com JPEG2000. Neste contexto, a RoI utilizada pelos métodos da literatura é praticamente toda a imagem, como demonstrado pelos círculos verdes, diferente do que é destacado em [122]. Já a RoI utilizada pelo PEL está mais próxima do que é previsto em [122] quando consideradas as bordas fortes, mas, por outro lado, o método PEL diminui sua capacidade de ser um pouco mais eficaz por não considerar diversas bordas fracas pertencentes à RoI. Isso inclusive ajuda a explicar, conforme os dados da Tabela 5, o motivo pelo qual o método PEL*, que não utilizou a etapa de LSD, obteve melhor eficácia no BID MITC. Nesse sentido, então poderia ser considerado o uso de PEL* para cenários de aplicações que utilizem apenas imagens corrompidas com JPEG ou JPEG2000.

Diferente dos dois cenários comparados anteriormente, o método PEL, em geral, obteve vantagens por receber mapas que destacaram bordas fortes e desconsideraram bordas fracas em RoI quanto às imagens contendo as demais distorções. Para exemplificar tal afirmação, foram destacados círculos verdes nos mapas representados na Fig. 45(o) e na Fig. 45(s) conforme a RoI apontada por [122], oriundos da imagem da Fig. 45(k) que foi corrompida com ruído gaussiano branco. Nota-se que na Fig. 45(o) as bordas fortes e fracas extrapolam os círculos verdes, enquanto que na Fig. 45(s) são concentradas a maioria das bordas fortes na RoI, bem como são desconsideradas grande parte das bordas fracas, favorecendo uma tendência mais próxima da RoI apontada em [122]. Para imagens distorcidas com gaussiano bidimensional circular simétrico, as RoI destacadas nos mapas da Fig. 45(p) e na Fig. 45(t) seguem praticamente o mesmo comportamento do cenário anteriormente descrito.

5.2.4 Diferença Absoluta (Crominância, Luminância e Gradiente)

O conceito de diferença absoluta (*Absolute Difference*, AD) desempenha um papel fundamental na análise de imagens e na visão computacional. A AD de duas imagens é definida como a soma da diferença absoluta em cada píxel [155]. Embora sensível a variações de ruído e iluminação, o conceito fornece percepções valiosas sobre variações e mudanças na imagem [153], além de exigir menos recursos computacionais [155].

Nesta tese, a AD foi aplicada aos mapas de crominância (I_r/I_t e Q_r/Q_t), luminância (Y_r/Y_t) e gradiente (GM_r/GM_t), gerando, respectivamente, os mapas de Diferença Absoluta de Crominância (AD_I e AD_Q), Diferença Absoluta de Luminância (AD_L) e Diferença Absoluta de Gradiente (AD_G) conforme as seguintes equações:

$$AD_I(i) = |I_r(i) - I_t(i)|, \tag{5.4}$$

$$AD_Q(i) = |Q_r(i) - Q_t(i)|, \tag{5.5}$$

$$AD_L(i) = |Y_r(i) - Y_t(i)|, \tag{5.6}$$

$$AD_G(i) = |GM_r(i) - GM_t(i)|. \tag{5.7}$$

A Fig. 46 apresenta os resultados desses mapas gerados após a aplicação do AD a imagens da BDI LIVE. As duas primeiras colunas são as imagens de referência (caps, plane, parrots e lighthouse2) e suas imagens de teste correspondentes, corrompidas com Rayleigh de Desvanecimento Rápido (*Fast Fading Rayleigh*), Borramento Gaussiano (*Gaussian Blur*), JPEG2000 e JPEG, respectivamente.

Vários testes empíricos foram conduzidos para encontrar a combinação mais precisa possível para os pares de mapas AD_I/AD_Q e AD_L/AD_G nesta tese. Isso se deve ao fato de que, individualmente, esses mapas, como mostrado nas imagens das colunas três, quatro, seis e sete da Fig. 46, não forneceram resultados significativos que pudessem contribuir para o aprimoramento do método proposto. Nesse sentido, as combinações AD_IxAD_Q, exemplificadas nos mapas da quinta-coluna da Fig. 46, apresentam notavelmente resultados visuais que contribuíram positivamente para o aprimoramento da percepção das regiões coloridas das imagens. Esses resultados provavelmente foram aprimorados devido à combinação originada de dois canais de crominância diferentes. Além disso, vale destacar que as combinações AD_IxAD_Q conseguiram melhorar a RoI a ser percebida pelo HVS nas imagens.

5.2.5 Energia Local e Constrate

Para melhorar a eficácia do método proposto sem comprometer sua eficiência, nesta tese foi considerado o uso de mapas de características com baixa complexidade de processamento, como *energia local* e *contraste*, mas que ainda podem extrair características altamente informativas.

A energia local fornece informações sobre a homogeneidade da imagem. Ela apresenta valores baixos quando as probabilidades dos pares de níveis de cinza são bastante semelhantes, e valores altos caso contrário [156]. Então, o mapa de energia local proposto nesta tese é dado por:

$$LE(i) = \frac{GM_t(i)}{\left(GM_r(i)^2 + GM_t(i)^2 + AD_G(i)^2 + K\right)^{1/2}}, \tag{5.8}$$

onde GM_r e GM_t são mapas de referência e de magnitude de gradiente de teste, AD_G é especificado em (5.7), K é uma constante positiva para manter a estabilidade da fração e seu valor proposto é 1.

É importante enfatizar que *LE na Fig. 47 representa os mapas de características de energia local existentes na literatura, que não utilizam o AD_G conforme especificado em (5.8). A razão para adicionar AD_G nesta tese foi para reduzir a intensidade dos valores derivados da energia local conforme o conteúdo das próprias imagens. O mapa de características LE_G proposto contribui efetivamente como um peso para ajustar tanto o mapa de Similaridade de Superpixels em (5.13) quanto o mapa de Erro Perceptual em (5.15). Portanto, o peso W_{LE} é o resultado do desvio padrão aplicado ao mapa LE_G, definido nesta tese como:

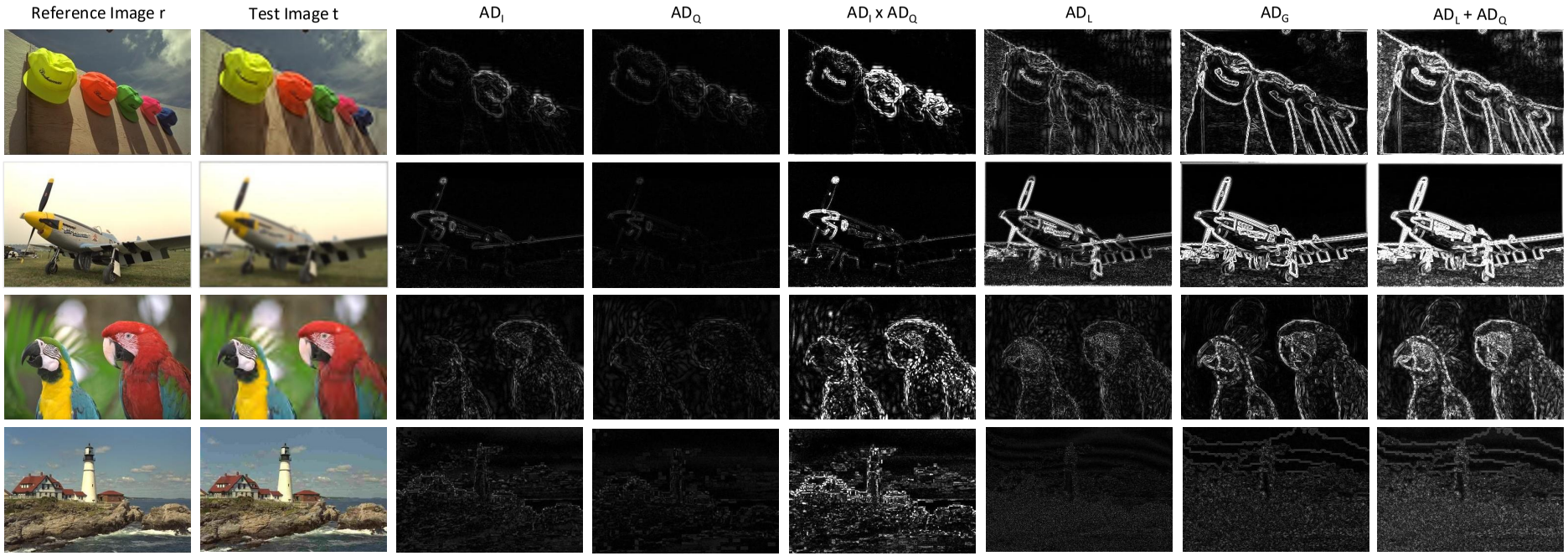


Figura 46 – Exemplos de Diferença Absoluta (AD) aplicada a mapas de características de crominância (I_r/I_t e Q_r/Q_t), luminância (Y_r/Y_t) e gradiente (GM_r/GM_t) em imagens da BDI LIVE [12]. As duas primeiras colunas são as imagens de referência (caps, plane, parrots e lighthouse2) e suas imagens de teste correspondentes (corrompidas com Rayleigh de Desvanecimento Rápido, Borramento Gaussiano, JPEG2000 e JPEG, respectivamente). AD_I e AD_Q: Diferenças Absolutas de Crominância; AD_L: Diferença Absoluta de Luminância; AD_G: Diferença Absoluta de Gradiente; $AD_I x AD_Q$ e $AD_L + AD_G$: combinações que contribuem para alcançar melhores resultados em termos de percepção de erro.
Fonte: Autoria própria.

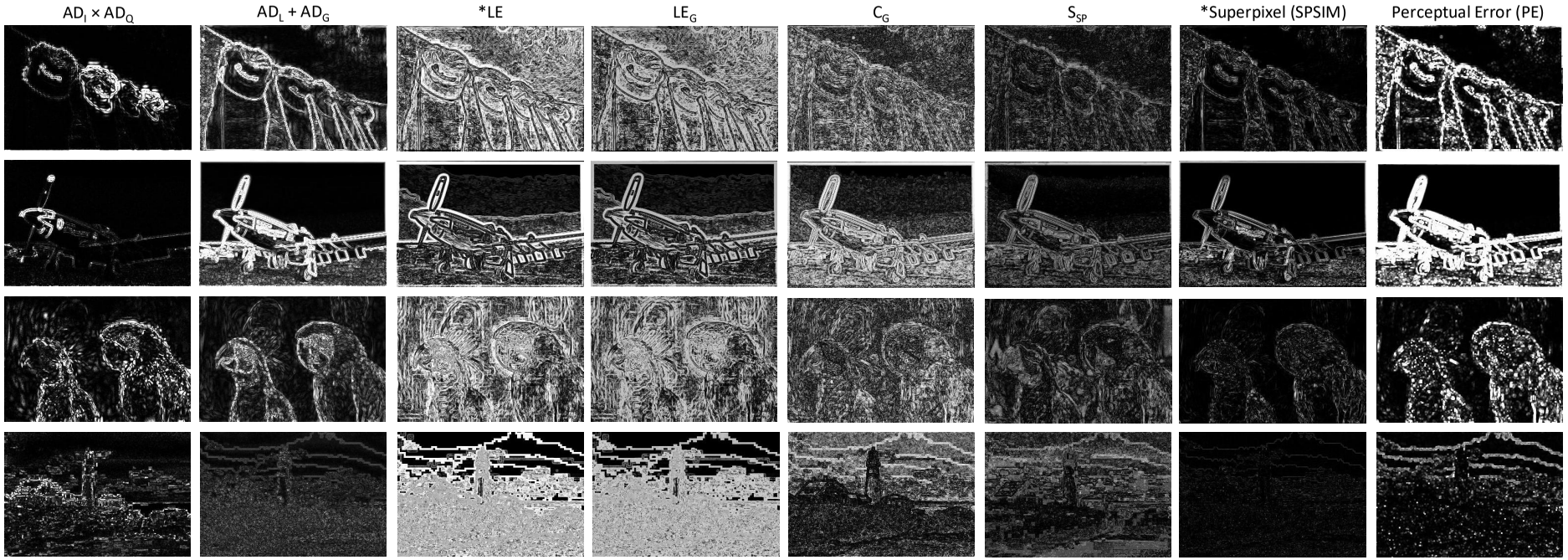


Figura 47 – Exemplos de mapas resultantes das características que contribuem para a formação do mapa final de Erro Perceptual (EP) do método proposto. *LE: característica de energia local existente na literatura; LE_G: característica de energia local proposta; C_G: contraste de gradiente; S_{SP}: superpixel modificado; *$SPSIM$: superpixel da literatura.
Fonte: Autoria própria.

$$W_{LE} = \Big(\frac{1}{N-1}\sum_{i=1}^{N}(LE'_G(i) - \overline{LE'_G})^2\Big)^{1/2}, \tag{5.9}$$

onde LE'_G é LE^2_G, que contribuiu de forma mais eficaz do que a simples aplicação do mapa LE_G, e i representa o i-ésimo píxel. Observe nos mapas LE_G da Fig. 47 que a estrutura das imagens não é alterada em comparação com os mapas LE, mas apenas a intensidade da energia local.

O contraste é uma característica importante do HVS, que tem sido amplamente utilizada em métodos de avaliação de qualidade de imagem [34], [126], [73]. Foi utilizada como uma das três características (luminância, estrutura e contraste) em SSIM [21]. O contraste contribui para diferenciar objetos de seus fundos [87]. Nesta tese, o contraste de Michelson (CM) [92] é utilizado para calcular o contraste. O CM foi reescrito como:

$$C_G(i) = \frac{AD(i)}{GM_r(i) + GM_t(i) + K}, \tag{5.10}$$

onde C_G é o contraste do gradiente, GM_r e GM_t são, respectivamente, mapas de magnitude do gradiente de referência e de teste, i representa o i-ésimo píxel, enquanto K é uma constante positiva usada para manter a estabilidade da fração, e seu valor é 1.

É importante destacar que o recurso de contraste, embora amplamente utilizado na literatura, é aplicado aqui para extrair informações relevantes das imagens, como exemplificado na quinta coluna da Fig. 47, atuando sobre os mapas de magnitude de gradiente derivados dos mapas LSD. O mapa C_G contribui diretamente para a formação do mapa de erro perceptual, exercendo influência exponencial na combinação AD_L+AD_G.

5.2.6 Similaridade (Gradiente e Crominância)

Esta seção apresenta as similaridades de gradiente e crominância entre as imagens de referência e de teste, usadas para formar a similaridade de super-píxeis (*superpixel similarity*). Vale ressaltar que a similaridade de gradiente e crominância foi extensivamente investigada na literatura [24], [74], [78], [120] e [157] para avaliação de distorções de imagem.

A similaridade do gradiente (*gradient similarity*) é definida como a similaridade da magnitude do gradiente em cada píxel entre as imagens de referência e de teste, dada por:

$$S_G(i) = \frac{2GM_r(i)\cdot GM_t(i) + K}{GM_r(i)^2 + GM_t(i)^2 + K}, \tag{5.11}$$

onde GM_r e GM_t são, respectivamente, mapas de magnitude de gradiente de referência e teste, K é uma constante positiva para manter a estabilidade da fração e seu valor proposto é 1.

Sejam $I_r|I_t$ e $Q_r|Q_t$ os canais cromáticos, respectivamente, I e Q das imagens r e t, como em (5.1). Semelhantemente à definição de S_G, a similaridade entre canais cromáticos é definida como:

$$S_C(i) = \frac{2I_r(i)\cdot I_t(i) + K}{I_r(i)^2 + I_t(i)^2 + K} \cdot \frac{2Q_r(i)\cdot Q_t(i) + K}{Q_r(i)^2 + Q_t(i)^2 + K}, \tag{5.12}$$

onde K é uma constante positiva para manter a estabilidade da fração e seu valor proposto é 1.

5.2.7 Similaridade de Super-píxeis

Um grupo de píxeis que compartilham características semelhantes, como intensidade, textura e cor, é chamado de super-píxel (*superpixel*) [158]. Ren e Malik propuseram o conceito de super-píxel em [159]. Devido às suas propriedades inerentes, o conceito foi rapidamente adotado em uma ampla gama de aplicações [160], como detecção de objetos [161], abordagens de segmentação de super-píxeis para imagens de radar de abertura sintética (SAR) polarimétricas [162], [163], lesões cutâneas [164], avaliação de métodos para permitir a geração de conjuntos de dados de treinamento de aprendizado de máquina [165], reconstrução de imagens [158] e super-resolução de imagens [166].

Com base nas vantagens dos super-píxeis, pesquisadores desenvolveram recentemente métodos voltados para a análise de qualidade de imagem (IQA). Fang *et al.* [167] propuseram um modelo de IQA objetivo para imagens de fusão de imagens multiexposição (*Multi-exposure image fusion*, MEF) de cenas estáticas e dinâmicas, baseado em super-píxeis e uma estratégia de agrupamento adaptativo baseada na teoria da informação. Em nosso trabalho, os métodos [75], SSVD [76] e [78] de FR, descritos na Seção 3.1, utilizaram os super-píxeis para IQA. Outros pesquisadores utilizaram o super-píxel para IQA em métodos NR [168], [169].

A similaridade de super-píxeis (*superpixel similarity*, S_{SP}), neste trabalho, foi inspirada em [75]. No entanto, vale destacar que a similaridade de luminância não foi considerada aqui no cálculo da similaridade de super-píxeis, conforme proposto originalmente por [75] e utilizado por [78]. A similaridade de super-píxeis foi reescrita aqui como:

$$S_{SP}(i) = [S_G(i)]^{\alpha+W_{LE}} \cdot exp[\beta(S_C(i) - 1)], \tag{5.13}$$

onde α e β são parâmetros para ajustar os pesos das similaridades de gradiente e crominância, respectivamente, 0,05 e 0,35, conforme [75]. Vale ressaltar que aqui, além do parâmetro α, o S_G também possui, como parâmetro de ajuste, o peso W_{LE} definido em (5.9).

A Fig. 47 apresenta, respectivamente, nas colunas 6 e 7, uma comparação entre os mapas resultantes da similaridade de super-píxeis (S_{SP}) proposta em relação à SPSIM [75]. Por exemplo, nos mapas da coluna 7 da Fig. 47, o método proposto em [75] falha em detectar as bordas suprimidas dos chapéus nos bonés e as existentes nas rodas do avião, bem como as folhas verdes ao redor dos papagaios. Além disso, tal método falha em detectar efetivamente, em áreas homogêneas, mudanças repentinas na intensidade luminosa, como na imagem do avião (plane), e o aparecimento de distorções, como o efeito escada (*staircase effect*), corroborada pela blocagem (*blockness*) e o padrão de mosaico (*mosaic pattern*), na imagem lighthouse2. Pode-se observar que os mapas S_{SP} são mais eficazes na captura dessas distorções, além de outras destacadas nesses mapas, como borramento (*blurring*), vibramento (*ringing*), bordas falsas (*false edges*) e efeito mosquito (*mosquito effect*).

Nesta tese, a similaridade não é a protagonista para contabilizar o cálculo do método proposto, como evidenciado na maioria dos métodos da literatura, mas contribui como um peso, definido da

seguinte forma:

$$W_{SP} = \frac{1}{N}\sum_{i=1}^{N} S_{SP}(i), \tag{5.14}$$

onde N denota o número de i píxeis no S_{SP}. O W_{SP} é combinado com o W_{LE} em (5.16) e (5.17) para ajustar os mapas de características em (5.15), conforme demonstrado na próxima seção.

5.2.8 Cálculo da Avaliação da Qualidade

Esta seção detalha como todos os outros mapas de características apresentados anteriormente foram combinados para formar o mapa de características final e o índice de qualidade visual do método proposto.

O mapa de características do método proposto, denominado Erro Perceptual (*Perceptual Error*, EP), combina as características simples, representadas pelos mapas de diferença absoluta em (5.4), (5.5), (5.6) e (5.7), com as características mais complexas, representadas pelo potencial do mapa de contraste C_G e pelos pesos W_{LE} e W_{SP}, como a seguir:

$$PE = \alpha \cdot \Big(LSD\big(\left[e^{C_G}\right]^{\beta} \cdot (AD_G + AD_L)\big) + AD_I \cdot AD_Q \Big) \tag{5.15}$$

onde α e β são fatores de ajuste definidos nesta tese como:

$$\alpha = \frac{1}{e^{C_1 \cdot W_{SP}[W'_{LE}]^2}} \tag{5.16}$$

$$\beta = C_2 \cdot W'_{LE} - W^2_{SP} \tag{5.17}$$

onde $e^{W_{LE}}$ é W'_{LE} para simplificação em (5.16) e (5.17). C_1 e C_2 são constantes, respectivamente, 2 e 6, que contribuem para otimizar os valores ajustáveis de α e β.

Em (5.15), os mapas de características AD_G e AD_L foram aprimorados pelo mapa de contraste C_G para correlacionar quantitativamente os erros percebidos pelo HVS em relação ao canal de luminância. Nesse contexto, para quantificar o poder do mapa de contraste C_G, o fator de ajuste β foi definido em (5.17), representando uma força de correlação formada a partir da combinação antagônica entre os pesos dos mapas de energia e super-píxel. A função LSD, definida em (5.2), também foi aplicada aos mapas de luminância (AD_L / AD_G) utilizados em (5.15), resultando em melhorias significativas quanto ao método proposto. Nesse sentido, é importante destacar que a função LSD não contribui para melhorias quando aplicada a mapas de crominância. Por outro lado, o fator de ajuste α, definido em (5.16), representando uma força de correlação proporcionalmente combinada entre os pesos dos mapas de energia e super-píxel, contribui efetivamente para os mapas de luminância e crominância.

Fazendo uma simples comparação entre os mapas dos métodos SPSIM e PEL, apresentados respectivamente na penúltima e última colunas da Fig. 47, dois pontos merecem destaques: 1) os mapas PE conseguem capturar os erros existentes nas imagens de teste de forma mais eficaz, mesmo quando tais erros são classificados em diferentes tipos de distorções; 2) os mapas PE exibem naturalmente

super-píxeis com dimensões maiores e, em geral, até mesmo destacando as Regiões de Interesse (*Regions of Interest*, RoIs).

As Tabelas 7 e 9 demonstram a eficácia alcançada pelos mapas PE quando comparados com aqueles produzidos por outros métodos da literatura. Para realizar tais comparações, é necessário mensurar os escores dos métodos quando aplicados a bancos de dados de imagens e, posteriormente, analisá-los por meio de métricas estatísticas, conforme apresentado na Seção 4.3.2.

Nesse sentido, a última etapa do método proposto fornece a pontuação (*score*), geralmente chamada de valor ou índice, que representa a avaliação da qualidade perceptual a respeito da imagem analisada. Diferentemente da maioria dos métodos propostos na literatura, que utilizam a similaridade como medida principal para calcular seu índice de qualidade, nesta tese, foi adotada uma abordagem diferente. O método proposto utiliza a função logarítmica como escala de medida para calcular a qualidade geral da imagem sobre o mapa PE. A razão para isso é fazer com que haja um alinhamento das predições do modelo PEL à percepção humana. Essa prática fundamenta-se na Lei de Weber-Fechner, que estabelece que a resposta do sistema visual humano a estímulos físicos é não-linear e de natureza logarítmica [170]. Além disso, tal procedimento segue as diretrizes da recomendação ITU-T P.1401 para a avaliação estatística de modelos de qualidade objetiva [171].

O índice de PEL é calculado como a seguir:

$$PEL = log_{10}(\frac{1}{N}\sum_{i=0}^{N-1} PE(i) + K) \tag{5.18}$$

onde N denota o número de i píxeis no PE. K é uma constante positiva para manter a estabilidade do logaritmo, e seu valor proposto é 1. Os valores da pontuação de PEL permanecem em uma faixa entre 0 e 5, em virtude da aplicação da base 10 na função logarítmica, bem como convergem para a mesma faixa de valores da escala DMOS [10], [11], sendo o menor valor indicativo de alta qualidade em ambas as faixas. É importante observar que a PEL apresenta valores em uma escala que facilita a análise subjetiva.

6

EXPERIMENTOS, ANÁLISES E DISCUSSÕES

Este capítulo apresenta os experimentos, as análises e as discussões referentes ao método PEL descrito no Cap. 5 e as especificações metodológicas estabelecidas no Cap. 4. O código-fonte em MATLAB e os resultados do método PEL, bem como de seus dezenove concorrentes, podem ser obtidos em https://github.com/sergioacbezerra/PEL.

6.1 Visão Geral

Neste capítulo, são descritas as análises e discussões baseadas nas orientações definidas no Capítulo 4 sobre os experimentos, a fim de comprovar as habilidades perceptuais do método proposto PEL em comparação com 19 (dezenove) métodos FR-IQA do estado da arte, a saber: SSIM [21], MS-SSIM [22], FSIM$_C$ [24], GMSD [71], VSI [120], MCSD [16], SCQI [126], PSIM [125], RVSIM [73], SSRM [127], HaarPSI [128], EFS [74], SPSIM [75], CVSS [72], SSVD [76], SUMMER [133], VCGS [77], DSSIM [129] e SSVS [78].

Os experimentos nesta tese envolveram oito bases de dados de imagens (BDIs) amplamente utilizadas na literatura, a saber: IVC [123], LIVE [12], CSIQ [134], MICT [124], VCL [135], TID13 [136], CIDIQ [137] e KADID [138].

6.2 Custo Computacional

Na Seção 4.3.1 são abordadas as razões pelas quais o Custo Computacional foi definido como medida de eficiência.

Nesta tese, foi adotado o tempo médio de execução como medida de eficiência. O tempo médio de execução de todos os vinte métodos foi medido em milissegundos sobre todas as imagens das BDIs, a fim de comparar o custo computacional dos métodos adotados nos experimentos.

Os experimentos foram repetidos dez vezes para cada distorção das BDIs em um laptop com CPU Intel Core i5-10300H de 2,50 GHz e 20 GB de RAM. Todos os códigos foram implementados com

MATLAB R2015a no Windows 11 de 64 bits [172].

A Tabela 6 apresenta a classificação do custo computacional para os resultados dos experimentos quanto aos métodos avaliados, com um índice subscrito de 1 (melhor) a 20 (pior). Os resultados são tabulados em termos de velocidades de execução em milissegundos (ms) por imagem e quadros por segundo (*Frames Per Second*, fps). O método proposto alcança a quarta posição, ligeiramente superado apenas pelos métodos GMSD, MCSD e CVSS, respectivamente, 2,10, 1,67 e 1,24 vezes. No entanto, vale destacar que esses métodos não utilizam recursos de crominância, bem como demonstram desvantagens em comparação com o método PEL quanto à habilidade perceptual em nível de desempenho e custo-benefício, como demonstrado nas próximas seções.

Tabela 6 – Comparações do custo médio de computação em milissegundos/imagem e velocidade de execução (*Running speed*) em quadros por segundo (*Frames Per Second*, fps) para vinte métodos FR-IQA em oito bancos de imagens.

Método FR-IQA	**Velocidade de Execução (*ms*)↓**	**Quadros por Sec. (*fps*)↑**
$SSIM_5$	27.02	37.0
$MS\text{-}SSIM_6$	35.58	28.1
$FSIM_{C15}$	171.73	5.8
$GMSD_1$	11.30	88.5
VSI_{13}	103.14	9.7
$MCSD_2$	14.22	70.3
$SCQI_{11}$	76.18	13.1
$PSIM_7$	41.38	24.2
$RVSIM_{19}$	854.81	1.2
$SSRM_{12}$	99.37	10.1
$HaarPSI_9$	55.97	17.9
EFS_{18}	409.29	2.4
$SPSIM_{16}$	175.31	5.7
$CVSS_3$	19.07	52.4
$SSVD_{14}$	151.50	6.6
$SUMMER_8$	47.46	21.1
$VCGS_{17}$	318.63	3.1
$DSSIM_{10}$	64.85	15.4
$SSVS_{20}$	906.09	1.1
PEL_4	23.73	42.1

As taxas de quadros de vídeo em tempo real foram definidas como 15 a 30 fps por diversos estudos de fatores humanos [173–176]. A taxa de 30 fps é considerada o padrão de fato para vídeo em tempo real, mas taxas de quadros mais baixas são aceitáveis com base na necessidade de evitar oscilações e na representação precisa do movimento [177]. Diversos estudos recentes voltados ao processamento digital de imagens e vídeos têm utilizado a taxa de 30 fps como valor mínimo para que uma aplicação ou modelo seja considerado sua execução em tempo real, como em [178–183].

Nesse contexto, o método PEL consegue processar em média, para todos as oito BDIs, 42,1 fps, sendo superado apenas pelos métodos GMSD, MCSD e CVSS. PEL consegue ser mais eficiente que o método SSIM, que alcança em média 37 fps. Já os demais métodos alcançam uma velocidade de execução em quadro por segundo menor que 30 fps.

6.3 Análise de Desempenho

Com base nos procedimentos recomendados pelo VQEG [10], [11], três etapas, como descritas na Seção 4.3.2, foram aplicadas para analisar o desempenho do método proposto PEL em relação aos seus concorrentes. As análises e discussões dos resultados desse desempenho são descritas nas Seções de 6.3.1 a 6.3.3.

6.3.1 Comparação entre Valores Subjetivos e Objetivos Normalizados

Os fundamentos teóricos referentes aos Valores DMOS/MOS e Mapeamento para a Escala Subjetiva foram abordados na Seção 4.3.2.1. Nesta seção, os resultados da correlação dos métodos quanto a esse mapeamento são apresentados com base em uma visão específica ou geral, respectivamente, nas Seções 6.3.1.1 e 6.3.1.2.

6.3.1.1 Visão Específica

Em geral, o método proposto apresenta destaque diferenciado em relação aos seus concorrentes nas BDIs IVC, CSIQ, MICT, VCL e CIDIQ, como pode ser visto nas Figs. 48, 50, 51, 52 e 54.

Já na BDI LIVE, Fig 49, os métodos HaarPSI, PEL e CVSS, e na BDI KADID, Fig. 55, os métodos HaarPSI, PEL e SSVD apresentam poucas diferenças quanto às suas distribuições.

Em relação à BDI TID2013, Fig. 53, os métodos SSVS, SCQI, PSIM, SPSIM e PEL alternam entre si quanto às primeiras posições, como pode ser atestado na Fig. 53, com leve vantagem para o método SSVS.

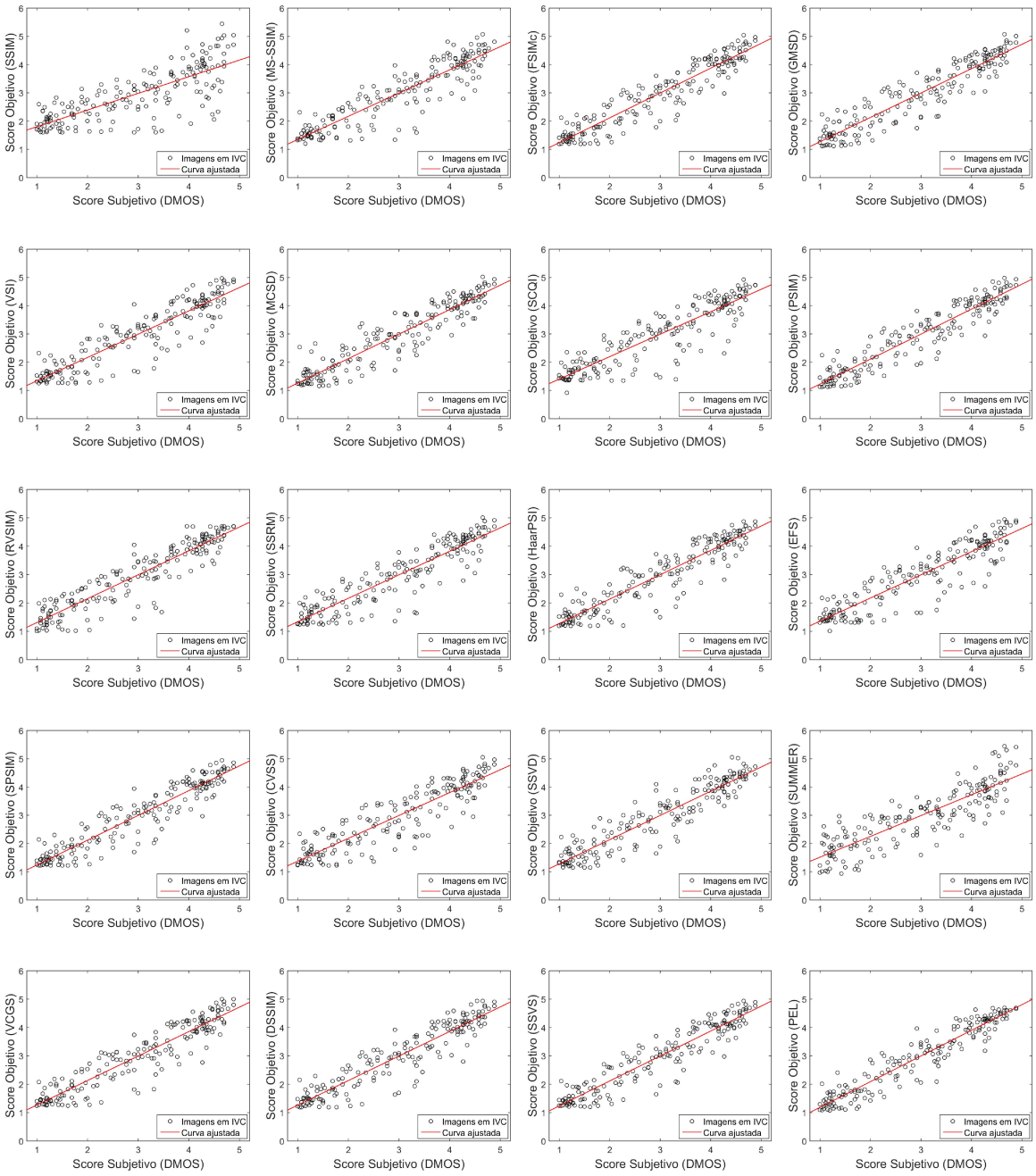


Figura 48 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens IVC.
Fonte: Autoria própria.

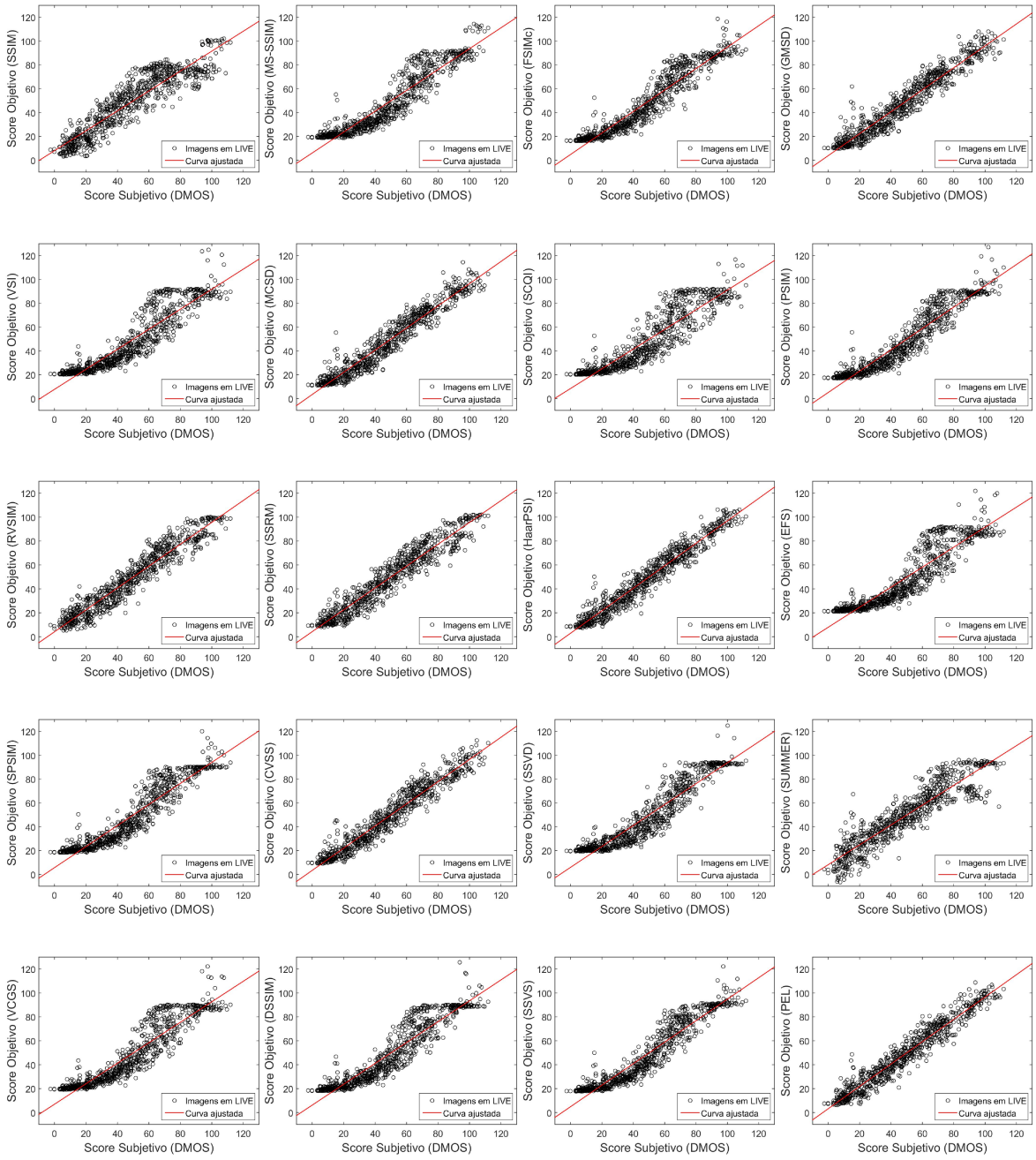


Figura 49 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens LIVE.
Fonte: Autoria própria.

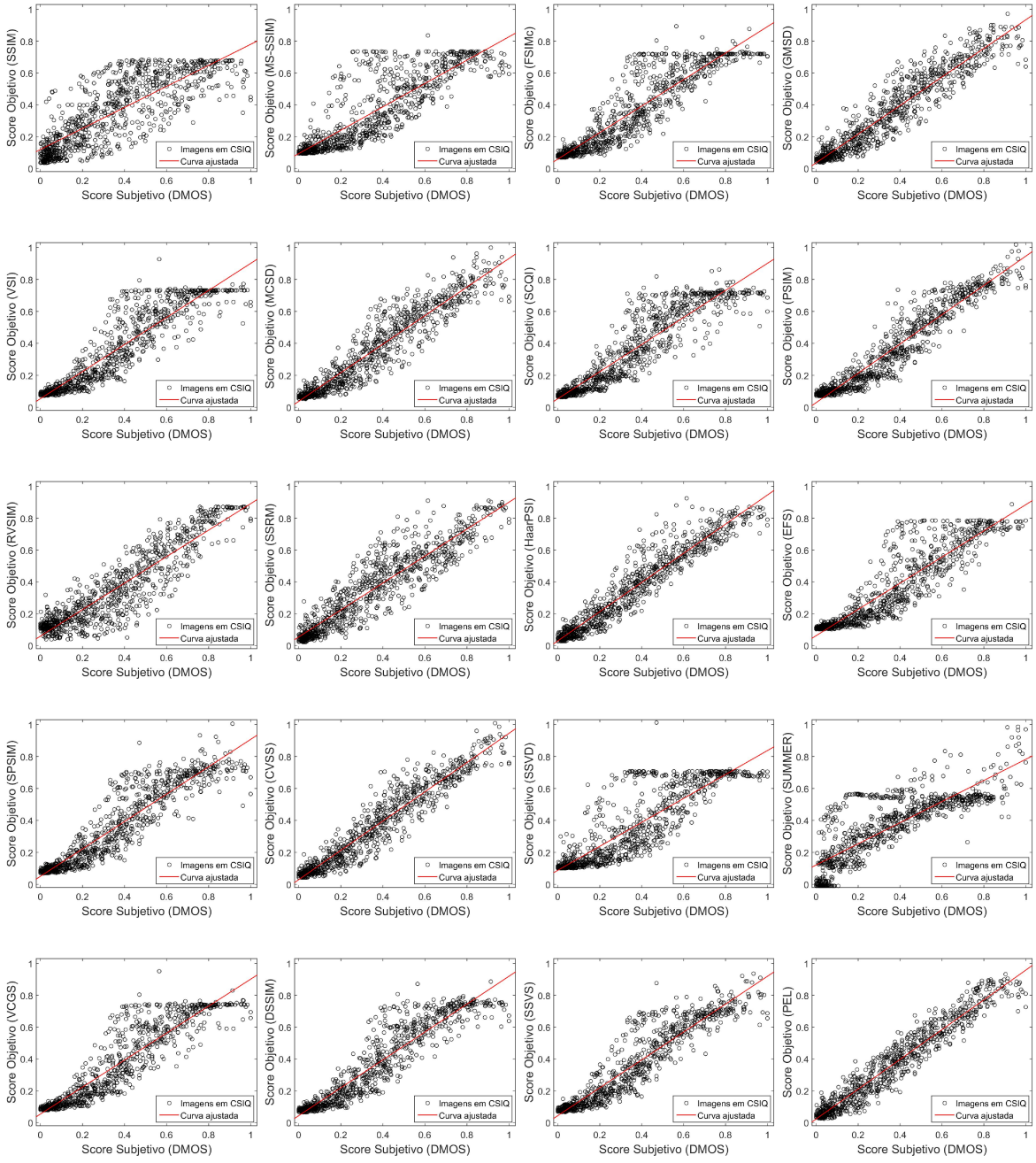


Figura 50 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens CSIQ.
Fonte: Autoria própria.

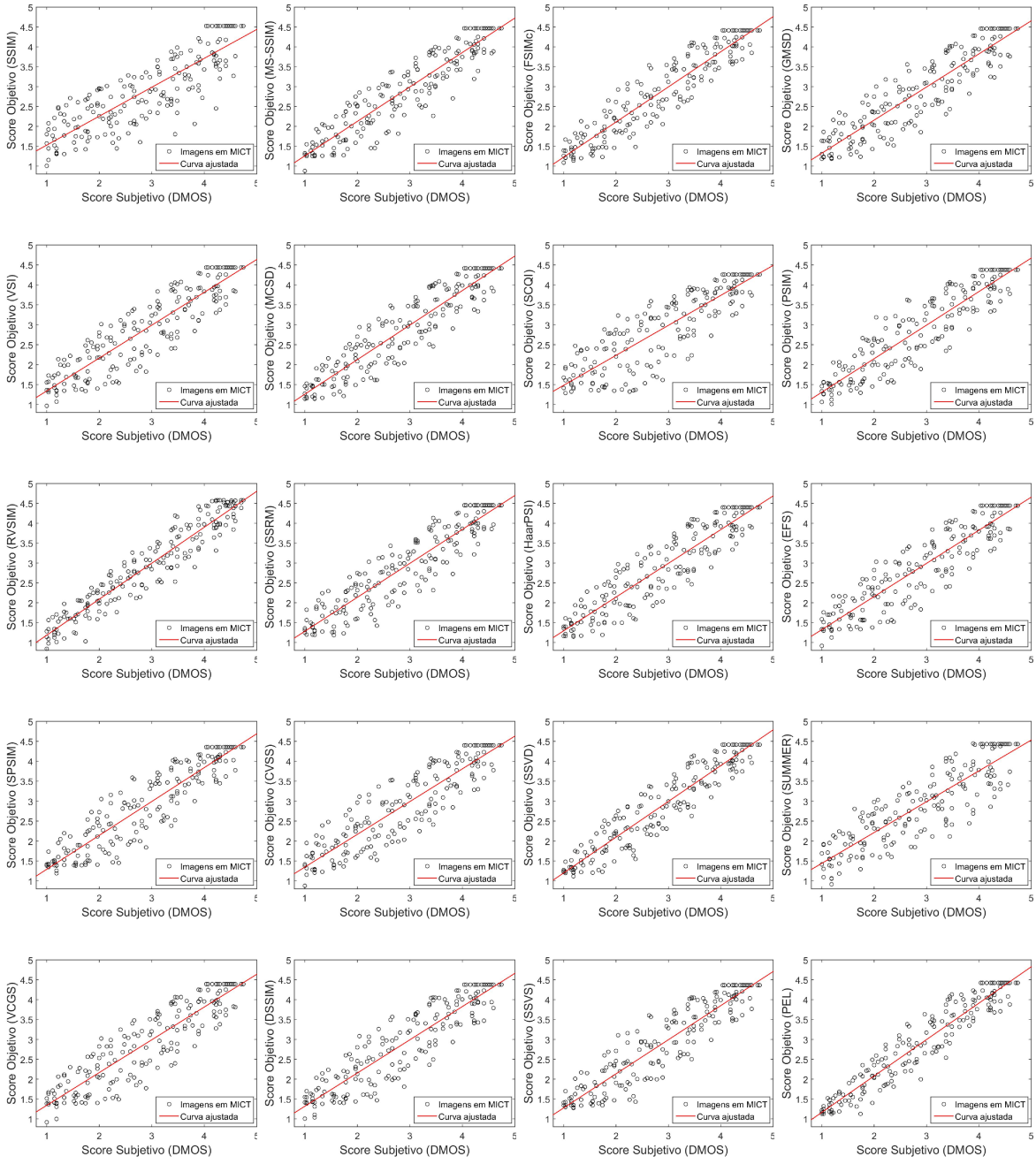


Figura 51 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens MICT.
Fonte: Autoria própria.

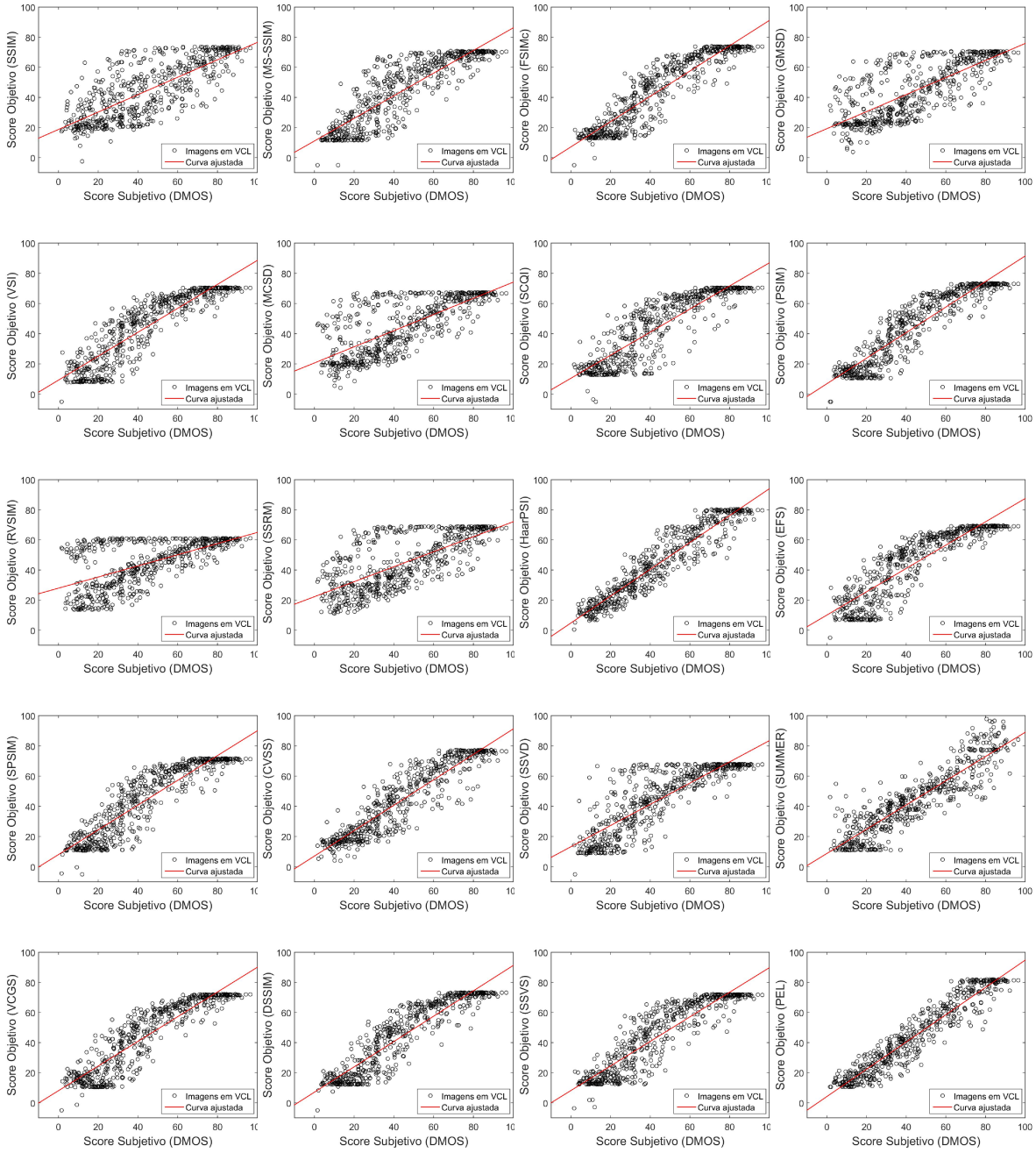


Figura 52 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens VCL.
Fonte: Autoria própria.

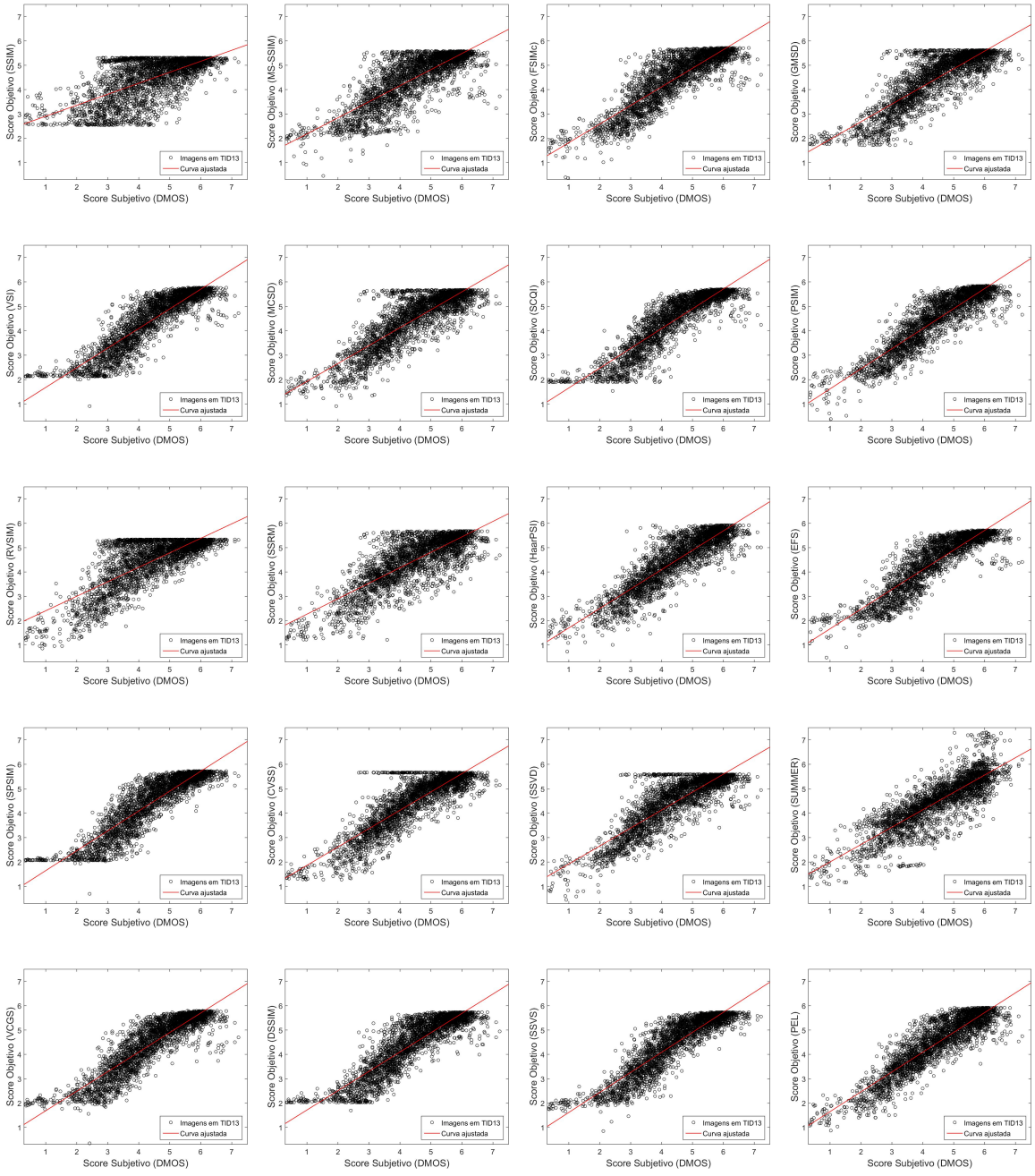


Figura 53 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens TID13.
Fonte: Autoria própria.

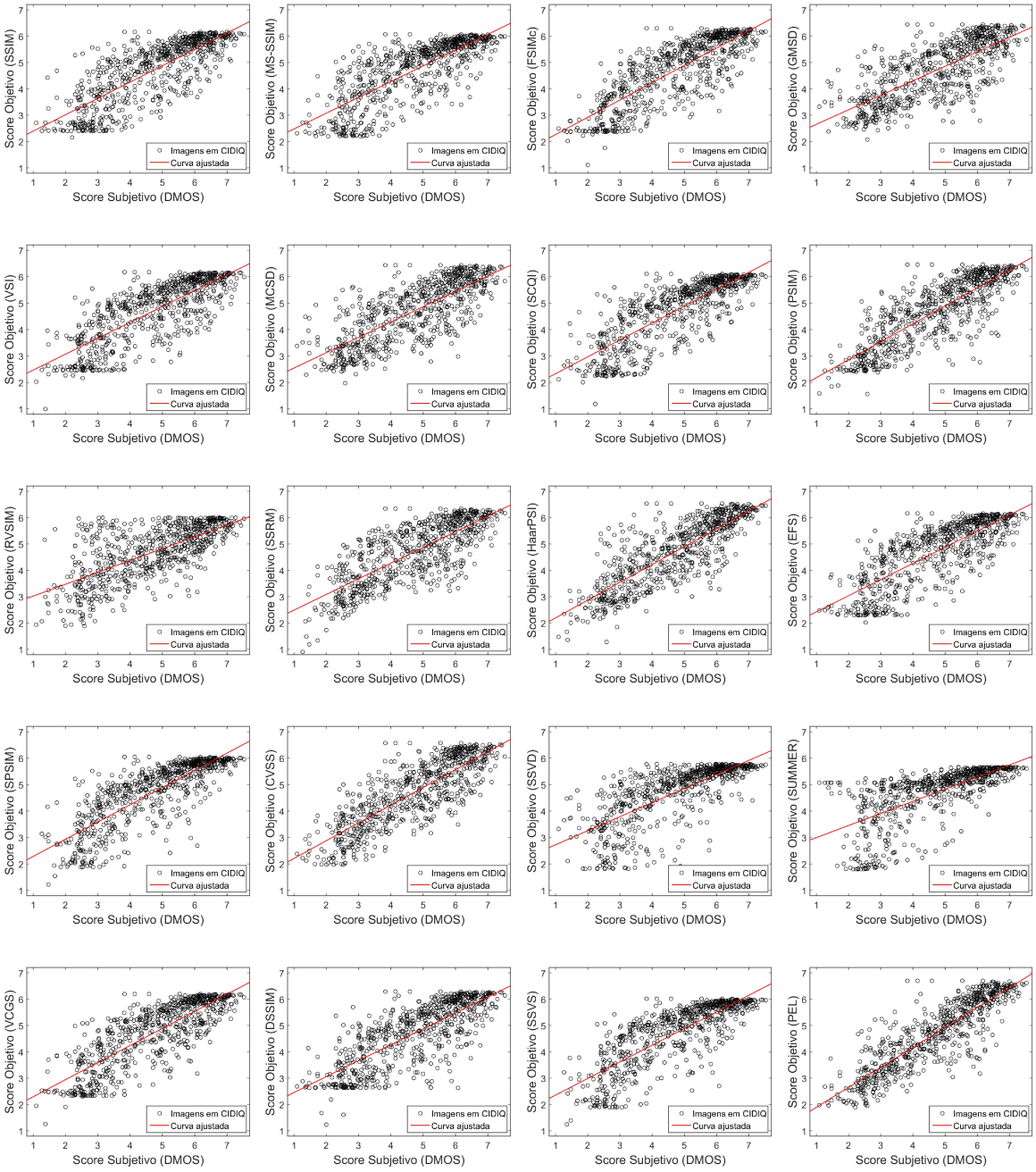


Figura 54 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens CIDIQ.
Fonte: Autoria própria.

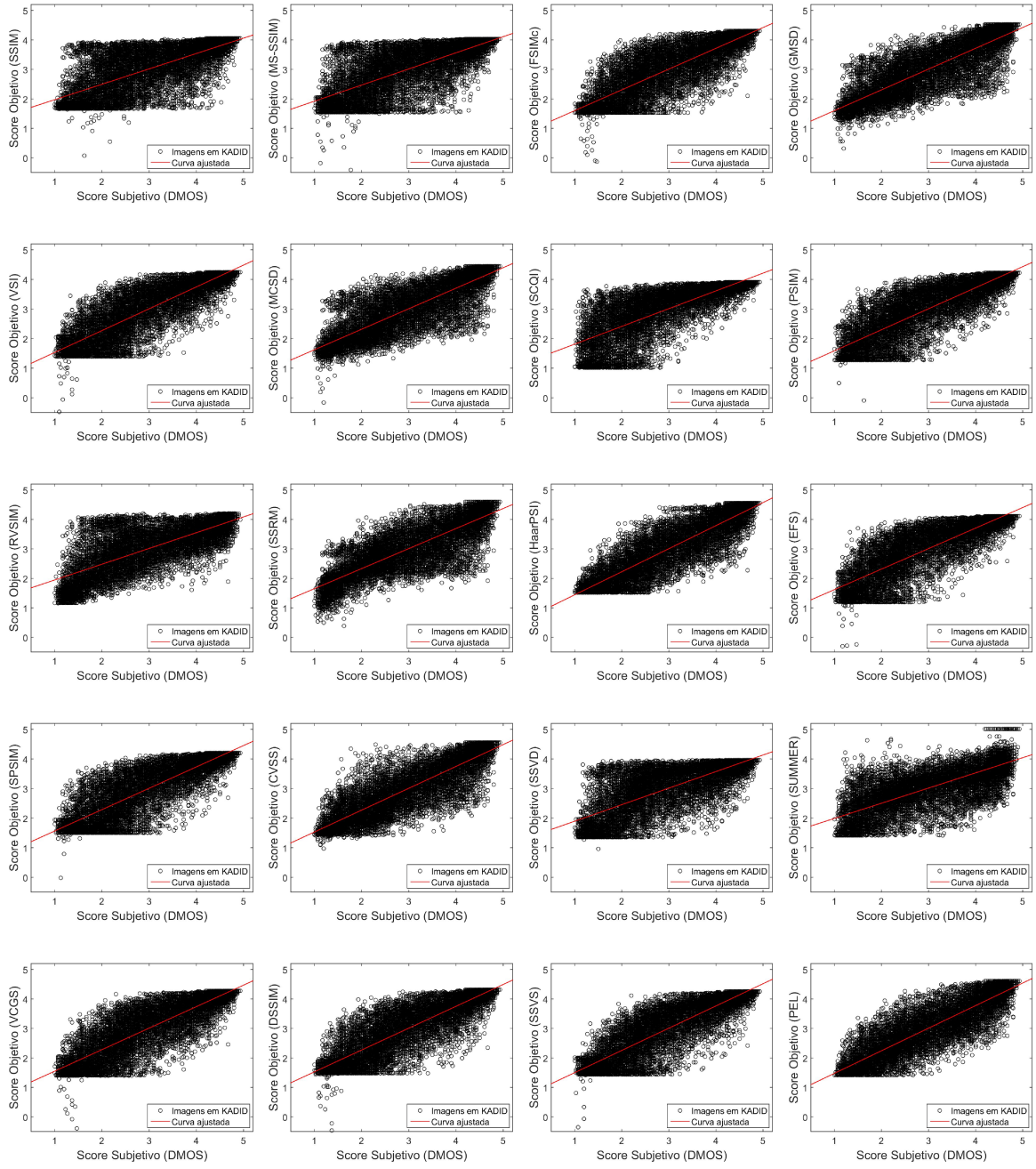
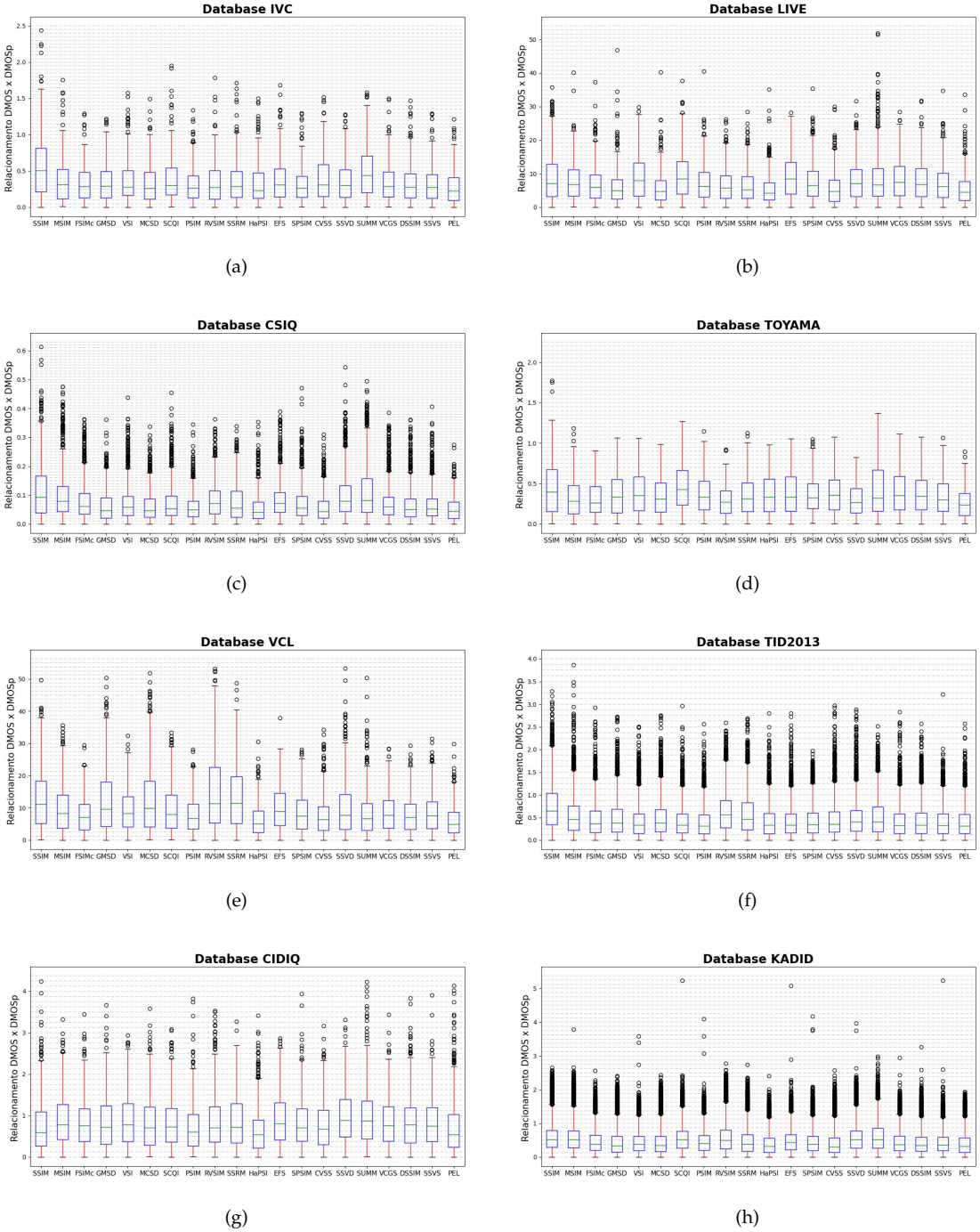


Figura 55 – Gráficos de dispersão de DMOS subjetivo versus pontuações obtidas por métodos para FR-IQA para todas as distorções no banco de imagens KADID.
Fonte: Autoria própria.

6.3.1.2 Visão Geral

A Fig. 56 apresenta gráficos de caixa (box plots) das relações $DMOS$ x $DMOS_p$, onde são demonstradas comparações de desempenho dos 20 métodos em estudo nas oito BDIs nesta tese. É importante destacar que um desvio padrão menor com um valor mediano menor indica melhor desempenho.



(a) (b) (c) (d) (e) (f) (g) (h)

Figura 56 – Box plots (Gráficos de caixa) de DMOS subjetivo versus pontuações obtidas por métodos FR-IQA para todas as distorções nas oito bases de dados de imagens (BDIs).
Fonte: Autoria própria.

De acordo com os gráficos de caixa mostrados na Fig. 56, em geral, os resultados experimentais demonstram que as pontuações objetivas previstas pelo método proposto são muito mais consistentes com as avaliações subjetivas do que todos os dezenove métodos de IQA da literatura avaliados. Além disso, o PEL demonstra consistência e robustez com HVS para analisar os diferentes tipos de distorções existentes nas BDIs utilizadas. Para demonstrar essa habilidade perceptual superior do método PEL em relação aos seus concorrentes, algumas análises e discussões dos resultados experimentais sobre as oito BDIs são descritas a seguir.

Para a BDI IVC, o método proposto alcança um desempenho estatisticamente superior a todos os métodos. Conforme a Tabela 7, os métodos PEL, PSIM, FSIM/SSVS e SPSIM são os mais bem classificados, respectivamente. Nesse contexto, é interessante observar, como mostrado na primeira linha da Tabela 4, que PEL apresenta significância superior aos métodos PSIM, SSVS, FSIM e SPSIM, com F% variando entre 20 e 30% na BDI IVC, com valores de 22,16, 26,47, 27,59 e 29,74, respectivamente. Para uma visão mais detalhada dessa superioridade, pode-se observar no gráfico de caixa da Fig. 56(a) que PEL se mostra mais eficaz do que esses métodos, alcançando os menores valores para a mediana, Q1 e Q2. Por outro lado, os valores de IQR e *outliers* do método PEL são muito semelhantes aos dos métodos PSIM, SSVS, FSIM e SPSIM, demonstrando, portanto, comportamento semelhante em relação à distribuição de seus valores de pontuação. Vale ressaltar também, como mostrado na Tabela 6, que, dentre esses métodos, apenas os métodos PEL e PSIM apresentam eficiência em tempo real, com PEL sendo aproximadamente 1,74 vezes mais rápido que o método PSIM.

Para a BDI de LIVE, há uma competição muito próxima, como mostrado na Tabela 7, entre os métodos HaarPSI, CVSS, PEL e MCSD, respectivamente. Esta classificação também pode ser observada na Fig. 56(b) por meio da mediana. No entanto, pode ser observado que o método HaarPSI tem apenas uma pequena vantagem sobre PEL em termos de valores de IQR, Q2, whisker e outlier, enquanto para o valor Q1, essa vantagem é em favor do método PEL. Neste cenário, nota-se que o melhor valor de Q1 pertence ao método CVSS, que por sua vez possui desvantagens em relação aos outros valores pertencentes aos métodos HaarPSI e PEL. O método CVSS também possui muita semelhança em relação ao método MCSD. Consoante à segunda linha da Tab. 9, HaarPSI tem uma pequena vantagem significativa sobre o método PEL, que por sua vez não mostra diferença significativa comparado aos métodos CVSS e MCSD. Vale a pena observar, no entanto, como mostrado na Tabela 6, que MCSD, CVSS e são mais eficientes que o método HaarPSI.

Para a BDI CSIQ, o método proposto obtém um desempenho estatisticamente superior a todos os métodos. Conforme a Tabela 7, os métodos PEL, PSIM, HaarPSI e CVSS são os mais bem classificados, respectivamente. É interessante observar, como mostrado na terceira linha da Tabela 9, que PEL apresenta significância superior em comparação aos métodos PSIM, HaarPSI e CVSS, com valores de 20,06, 29,06 e 30,42, respectivamente. Apesar do aparente equilíbrio entre os métodos HaarPSI e PEL em termos dos valores de mediana, Q1 e Q2 apresentados na Fig. 56(c), a superioridade do método PEL, como evidenciado na Tabela 7, pode ser notada pela vantagem em termos do valor do IQR e da centralização

da mediana na caixa, bem como pelos poucos valores *outliers* em comparação aos seus concorrentes. No geral, o método PEL é mais eficaz que seus concorrentes, alcançando alto desempenho na BDI CSIQ, conforme mostrado na Tabela 6. O PEL só é superado em eficiência pelo método CVSS, que é 1,24 vezes mais rápido.

Para a BDI MICT, o método proposto obtém um desempenho estatisticamente melhor do que todos os métodos. Com base na Tabela 7, os métodos PEL, RVSIM e SSVD são os mais bem classificados, respectivamente. Conforme a quarta linha da Tabela 9, PEL tem significância superior em comparação aos métodos RVSIM e SSVD, com valores de 12,54 e 23,34, respectivamente. A superioridade do PEL é mais eficaz do que esses métodos, como visto no gráfico de caixa na Fig. 56(d), pois alcança os melhores valores para mediana, Q1, Q2 e IQR, apresentando valores semelhantes apenas em termos de *whiskers* e *outliers*. Também vale a pena notar, como mostrado na Tabela 6, que o método PEL possui eficiência em tempo real, enquanto RVSIM e SSVD requerem um tempo computacional exorbitante.

Para a BDI VCL, o método proposto também alcança um desempenho estatisticamente superior a todos os métodos. Conforme a Tabela 7, os métodos PEL e HaarPSI são os mais bem classificados, respectivamente. Conforme mostrado na quinta linha da Tabela 9, PEL apresenta uma superioridade de 15,56 vezes em significância em comparação ao método HaarPSI. A superioridade do PEL é ligeiramente mais eficaz do que o método HaarPSI, como visto no gráfico de caixa da Fig. 56(e), ao alcançar melhores valores para mediana, Q1, Q2, IQR e *whiskers*, apesar de apresentar valores semelhantes para *outliers*. Também vale a pena notar, consultando a Tabela 6 referente à BDI VCL, que a eficiência do método PEL é aproximadamente 2,36 vezes maior que a do método HaarPSI.

Para a BDI TID2013, como mostrado na Tabela 7, há uma competição de classificação muito acirrada e difícil entre os métodos SSVS, PSIM, PEL, SPSIM, SCQI, VSI, EFS, VCGS, HaarPSI e DSSIM. Na Tabela 7, os resultados mostram uma ligeira vantagem numérica do SSVS sobre os outros métodos, o que também é demonstrado na Tabela 9. Nesse contexto, conforme a sexta linha desta tabela, o SSVS tem uma ligeira vantagem significativa sobre o PEL (ou seja, 5,40). Comparado aos outros métodos, o método PEL não apresenta diferença significativa em relação aos métodos SCQI, PSIM, EFS e SPSIM. Por outro lado, o PEL tem uma ligeira vantagem significativa sobre os métodos VSI, HaarPSI, VCGS e DSSIM. Essa grande dificuldade em classificar os métodos referentes à BDI TID2013 pode ser observada nos valores apresentados na Fig. 56(f). Dada essa competição extremamente acirrada em termos de eficácia, a eficiência de um método deve ser considerada, sendo que, neste caso, o método PEL apresenta o melhor resultado, seguido pelo método PSIM.

Para a BDI CIDIQ, o método proposto alcança um desempenho estatisticamente superior a todos os seus concorrentes. Conforme a Tabela 7, os métodos PEL, HaarPSI, PSIM, SPSIM, CVSS e SSIM são os mais bem classificados, respectivamente. Conforme mostrado na sétima linha da Tabela 9, PEL apresenta significância superior em comparação a todos os métodos. É importante destacar que PEL e o HaarPSI são os métodos mais eficazes. A Fig. 56(g) mostra valores favoráveis para o método HaarPSI em termos de Q1, Q2, IQR e *whiskers*, enquanto em termos de mediana e *outliers*, ambos são semelhantes. Deve-se

notar, no entanto, que PEL apresenta maior eficiência do que o método HaarPSI.

Para a BDI KADID, há uma competição muito acirrada, como mostrado na Tabela 7, entre os métodos HaarPSI, PEL e SSVS, respectivamente. Consoante a oitava linha da Tabela 9, o HaarPSI apresenta uma ligeira superioridade em significância em comparação com o método PEL, que, por sua vez, não apresenta diferença significativa em comparação com o método SSVS. A Fig. 56(h) apresenta valores muito semelhantes entre esses métodos, mas com uma ligeira vantagem para os métodos HaarPSI e PEL, respectivamente, em termos de IQR e *outliers*. No geral, esses três métodos são muito eficazes devido aos seus desempenhos, mas o PEL é claramente superior em termos de eficiência.

6.3.2 Predição da Avaliação Estatística dos Métodos

Na Seção 4.3.2.2 são descritas as métricas estatísticas adotadas nesta tese para medir as habilidades perceptuais dos métodos quanto ao seu desempenho da predição em IQA. Nesse sentido, os experimentos, análises e discussões foram separados em duas partes, como descritas a seguir.

6.3.2.1 Comparação de Desempenho sobre as BDIs

Nesta seção, é apresentado o desempenho de predição medido por métricas de avaliação (SROCC, KROCC, PLCC e RMSE) de cada método de IQA concorrente nas imagens nas BDIs IVC, LIVE, CSIQ, MICT, VCL, TID13, CIDIQ e KADID.

Os resultados estão listados na Tabela 7. Os três melhores resultados estão destacados em negrito para cada métrica estatística de avaliação. Além disso, na Tabela 7, os índices subscrito de 1 a 20 são usados para facilitar a classificação em ordem decrescente dos métodos de IQA de acordo com seu desempenho. A antepenúltima seção desta tabela apresenta as médias dos resultados de SROOC, KROOC e PLCC em oito BDIs para avaliar o desempenho geral.

Similarmente aos trabalhos [73], [75], [76], [77], [126], o RMSE geral não foi calculado porque cada BDI possui escalas diferentes para as pontuações de avaliação da qualidade subjetiva. As últimas seções apresentam a quantidade de Top-1 e Top-3 alcançada pelos métodos de IQA para todas as métricas de avaliação.

O método PEL apresenta desempenho consistente em todos os bancos de dados. Em termos de todas as métricas de avaliação (SROOC, KROOC, PLCC e RMSE), ele foi classificado 19 vezes como Top-1, seguido por HaarPSI (8 vezes), SSVS (3 vezes) e PSIM (uma vez). O método PEL foi classificado 30 vezes como Top-3, seguido por HaarPSI (20 vezes), PSIM (16 vezes) e SSVS (10 vezes).

Tabela 7 – Comparação de métricas de avaliação de desempenho para métodos IQA em oito BDI.

BDI	Métrica	SSIM	MSSIM	FSIM$_C$	GMSD	VSI	MCSD	SCQI	PSIM	RVSIM	SSRM	HaPSI	EFS	SPSIM	CVSS	SSVD	SUMER	VCGS	DSSIM	SSVS	PEL
	SROCC↑	0.7677_{20}	0.8967_{15}	$\mathbf{0.9286}_{3}$	0.9212_{7}	0.8983_{14}	0.9243_{6}	0.8836_{18}	$\mathbf{0.9333}_{2}$	0.9124_{12}	0.9021_{13}	0.9163_{11}	0.8912_{17}	0.9268_{5}	0.8942_{16}	0.9196_{9}	0.8725_{19}	0.9171_{10}	0.9207_{8}	0.9278_{4}	$\mathbf{0.9443}_{1}$
IVC	KROCC↑	0.5844_{20}	0.7197_{14}	$\mathbf{0.7619}_{3}$	0.7475_{8}	0.7190_{15}	0.7541_{6}	0.7039_{18}	$\mathbf{0.7714}_{2}$	0.7454_{10}	0.7290_{13}	0.7460_{9}	0.7074_{17}	0.7600_{5}	0.7076_{16}	0.7427_{12}	0.6782_{19}	0.7442_{11}	0.7511_{7}	0.7600_{4}	$\mathbf{0.7910}_{1}$
2005	PLCC↑	0.7680_{20}	0.9068_{15}	0.9376_{4}	0.9287_{9}	0.9088_{14}	0.9312_{6}	0.8918_{18}	$\mathbf{0.9404}_{2}$	0.9178_{12}	0.9102_{13}	0.9267_{10}	0.9041_{16}	0.9366_{5}	0.8998_{17}	0.9260_{11}	0.8584_{19}	0.9288_{8}	0.9306_{7}	$\mathbf{0.9380}_{3}$	$\mathbf{0.9515}_{1}$
	RMSE↓	0.7889_{20}	0.5192_{15}	0.4281_{4}	0.4567_{9}	0.5138_{14}	0.4489_{6}	0.5574_{18}	$\mathbf{0.4189}_{2}$	0.4890_{12}	0.5102_{13}	0.4628_{10}	0.5265_{16}	0.4317_{5}	0.5374_{17}	0.4651_{11}	0.6320_{19}	0.4564_{8}	0.4509_{7}	$\mathbf{0.4269}_{3}$	$\mathbf{0.3790}_{1}$
	SROCC↑	0.9164_{20}	0.9550_{13}	0.9645_{6}	0.9603_{9}	0.9488_{17}	0.9655_{4}	0.9374_{18}	0.9615_{7}	0.9588_{10}	0.9584_{11}	$\mathbf{0.9683}_{1}$	0.9507_{16}	0.9606_{8}	$\mathbf{0.9672}_{2}$	0.9526_{15}	0.9304_{19}	0.9532_{14}	0.9559_{12}	0.9647_{5}	$\mathbf{0.9657}_{3}$
LIVE	KROCC↑	0.7395_{20}	0.8124_{13}	0.8363_{6}	0.8262_{9}	0.7966_{17}	0.8365_{5}	0.7755_{18}	0.8268_{7}	0.8167_{10}	0.8162_{11}	$\mathbf{0.8448}_{1}$	0.8006_{16}	0.8262_{8}	$\mathbf{0.8406}_{2}$	0.8038_{15}	0.7649_{19}	0.8073_{14}	0.8132_{12}	0.8370_{4}	$\mathbf{0.8374}_{3}$
2005	PLCC↑	0.9144_{18}	0.9360_{13}	0.9496_{8}	0.9600_{5}	0.9172_{16}	0.9649_{4}	0.9091_{20}	0.9464_{10}	0.9567_{6}	0.9560_{7}	$\mathbf{0.9696}_{1}$	0.9150_{17}	0.9404_{11}	$\mathbf{0.9653}_{3}$	0.9376_{12}	0.9126_{19}	0.9257_{15}	0.9343_{14}	0.9482_{9}	$\mathbf{0.9665}_{2}$
	RMSE↓	11.087_{18}	9.6407_{13}	8.5893_{8}	7.6676_{5}	10.911_{16}	7.1889_{4}	11.409_{20}	8.8499_{10}	7.9690_{6}	8.0378_{7}	$\mathbf{6.7023}_{1}$	11.052_{17}	9.3184_{11}	$\mathbf{7.1487}_{3}$	9.5279_{12}	11.197_{19}	10.360_{15}	9.7661_{14}	8.6984_{9}	$\mathbf{7.0234}_{2}$
	SROCC↑	0.8332_{19}	0.8840_{17}	0.9307_{14}	0.9516_{7}	0.9419_{12}	0.9519_{5}	0.9434_{11}	$\mathbf{0.9621}_{2}$	0.8831_{18}	0.9255_{15}	$\mathbf{0.9603}_{3}$	0.9346_{13}	0.9434_{10}	0.9580_{4}	0.8849_{16}	0.8266_{20}	0.9442_{9}	0.9516_{6}	0.9492_{8}	$\mathbf{0.9627}_{1}$
CSIQ	KROCC↑	0.6305_{20}	0.7055_{16}	0.7683_{14}	0.8026_{6}	0.7842_{12}	0.8033_{5}	0.7863_{10}	$\mathbf{0.8265}_{1}$	0.7021_{18}	0.7589_{15}	$\mathbf{0.8234}_{3}$	0.7739_{13}	0.7859_{11}	0.8173_{4}	0.7026_{17}	0.6528_{19}	0.7899_{9}	0.8013_{8}	0.8013_{7}	$\mathbf{0.8264}_{2}$
2010	PLCC↑	0.8131_{20}	0.8569_{18}	0.9126_{14}	0.9479_{5}	0.9202_{13}	0.9454_{6}	0.9206_{11}	$\mathbf{0.9574}_{3}$	0.9126_{15}	0.9204_{12}	$\mathbf{0.9574}_{2}$	0.9072_{16}	0.9259_{9}	0.9569_{4}	0.8657_{17}	0.8200_{19}	0.9207_{10}	0.9360_{7}	0.9358_{8}	$\mathbf{0.9672}_{1}$
	RMSE↓	0.1532_{20}	0.1357_{18}	0.1076_{14}	0.0839_{5}	0.1030_{13}	0.0857_{6}	0.1028_{11}	$\mathbf{0.0760}_{3}$	0.1076_{15}	0.1029_{12}	$\mathbf{0.0760}_{2}$	0.1107_{16}	0.0994_{9}	0.0764_{4}	0.1317_{17}	0.1506_{19}	0.1027_{10}	0.0926_{7}	0.0928_{8}	$\mathbf{0.0669}_{1}$
	SROCC↑	0.8542_{20}	0.9303_{5}	0.9401_{4}	0.9117_{13}	0.9066_{16}	0.9278_{6}	0.8741_{19}	0.9161_{11}	$\mathbf{0.9500}_{2}$	0.9237_{8}	0.9169_{10}	0.9125_{12}	0.9226_{9}	0.9051_{17}	$\mathbf{0.9447}_{3}$	0.8798_{18}	0.9073_{15}	0.9112_{14}	0.9272_{7}	$\mathbf{0.9536}_{1}$
MICT	KROCC↑	0.6656_{20}	0.7722_{5}	0.7854_{4}	0.7377_{12}	0.7272_{17}	0.7663_{6}	0.6860_{19}	0.7452_{10}	$\mathbf{0.8020}_{2}$	0.7549_{9}	0.7393_{11}	0.7373_{13}	0.7572_{8}	0.7296_{16}	$\mathbf{0.7948}_{3}$	0.6952_{18}	0.7300_{15}	0.7351_{14}	0.7645_{7}	$\mathbf{0.8154}_{1}$
2011	PLCC↑	0.8534_{20}	0.9313_{5}	0.9403_{4}	0.9128_{14}	0.9078_{16}	0.9307_{6}	0.8668_{19}	0.9169_{11}	$\mathbf{0.9517}_{2}$	0.9246_{8}	0.9212_{10}	0.9131_{13}	0.9216_{9}	0.9062_{17}	$\mathbf{0.9469}_{3}$	0.8789_{18}	0.9083_{15}	0.9147_{12}	0.9263_{7}	$\mathbf{0.9572}_{1}$
	RMSE↓	0.5976_{20}	0.4175_{5}	0.3902_{4}	0.4683_{14}	0.4808_{16}	0.4193_{6}	0.5717_{19}	0.4574_{11}	$\mathbf{0.3521}_{2}$	0.4367_{8}	0.4460_{10}	0.4673_{13}	0.4448_{9}	0.4849_{17}	$\mathbf{0.3686}_{3}$	0.5469_{18}	0.4795_{15}	0.4633_{12}	0.4320_{7}	$\mathbf{0.3319}_{1}$
	SROCC↑	0.7584_{16}	0.8861_{14}	0.9263_{5}	0.7547_{17}	0.9222_{9}	0.7330_{18}	0.8978_{12}	$\mathbf{0.9338}_{3}$	0.5912_{20}	0.6994_{19}	$\mathbf{0.9465}_{2}$	0.9220_{10}	0.9248_{6}	0.9230_{8}	0.8555_{15}	0.8974_{13}	0.9235_{7}	0.9319_{4}	0.9189_{11}	$\mathbf{0.9483}_{1}$
VCL	KROCC↑	0.5592_{17}	0.6964_{14}	0.7541_{5}	0.5646_{16}	0.7447_{9}	0.5525_{18}	0.7141_{13}	$\mathbf{0.7634}_{3}$	0.4551_{20}	0.5150_{19}	$\mathbf{0.7887}_{2}$	0.7434_{10}	0.7516_{6}	0.7510_{7}	0.6724_{15}	0.7182_{12}	0.7449_{8}	0.7605_{4}	0.7432_{11}	$\mathbf{0.7935}_{1}$
2012	PLCC↑	0.7607_{16}	0.8677_{14}	0.9173_{5}	0.7524_{17}	0.8920_{11}	0.7313_{18}	0.8743_{13}	$\mathbf{0.9210}_{3}$	0.6085_{20}	0.7059_{19}	$\mathbf{0.9441}_{2}$	0.8822_{12}	0.9054_{8}	0.9173_{6}	0.8385_{15}	0.8954_{10}	0.9072_{7}	0.9189_{4}	0.9030_{9}	$\mathbf{0.9518}_{1}$
	RMSE↓	15.999_{16}	12.253_{14}	9.8170_{6}	16.236_{17}	11.142_{11}	16.812_{18}	11.965_{13}	$\mathbf{9.6034}_{3}$	19.561_{20}	17.460_{19}	$\mathbf{8.1247}_{2}$	11.606_{12}	10.463_{8}	9.8156_{5}	13.431_{15}	10.975_{10}	10.373_{7}	9.7236_{4}	10.590_{9}	$\mathbf{7.5579}_{1}$
	SROCC↑	0.6125_{20}	0.7817_{17}	0.8510_{12}	0.8051_{15}	0.8965_{4}	0.8092_{14}	$\mathbf{0.9050}_{2}$	0.8926_{7}	0.6650_{19}	0.7481_{18}	0.8732_{10}	0.8948_{5}	$\mathbf{0.9043}_{3}$	0.8115_{13}	0.7999_{16}	0.8548_{11}	0.8926_{6}	0.8785_{9}	$\mathbf{0.9063}_{1}$	0.8796_{8}
TID13	KROCC↑	0.4425_{20}	0.5938_{17}	0.6665_{11}	0.6259_{16}	0.7183_{5}	0.6304_{14}	$\mathbf{0.7324}_{1}$	0.7162_{7}	0.4993_{19}	0.5622_{18}	0.6923_{10}	0.7200_{4}	$\mathbf{0.7249}_{3}$	0.6433_{13}	0.6275_{15}	0.6661_{12}	0.7166_{6}	0.6969_{9}	$\mathbf{0.7307}_{2}$	0.6981_{8}
2013	PLCC↑	0.6721_{20}	0.8128_{17}	0.8742_{11}	0.8518_{15}	0.8967_{7}	0.8565_{14}	0.9000_{5}	$\mathbf{0.9060}_{2}$	0.7717_{19}	0.7971_{18}	0.8932_{9}	0.8993_{6}	0.9019_{4}	0.8667_{12}	0.8574_{13}	0.8432_{16}	0.8965_{8}	0.8914_{10}	$\mathbf{0.9072}_{1}$	$\mathbf{0.9020}_{3}$
	RMSE↓	0.9185_{20}	0.7226_{17}	0.6024_{11}	0.6499_{15}	0.5491_{7}	0.6403_{14}	0.5408_{5}	$\mathbf{0.5250}_{2}$	0.7889_{19}	0.7491_{18}	0.5578_{9}	0.5424_{6}	0.5358_{4}	0.6189_{12}	0.6384_{13}	0.6668_{16}	0.5495_{8}	0.5622_{10}	$\mathbf{0.5218}_{1}$	$\mathbf{0.5357}_{3}$
	SROCC↑	0.7519_{11}	0.7465_{12}	0.7596_{8}	0.7157_{18}	0.7185_{17}	0.7342_{13}	0.7790_{7}	0.7912_{4}	0.6998_{19}	0.7205_{16}	$\mathbf{0.8339}_{2}$	0.7272_{15}	$\mathbf{0.8047}_{3}$	0.7806_{6}	0.7531_{10}	0.6881_{20}	0.7535_{9}	0.7338_{14}	0.7828_{5}	$\mathbf{0.8797}_{1}$
CIDIQ	KROCC↑	0.5657_{9}	0.5537_{12}	0.5670_{8}	0.5219_{19}	0.5277_{17}	0.5368_{14}	0.5812_{7}	0.6005_{4}	0.5235_{18}	0.5278_{16}	$\mathbf{0.6484}_{2}$	0.5325_{15}	$\mathbf{0.6137}_{3}$	0.5813_{6}	0.5615_{10}	0.5185_{20}	0.5553_{11}	0.5435_{13}	0.5912_{5}	$\mathbf{0.7002}_{1}$
2014	PLCC↑	0.7501_{6}	0.7081_{15}	0.7479_{7}	0.7105_{14}	0.7009_{18}	0.7284_{11}	0.7449_{8}	$\mathbf{0.7857}_{3}$	0.7073_{17}	0.7169_{13}	$\mathbf{0.8314}_{2}$	0.7077_{16}	0.7553_{5}	0.7790_{4}	0.6642_{19}	0.6206_{20}	0.7377_{9}	0.7235_{12}	0.7365_{10}	$\mathbf{0.8776}_{1}$
	RMSE↓	1.0183_{6}	1.0871_{15}	1.0221_{7}	1.0834_{14}	1.0981_{18}	1.0550_{11}	1.0271_{8}	$\mathbf{0.9524}_{3}$	1.0883_{17}	1.0734_{13}	$\mathbf{0.8555}_{2}$	1.0877_{16}	1.0090_{5}	0.9654_{4}	1.1509_{19}	1.2103_{20}	1.0394_{9}	1.0629_{12}	1.0414_{10}	$\mathbf{0.7198}_{1}$
	SROCC↑	0.6271_{20}	0.7610_{17}	0.8509_{11}	0.8413_{13}	0.8747_{4}	0.8367_{14}	0.8489_{12}	0.8664_{9}	0.7165_{19}	0.8231_{15}	$\mathbf{0.8849}_{1}$	0.8685_{6}	0.8681_{7}	0.8606_{10}	0.7718_{16}	0.7235_{18}	0.8677_{8}	0.8717_{5}	$\mathbf{0.8839}_{2}$	$\mathbf{0.8762}_{3}$
KADID	KROCC↑	0.4552_{20}	0.5710_{17}	0.6604_{11}	0.6573_{12}	0.6839_{5}	0.6528_{14}	0.6560_{13}	0.6755_{10}	0.5403_{19}	0.6391_{15}	$\mathbf{0.6994}_{1}$	0.6785_{7}	0.6773_{9}	0.6775_{8}	0.5847_{16}	0.5406_{18}	0.6780_{7}	0.6842_{4}	$\mathbf{0.6983}_{2}$	$\mathbf{0.6908}_{3}$
2019	PLCC↑	0.6192_{20}	0.7401_{17}	0.8396_{11}	0.8400_{10}	0.8592_{6}	0.8330_{13}	0.7744_{15}	0.8432_{9}	0.7315_{18}	0.8249_{14}	$\mathbf{0.8846}_{1}$	0.8345_{12}	0.8497_{8}	0.8598_{5}	0.7476_{16}	0.7162_{19}	0.8545_{7}	0.8613_{4}	$\mathbf{0.8685}_{3}$	$\mathbf{0.8752}_{2}$
	RMSE↓	0.8503_{20}	0.7282_{17}	0.5882_{11}	0.5876_{10}	0.5541_{6}	0.5991_{13}	0.6851_{15}	0.5822_{9}	0.7383_{18}	0.6122_{14}	$\mathbf{0.5050}_{1}$	0.5967_{12}	0.5709_{8}	0.5529_{5}	0.7191_{16}	0.7557_{19}	0.5625_{7}	0.5502_{4}	$\mathbf{0.5367}_{3}$	$\mathbf{0.5238}_{2}$
	SROCC↑	0.7652_{20}	0.8552_{16}	0.8940_{8}	0.8577_{15}	0.8884_{9}	0.8603_{13}	0.8837_{12}	0.9071_{4}	0.7971_{19}	0.8376_{17}	$\mathbf{0.9125}_{2}$	0.8877_{10}	0.9069_{5}	0.8875_{11}	0.8603_{14}	0.8341_{18}	0.8949_{6}	0.8944_{7}	$\mathbf{0.9076}_{3}$	$\mathbf{0.9251}_{1}$
Média	KROCC↑	0.5803_{20}	0.6781_{16}	0.7250_{6}	0.6855_{15}	0.7127_{10}	0.6916_{13}	0.7044_{12}	0.7407_{4}	0.6356_{19}	0.6629_{17}	$\mathbf{0.7478}_{2}$	0.7117_{11}	0.7371_{5}	0.7185_{9}	0.6863_{14}	0.6543_{18}	0.7208_{8}	0.7232_{7}	$\mathbf{0.7408}_{3}$	$\mathbf{0.7682}_{1}$
	PLCC↑	0.7689_{20}	0.8450_{16}	0.8899_{7}	0.8630_{13}	0.8754_{10}	0.8652_{12}	0.8602_{14}	$\mathbf{0.9021}_{3}$	0.8197_{18}	0.8445_{17}	$\mathbf{0.9160}_{2}$	0.8704_{11}	0.8921_{6}	0.8939_{5}	0.8480_{15}	0.8182_{19}	0.8849_{9}	0.8888_{8}	0.8954_{4}	$\mathbf{0.9304}_{1}$
Top-1	Todas	0	0	0	0	0	0	1	1	0	0	8	0	0	0	0	0	0	0	3	19
Top-3	Todas	0	0	2	0	0	0	2	16	4	0	20	0	4	4	4	0	0	0	10	30

De acordo com os resultados experimentais apresentados na Tabela 7, PEL supera todos os métodos nas BDIs IVC, CSIQ, MICT, VCL e CIDIQ. Nas BDIs LIVE e KADID, PEL apresenta desempenho ligeiramente inferior apenas ao método HaarPSI. A BDI TID13 registra, por meio das métricas SROCC e KROCC, resultados de PEL menos eficazes, mas ainda assim alcança a oitava posição. Vale destacar, no entanto, que para esta mesma base, usando as métricas PLCC e RMSE, PEL é classificado como Top-3.

Nesse contexto, Mansouri e Aznaveh [76] destacam que o desempenho dos métodos de IQA do estado da arte é altamente dependente das bases de dados utilizadas para avaliação. Por exemplo, o método SCQI avalia a qualidade da imagem consistente com pontuações subjetivas na BDI TID2013. No entanto, este método apresenta baixo desempenho para as outras BDIs.

Contrariando tal afirmação, o método proposto nesta tese apresenta uma estabilidade muito alta, competindo consistentemente pelas primeiras posições em todas as bases de dados utilizadas, comprovando assim sua robustez. Diante do exposto, as pontuações objetivas previstas por PEL correlacionam-se mais fortemente com as avaliações subjetivas do que todos os outros métodos, demonstrando assim sua grande acurácia, monotonicidade e consistência.

6.3.2.2 Comparação de Desempenho sobre os Tipos Individuais de Distorções da BDI TID2013

Esta seção apresenta as comparações de desempenho para os métodos Top-5, conforme a Tab. 7, quanto às distorções individuais da BDIs TID2013.

A necessidade desse detalhamento de desempenho objetiva entender o motido pelo qual o método PEL ter sido menos eficaz para a BDI TID2013, a fim que futuras melhorias possam ser aplicadas.

A Tabela 8 apresenta os resultados de desempenho, onde na primeira coluna são descritas as distorções individuais existentes na BDI TID2013, bem como nas colunas seguintes são listados os valores dos métodos Top-5 conforme a métrica estatística PLCC. Nesta tabela também são usados os índices subscritos de classificação, bem como os Top-1 e Top-3 para os cinco métodos avaliados.

Conforme os resultados, nota-se que o método PEL possui desempenho bem abaixo dos outros métodos para as distorções Masked noise (MN), Impulse Noise (IN), Mean Shift (MS) e Contrast change (CTC), provocando dessa forma uma diferença significativa a ponto de impactar o resultado geral de PEL sobre a BDI TID2013.

Esse impacto pode ser melhor entendido com base nos mapas da Fig. 57 produzidos por PEL e demais métodos comparados para essas quatro distorções citadas anteriomente. Vale ressaltar, que os autores do método PSIM não disponibilizaram o código-fonte, mas apenas o executável, logo não foi possível adicionar a linha de código necessária para gerar seus mapas sobre tais distorções.

A primeira coluna da Fig. 57 mostra os mapas referente à distorção Masked noise (MN). Nota-se que o método PEL se mostrou sensível às distorções MN fora da RoI. Conforme o mapa PEL em (r) da Fig. 57, o método PEL se mostrou bastante sensível as distorções Impulse Noise (IN), a ponto de considerar os píxeis vizinhos como também sendo ruidososos, influenciando, dessa forma, seu valor total

de índice de qualidade. A distorção existente em (c) da figura em análise possui uma diferença sensível de crominância ao HVS, mas apesar do método PEL conseguir perceber tal mudança, percebe-se que a intensidade de tal distorção foi pontencilizada no resultado final do mapa, e conseguentemente, do seu índice. A imagem em (d) foi afetada apenas por uma mudança de contraste, mas o método PEL quase não percebe. Neste caso, SPSIM e, sua extensão, SSVS conseguem perceber bem tal mudança de contraste, como podem ser vistos nos mapas (l) e (p) da Fig. 57, respectivamente.

Pode-se inferir dois pontos principais das análises anterior. O primeiro ponto é que as características de crominância estão sendo superdimensinadas pelo método PEL para distorções ruidosas coloridas do tipo MN e IN, logo o peso α precise ser ajustado exclusivamente para o produto $AD_I \cdot AD_Q$ na Eq. 5.15. Segundo, que o potencial de contraste na Eq. 5.15, quando combinado com a soma das diferenças de luminância de AD_G e AD_L, não está conseguindo funcionar de forma eficaz para capturar pequenas diferenças de contrastes, como aconteceram para imagens com distorções do tipo MS e CTC como em (c) e (d) da Fig. 57.

Tabela 8 – Comparação de avaliação de desempenho conforme a métrica PLCC para os métodos IQA Top-5 na BDI TID2013.

Distorção	**PSIM**	**HaarPSI**	**SPSIM**	**SSVS**	**PEL**
Additive white Gaussian noise (AGN)	**0.9532**$_1$	**0.9453**$_2$	0.9328$_5$	0.9349$_4$	**0.9364**$_3$
Additive noise in color (ANC)	**0.9200**$_1$	**0.9094**$_2$	0.8891$_5$	0.8953$_4$	**0.9071**$_3$
Spatially correlated noise (SCN)	**0.9510**$_1$	**0.9388**$_2$	0.9207$_5$	0.9222$_4$	**0.9376**$_3$
Masked noise (MN)	**0.8249**$_1$	**0.8204**$_2$	**0.8040**$_3$	0.7958$_4$	0.7573$_5$
High frequency noise (HFN)	**0.9699**$_1$	**0.9603**$_2$	0.9450$_5$	**0.9507**$_3$	0.9468$_4$
Impulse noise (IN)	0.8038$_4$	**0.8625**$_1$	**0.8132**$_3$	**0.8391**$_2$	0.6885$_5$
Quantization noise (QN)	**0.9216**$_1$	0.8906$_4$	**0.8918**$_3$	0.8691$_5$	**0.8993**$_2$
Gaussian blur (GB)	0.9148$_4$	0.9121$_5$	**0.9282**$_2$	**0.9407**$_1$	**0.9231**$_3$
Image denoising (DEN)	**0.9787**$_1$	0.9718$_4$	**0.9746**$_2$	0.9711$_5$	**0.9726**$_3$
JPEG compression (JPEG)	**0.9793**$_1$	**0.9791**$_2$	0.9777$_4$	**0.9780**$_3$	0.9708$_5$
JPEG2000 compression (JP2K)	**0.9824**$_3$	**0.9825**$_2$	0.9809$_4$	**0.9842**$_1$	0.9797$_5$
JPEG transmission errors (JGTE)	0.9200$_4$	**0.9220**$_2$	0.9179$_5$	**0.9427**$_1$	**0.9214**$_3$
JPEG2000 transmission errors (J2TE)	0.9067$_4$	**0.9102**$_2$	0.9014$_5$	**0.9072**$_3$	**0.9134**$_1$
Non eccentricity pattern noise (NEPN)	0.8387$_4$	**0.8419**$_3$	0.8165$_5$	**0.8424**$_2$	**0.8474**$_1$
Local block-wise distortions (Block)	**0.5634**$_2$	**0.4952**$_3$	0.3376$_5$	**0.6390**$_1$	0.4215$_4$
Mean shift (MS)	**0.7980**$_3$	0.7373$_4$	**0.8088**$_2$	**0.8213**$_1$	0.7206$_5$
Contrast change (CTC)	0.5832$_4$	**0.7089**$_3$	**0.8261**$_1$	**0.8244**$_2$	0.4683$_5$
Change of color saturation (CCS)	**0.7686**$_3$	0.6840$_5$	**0.7874**$_2$	**0.8000**$_1$	0.7413$_4$
Multiplicative Gaussian noise (MGN)	**0.9151**$_1$	**0.8967**$_2$	0.8677$_5$	**0.8883**$_3$	0.8842$_4$
Comfort noise (CN)	0.9546$_4$	**0.9579**$_2$	0.9442$_5$	**0.9549**$_3$	**0.9585**$_1$
Lossy compression of noisy images (LCNI)	0.9674$_4$	**0.9692**$_2$	0.9598$_5$	**0.9678**$_3$	**0.9765**$_1$
Image color quantization with dither (TCOD)	**0.9351**$_1$	0.9101$_4$	**0.9195**$_3$	0.9018$_5$	**0.9253**$_2$
Chromatic aberrations (CHA)	0.9715$_4$	0.9619$_5$	**0.9776**$_3$	**0.9790**$_1$	**0.9783**$_2$
Sparse sampling and reconstruction (SSR)	**0.9826**$_2$	0.9813$_5$	**0.9835**$_1$	0.9814$_4$	**0.9822**$_3$
Top-1	10	1	2	7	4
Top-3	15	16	11	16	14

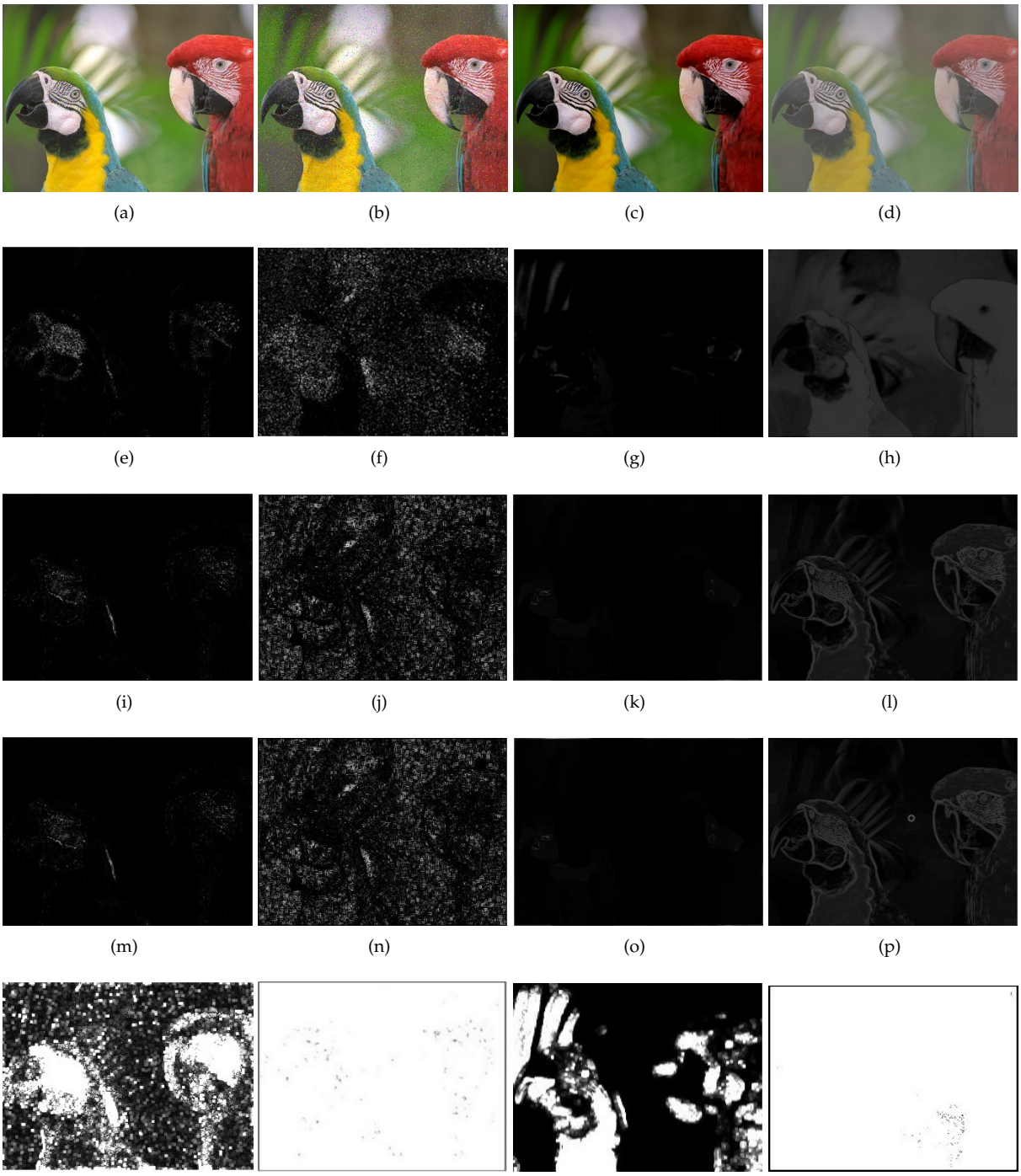


Figura 57 – (a-d) são imagens distorcidas a partir da imagem original I23 (parrots) da BDI TID2013 [136], que foram corrompidas, respectivamente, com Masked noise (MN), Impulse Noise (IN), Mean Shift (MS) e Contrast change (CTC). Mapas de Saída (e), (i), (m) e (q) produzidos, respectivamente, pelos métodos HaarPSI, SPSIM, SSVS e PEL sobre a imagem (a). Mapas de saída (f), (j), (n) e (r) produzidos, respectivamente, pelos métodos HaarPSI, SPSIM, SSVS e PEL sobre a imagem (b), e assim por diante.
Fonte: Autoria própria.

6.3.3 Significância Estatística dos Métodos

A significância estatística dos resultados (F%) de PEL em relação aos seus dezenove métodos concorrentes é apresentada na Tabela 9. Esses resultados foram calculados de acordo com as especificações descritas na Seção 4.3.3.

Os resultados de pontuação obtidos por todos os métodos estão resumidos no final da Tabela 9, de acordo com a tríade *PEL:≅:IQM*$_L$, que representa a comparação de significância entre os métodos. Os valores em PEL da tríade indicam a superioridade do método proposto em relação aos métodos da literatura (IQM_L); caso contrário, os valores em IQM_L da tríade são especificados. Por fim, os valores na parte central da tríade (≅) indicam o número de vezes em que não há diferença significativa entre os métodos comparados (ou seja, $F_{\%} \leq 5\%$).

Por exemplo, a última coluna da Tabela 9, que registra a comparação entre os métodos PEL e SSVS, apresenta a tríade 6:1:1, indicando que o método proposto é superior em 6 BDIs, respectivamente, IVC (26,87), LIVE (53,39), CSIQ (92,42), MICT (69,42), VCL (96,33) e CIDIQ (109,32), enquanto o método SSVS é superior apenas para a BDI TID13 (-5,40). Por outro lado, não há diferença significativa entre a eficácia dos métodos PEL e SSVS em relação à BDI KADID ($F_{\%} = 4.99 \leq 5\%$).

Em geral, o método proposto é superior para todos as BDIs quando comparado aos métodos SSIM, MS-SSIM, FSIM, GMSD, VSI, RVSIM, SSRM, SSVD, SUMMER, VCGS e DSSIM. Em relação aos métodos MCSD, SCQI, PSIM, EFS, SPSIM e CVSS, PEL é superior em sete BDIs e empata em apenas um. Esses empates indicam que não há diferença significativa entre PEL e esses métodos para as BDIs LIVE (4,77 e 3,60), TID13 (1,91, -4,12, 2,52 e 0,04) e KADID (4,99). Comparando os métodos PEL e HaarPSI, observa-se que PEL é significativamente melhor que o HaarPSI nos bancos de dados IVC, CSIQ, MICT, VCL, TID13 e CIDIQ, exceto para LIVE e KADID.

Na Tab. 9, o total de comparações estatísticas é 152, onde PEL supera 142 vezes todos os outros métodos, ou seja, o método proposto supera todas as outras abordagens em 93,42% dos testes de significância.

6.4 Custo Computacional versus Análise de Desempenho

Na Seção 4.3.3 são detlhados os fundamentos referentes a necessidade de uma análise para uma comparação entre a análise de desempenho e o custo computacional entre os métodos, bem como os tipos de gráficos utilizados para uma melhor compreensão dessas comparações. Nesse sentido, nesta seção são apresentados os resultados e as análises de tais comparações no contexto do método PEL proposto.

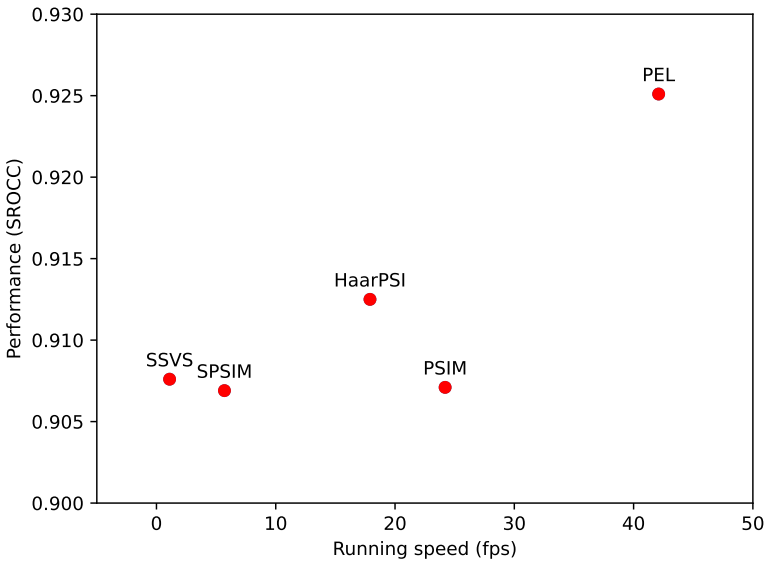
A Fig. 58 mostra a comparação entre os métodos FR-IQA da literatura e o método proposto em relação ao Top-5 em termos de desempenho (SROCC e KROCC gerais na Tabela 7) e custo computacional em relação à velocidade de execução (fps na Tabela 6).

Tabela 9 – Estatística $F_{\%}$ usada para a significância da diferença entre $RMSE_{MAX}$ / $RMSE_{MIN}$ dos IQMs sobre as oito bases de dados de imagens (BDIs) utilizadas.

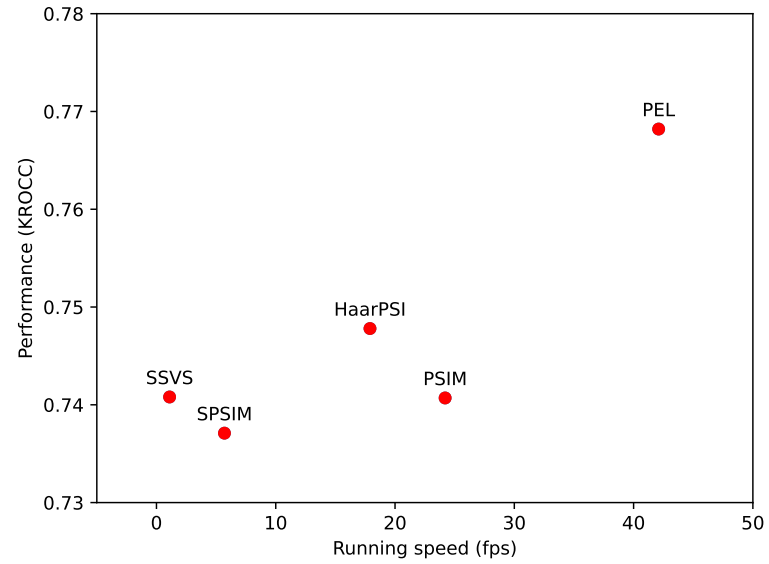
BDI	$F_{\%}$	**SSIM**	**MSSIM**	**FSIM$_C$**	**GMSD**	**VSI**	**MCSD**	**SCQI**	**PSIM**	**RVSIM**	**SSRM**	**HaPSI**	**EFS**	**SPSIM**	**CVSS**	**SSVD**	**SUMER**	**VCGS**	**DSSIM**	**SSVS**
IVC	**PEL**	333.28	87.67	27.59	45.21	83.78	40.29	116.30	22.16	66.47	81.22	49.11	92.98	29.74	101.06	50.60	178.07	45.01	41.54	26.87
LIVE	**PEL**	149.21	88.42	49.56	19.19	141.33	4.77	163.86	58.77	28.74	30.97	-9.81	147.62	76.03	3.60	84.03	154.15	117.60	93.35	53.39
CSIQ	**PEL**	424.40	311.44	158.69	57.28	137.04	64.10	136.12	29.06	158.69	136.58	29.06	173.81	120.76	30.42	287.54	406.75	135.66	91.59	92.42
MICT	**PEL**	224.20	58.23	38.22	99.08	109.85	59.60	196.70	89.92	12.54	73.12	80.57	98.23	79.60	113.45	23.34	171.52	108.72	94.85	69.42
VCL	**PEL**	348.13	162.84	68.70	361.47	117.32	394.80	150.62	61.45	569.83	433.70	15.56	135.79	91.65	68.67	215.81	110.87	88.35	65.52	96.33
TID13	**PEL**	193.98	81.95	26.45	47.18	5.07	42.86	1.91	-4.12	116.87	95.54	8.42	2.52	0.04	33.47	42.02	54.93	5.22	10.14	-5.40
CIDIQ	**PEL**	100.14	128.09	101.63	126.54	132.73	114.82	103.61	75.07	128.60	122.38	41.26	128.35	96.50	79.88	155.65	172.82	108.52	118.05	109.32
KADID	**PEL**	163.52	93.27	26.10	25.84	11.90	30.82	71.07	23.54	98.67	36.60	-7.58	29.77	18.79	11.42	88.47	108.15	15.32	10.33	4.99
Todas	**PEL:≅:IQM$_L$**	8:0:0	8:0:0	8:0:0	8:0:0	8:0:0	7:1:0	7:1:0	7:1:0	8:0:0	8:0:0	6:0:2	7:1:0	7:1:0	7:1:0	8:0:0	8:0:0	8:0:0	8:0:0	6:1:1

Como mostrado nas Figs. 58(a) e 58(b), o método proposto alcança o primeiro lugar entre os métodos Top-5. Vale ressaltar que o custo-benefício de PEL se sobressai em relação aos demais métodos comparados.

Por fim, é importante ressaltar que o método proposto apresenta alta estabilidade, competindo consistentemente pelas primeiras posições em todas as bases de dados de imagens utilizadas, comprovando assim a sua robustez, além de sua eficiência para avaliar as imagens em tempo real.



(a)



(b)

Figura 58 – Comparações dos 5 principais métodos FR-IQA em termos de desempenho (SROCC e KROCC geral) e custo computacional (fps): (a) SROCC e (b) KROCC.
Fonte: Autoria própria.

7

Conclusão e Trabalhos Futuros

Nesta tese, foi proposto um novo método analítico perceptual voltado para a avaliação da qualidade de imagens com referência completa (FR-IQA), denominado Logaritmo do Erro Perceptual (Perceptual Error Logarithm, PEL).

O novo método PEL atende a um dos chamados da comunidade científica, apontado por diversas pesquisas, como em [62], [63], [64], [65], [66], [67], [68], [69], [70], que é a necessidade da criação de novos métodos objetivos para IQA que sejam mais eficientes e eficazes do que os métodos descritos atualmente na literatura. Nesse sentido, foi constituído um processo metodológico com três passos para criação do método PEL, a saber: (1) Investigação de BDI e métodos existentes na literatura voltados à FR-IQA; (2) Descrição do Processo Computacional do método proposto; e (3) Avaliação do método proposto para verificar sua eficiência e eficácia competitiva em relação aos métodos do estado da arte sob as BDI voltadas à FR-IQA.

Nesse contexto, o Capítulo 4 apresenta, como resultado da investigação, os materiais utilizados nesta tese, constituídos por oito (8) BDI (IVC [123], LIVE [12], CSIQ [134], MICT [124], VCL [135], TID13 [136], CIDIQ [137] e KADID [138]) e dezenove (19) métodos objetivos (SSIM [21], MS-SSIM [22], FSIM$_C$ [24], GMSD [71], VSI [120], MCSD [16], SCQI [126], PSIM [125], RVSIM [73], SSRM [127], HaarPSI [128], EFS [74], SPSIM [75], CVSS [72], SSVD [76], SUMMER [133], VCGS [77], DSSIM [129] e SSVS [78]). Vale ressaltar que o Capítulo 3 apresenta uma visão geral desses métodos, incluindo seus processos computacionais e suas habilidades perceptuais preliminares.

O Capítulo 5 apresenta todos os detalhes do processo computacional do método PEL proposto. Cada etapa de PEL é fundamentada com base em características de métodos analíticos tradicionais e perceptuais existentes, utilizadas nesta tese em sua totalidade ou às vezes alteradas, a fim que fossem combinadas de forma harmoniosa.

Já para comparar PEL em relação aos dezenove métodos estudados da literatura, foram realizadas avaliações com base na Seção 4.3. Os resultados detalhados dessas avaliações, que se encontram no Capítulo 6, verificaram as habilidades perceptuais com relação à eficiência (custo computacional), eficácia (análise de desempenho) e custo-benefício (custo computacional versus análise de desempenho)

dos vinte métodos comparados frente às oito BDIs envolvidas.

A Seção 6.2 apresenta os resultados e discussões do custo computacional sobre os 20 métodos avaliados. O método proposto alcança a quarta posição, ligeiramente superado apenas pelos métodos GMSD, MCSD e CVSS, respectivamente, 2,10, 1,67 e 1,24 vezes. No entanto, vale destacar que esses métodos não utilizam recursos de crominância, bem como demonstram desvantagens, conforme as Seções 6.3 e 6.4, em comparação com o método PEL quanto à habilidade perceptual em nível de desempenho e custo-benefício. Além disso, foi demonstrado que o método PEL consegue processar, com uma velocidade média de execução em quadro por segundo de 42,1 fps, as avaliações sobre as imagens existentes nas oito BDIs. Com isso, o método proposto pode ser utilizado em aplicações e modelos que exijam processamento em tempo real sobre imagens e vídeos.

Para analisar a eficácia dos métodos avaliados, foram necessárias realizar análises de desempenho, como demonstrado na Seção 6.3, envolvendo três etapas: comparação entre valores subjetivos e objetivos normalizados; predição da avaliação estatística dos métodos; e significância estatística dos métodos.

Na primeira etapa, Seção 6.3.1, foram realizadas comparações entre os valores subjetivos e objetivos com base em uma visão específica e geral. Na *visão específica*, detalhada na Seção 6.3.1.1, os valores são apresentados visualmente por meio de gráficos de dispersão de DMOS subjetivo versus pontuações obtidas pelos métodos sobre todas as distorções conforme a BDI analisada. Em geral, o método proposto apresenta destaque diferenciado em relação aos seus concorrentes nas BDIs IVC, CSIQ, MICT, VCL e CIDIQ. Enquanto nas BDIs LIVE e KADID os métodos Top-3 apresentam poucas diferenças quanto às suas distribuições. Em relação à BDI TID2013, os métodos SSVS, SCQI, PSIM, SPSIM e PEL alternam entre si quanto às primeiras posições, como pode ser atestado na Fig. 53, com leve vantagem para o método SSVS. Já na *visão geral*, detalhada na Seção 6.3.1.2, foram realizadas comparações de desempenho dos 20 métodos em estudo nas oito BDIs investigadas. Tais comparações são apresentadas na Fig. 56, que mostra gráficos de caixa (*box plots*) das relações $DMOS$x$DMOS_p$, que comprovam, em geral, que as pontuações objetivas previstas pelo método proposto são muito mais consistentes com as avaliações subjetivas do que todos os dezenove métodos de IQA investigados da literatura.

Na segunda etapa, foram medidas as habilidades perceptuais dos métodos quanto ao seu desempenho da predição em IQA, detalhada na Seção 6.3.2, onde para isso foram utilizadas as métricas estatísticas adotadas nesta tese, sugeridas pelo VQEG, a saber: SROCC, KROCC, PLCC e RMSE. Os resultados dessa etapa são consolidados na Tab. 7, onde, em média, o método proposto alcança as primeiras posições para as métricas estatísticas. O método PEL alcança 19 posições como Top-1, frente ao segundo colocado HaarPSI, com 8 posições, e o terceiro colocado com apenas 3 posições. A Tab. 7 também apresenta os Top-3, inclusive com grandes vantagens para o método proposto.

A terceira parte, Seção 6.3.3, refere-se à significância estatística dos resultados (F%) de PEL em relação aos seus dezenove métodos concorrentes, onde as comparações são consolidadas na Tab. 9. O total de comparações estatísticas é 152, onde PEL supera 142 vezes todos os outros métodos, ou seja, o

método proposto supera todas as outras abordagens em 93,42% dos testes de significância.

Análise de desempenho versus custo computacional do método PEL e seus dezenove concorrentes foi apresentada na Seção 6.4. A Fig. 58 apresenta uma comparação entre os métodos Top-5, em termos de desempenho (SROCC e KROCC gerais na Tabela 7), e custo computacional, em relação à velocidade de execução (fps na Tabela 6). Com base nos dados apresentados, pode-se concluir que o método proposto alcança o primeiro lugar entre os métodos Top-5 quanto ao Custo-Benefício.

Diante do exposto, conclui-se que os resultados experimentais, análises e discussões demonstraram que o método proposto alcança habilidades perceptuais (eficácia e eficiência) acima de seus dezenove métodos concorrentes investigados.

Sendo assim, espera-se que PEL possa ser utilizado pela comunidade científica e industrial como método em IQA e VQA, bem como subsidiar diversas técnicas analíticas, em aplicações multimídia, ou baseadas em aprendizado, como CNNs e GANs, focadas em imagens e vídeos, que precisem garantir o processamento, a transmissão e o armazenamento de grandes volumes de dados, inclusive em tempo real.

7.1 Trabalhos Futuros

Durante o caminhar para a concepção desta tese, foram observadas algumas limitações do método PEL, que poderiam receber mais atenção, a fim de melhorar sua eficácia. Por outro lado, PEL poderia ser utilizado em outras áreas apontadas na Seção 1.1 para aprimorar modelos baseados em conhecimento. Nesse sentido, como trabalhos futuros, podem ser consideradas as seguintes sugestões:

i) Verificar como melhorar as características de crominância utilizadas pelo método PEL, que superdimensiona as distorções ruidosas coloridas do tipo MN e IN, como demonstrado na Seção 6.3.2.2. Uma sugestão seria utilizar o peso α para ajustar exclusivamente o produto $AD_I \cdot AD_Q$ na Eq. 5.15;

ii) O potencial de contraste na Eq. 5.15, quando combinado com a soma das diferenças de luminância AD_G e AD_L, não está conseguindo funcionar de forma eficaz para capturar pequenas diferenças de contrastes, como aconteceram para imagens com distorções do tipo MS e CTC como em (c) e (d) da Fig. 57. Uma sugestão seria considerar o contraste como parte da formação do mapa de Erro Perceptual (PE) na Eq. 5.15 ou mesmo ser combinado de uma outra forma com os mapas AD_G e AD_L;

iii) Na Seção 5.2.3 foram levantadas algumas limitações quanto ao tratamento de PEL em relação às RoI em imagens. Nesse sentido, seria de grande valia a inclusão de uma etapa no processo de PEL que pudesse detectar as RoI em imagens. Dessa forma, o método PEL poderia ser orientado com mais precisão a valorizar as bordas fortes e desconsiderar ou minimizar a influência das bordas fracas em RoI;

iv) O método PEL poderia ser utilizado tanto como IQA quanto função de perda (*loss function*) em modelos baseados em aprendizagem voltados para recuperação de imagens, super-resolução de imagens, imagens médicas e compressão de imagens;

v) Uma outra sugestão seria acrescentar uma etapa ao processo de PEL que pudesse medir o impacto de bordas em excesso ou ausentes em imagens distorcidas em relação as suas imagens de referência, inclusive em RoI;

vi) Uma sugestão bem diferenciada, em relação às apontadas para melhorar a eficácia do método PEL, seria considerar o uso de mapas com funções baseadas no HVS, como o potencial de ação de células ganglionares utilizadas em Bezerra *et al.* [184].

vii) Por fim, poderia-se utilizar modelos de aprendizagem para fornecer o índice de qualidade de PEL ao invés do cálculo analítico aplicado na Eq. 5.15.

REFERÊNCIAS

[1] Sung Cheol PARK, Min Kyu PARK, and Moon Gi KANG. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003. 20

[2] Younghyun JO, Sejong YANG, and Seon Joo KIM. Investigating loss functions for extreme super-resolution. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1705–1712, 2020. 20, 32

[3] Hongyu ZHU, Chao XIE, Yeqi FEI, and Huanjie TAO. Attention mechanisms in cnn-based single image super-resolution: A brief review and a new perspective. *Electronics*, 10(10), 2021. 20, 32

[4] Dawa Chyophel LEPCHA, Bhawna GOYAL, Ayush DOGRA, and Vishal GOYAL. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*, 91:230–260, 2023. 20, 23, 32

[5] Hanadi AL-MEKHLAFI and Shiguang LIU. Single image super-resolution: a comprehensive review and recent insight. *Frontiers of Computer Science*, 18(1), 2003. 20, 32

[6] Tiesong ZHAO, Yuting LIN, Yiwen XU, Weiling CHEN, and Zhou WANG. Learning-based quality assessment for image super-resolution. *IEEE Transactions on Multimedia*, 24:3570–3581, 2022. 20, 23

[7] S. WINKLER. *Digital Video Quality - Vision Models and Metrics*. John Wiley Sons, England, 2005. 20, 32

[8] U. ENGELKE and H. ZEPERNICK. Perceptual-based quality metrics for image and video services: A survey. In *2007 Next Generation Internet Networks*, pages 190–197, Trondheim, Norway, 2007. 20

[9] Philip CORRIVEAU. Video quality testing. In H. R. Wu and K. R. Rao, editors, *Digital Video Image Quality and Perceptual Coding*, page 125–153. CRC Press, USA, 2006. 20, 35, 37

[10] Video Quality Experts Group. *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II*. Video Quality Experts Group (VQEG), August 2003. 21, 38, 43, 64, 72, 74, 75, 90, 93

[11] Video Quality Experts Group. *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Multimedia Quality Assessment, Phase I*. Video Quality Experts Group (VQEG), September 2008. 21, 38, 43, 44, 64, 72, 75, 90, 93

[12] H. R. SHEIKH, M. F. SABIR, and A. C. BOVIK. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, Nov. 2006. 21, 47, 64, 66, 81, 82, 85, 91, 113

[13] Z. WANG and A. C. BOVIK. *Modern Image Quality Assessment*. Morgan & Claypool, USA, 2006. 21, 45

[14] Z. WANG, A. C. BOVIK, H. R. SHEIKH, and E. P. SIMONCELLI. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. 21, 39

[15] H. R. SHEIKH and A. C. BOVIK. Information theoretic approaches to image quality assessment. In Al Bovik, editor, *Handbook of image & Video Processing*, page 975–989. Elsevier, USA, 2 edition, 2005. 21

[16] Tonghan WANG, Lu ZHANG, Huizhen JIA, Baosheng LI, and Huazhong SHU. Multiscale contrast similarity deviation: An effective and efficient index for perceptual image quality assessment. *Signal Processing: Image Communication*, 45:1–9, 2016. 21, 43, 49, 50, 71, 91, 113

[17] T. N. PAPPAS, R. J. SAFRANEK, and J. CHEN. Perceptual criteria for image quality evaluation. In Al Bovik, editor, *Handbook of image & Video Processing*, page 939–959. Elsevier, USA, 2 edition, 2005. 21, 38

[18] A. M. ESKICIOGLU and P. S. FISHER. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995. 21, 38

[19] R. Freitas ZAMPOLO and R. SEARA. A measure for perceptual image quality assessment. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, volume 1, pages I–433, 2003. 21, 38

[20] Z. WANG, L. LU, and A. C. BOVIK. Video quality assessment based on structural distortion measurement. *Signal Processing: Image Communication*, 19(2):121–132, February 2004. 21, 38

[21] Z. WANG and A. C. BOVIK. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002. 22, 39, 71, 87, 91, 113

[22] Z. WANG, E. SIMONCELLI, and A. C. BOVIK. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, page 1398–1402, November 2003. 22, 39, 71, 91, 113

[23] H. R. SHEIKH and A. C. BOVIK. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006. 22, 39

[24] Lin ZHANG, Lei ZHANG, Xuanqin MOU, and David ZHANG. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011. 22, 25, 39, 42, 48, 49, 71, 73, 79, 87, 91, 113

[25] Krzysztof OKARMA. Combined full-reference image quality metric linearly correlated with subjective assessment. In Leszek RUTKOWSKI, Rafał SCHERER, Ryszard TADEUSIEWICZ, Lotfi A. ZADEH, and Jacek M. ZURADA, editors, *Artificial Intelligence and Soft Computing*, pages 539–546, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. 22, 40, 56

[26] Krzysztof OKARMA. Colour image quality assessment using the combined full-reference metric. In *Computer Recognition Systems 4*, 2011. 56

[27] Oleg IEREMEIEV, Vladimir V. LUKIN, Nikolay N. PONOMARENKO, Karen O. EGIAZARIAN, and Jaakko ASTOLA. Combined full-reference image visual quality metrics. In Sos S. AGAIAN, Karen O. EGIAZARIAN, and Atanas P. GOTCHEV, editors, *Image Processing: Algorithms and Systems XIV, San Francisco, California, USA, February 14-18, 2016*, pages 1–10. Society for Imaging Science and Technology, 2016.

[28] Domonkos VARGA. Full-reference image quality assessment based on an optimal linear combination of quality measures selected by simulated annealing. *Journal of Imaging*, 8(8), 2022.

[29] Domonkos VARGA. An optimization-based family of predictive, fusion-based models for full-reference image quality assessment. *J. Imaging*, 9(6):116, 2023. 22, 40

[30] Shuigen WANG, Chenwei DENG, Weisi LIN, Guang-Bin HUANG, and Baojun ZHAO. NMF-based image quality assessment using extreme learning machine. *IEEE Transactions on Cybernetics*, 47(1):232–243, 2017. 22, 40

[31] Ethem ALPAYDIN. *Introduction to Machine Learning*. Adaptive Computation and Machine Learning series. The MIT Press, Cambridge, MA, 4th edition, 2020. 22, 40

[32] Kamran SHAUKAT, Suhuai LUO, Vijay VARADHARAJAN, Ibrahim A. HAMEED, and Min XU. A survey on machine learning techniques for cyber security in the last decade. *IEEE Access*, 8:222310–222354, 2020. 22, 40

[33] Shuigen WANG, Chenwei DENG, Weisi LIN, Guang-Bin HUANG, and Baojun ZHAO. NMF-based image quality assessment using extreme learning machine. *IEEE Transactions on Cybernetics*, 47(1):232–243, 2017. 22, 40

[34] Zhisen TANG, Yuanlin ZHENG, Ke GU, Kaiyang LIAO, Wei WANG, and Miaomiao YU. Full-reference image quality assessment by combining features in spatial and frequency domains. *IEEE Transactions on Broadcasting*, 65(1):138–151, 2019. 22, 25, 41, 60, 79, 82, 87

[35] Manish NARWARIA and Weisi LIN. SVD-based quality metric for image and video using machine learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2):347–364, 2012. 22, 40

[36] Sebastian BOSSE, Dominique MANIRY, Klaus-Robert MÜLLER, Thomas WIEGAND, and Wojciech SAMEK. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing*, 27(1):206–219, 2018. 22, 42

[37] Woojae KIM, Anh-Duc NGUYEN, Sanghoon LEE, and Alan Conrad BOVIK. Dynamic receptive field generation for full-reference image quality assessment. *IEEE Transactions on Image Processing*, 29:4219–4231, 2020. 23, 42, 43

[38] Keyan DING, Kede MA, Shiqi WANG, and Eero P. SIMONCELLI. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2567–2581, 2022. 23, 42

[39] Fei ZHOU, Wei SHENG, Zitao LU, Bo KANG, Mianyi CHEN, and Guoping QIU. Super-resolution image visual quality assessment based on structure–texture features. *Signal Processing: Image Communication*, 117:117025, 2023. 23

[40] Yutao LIU, Baochao ZHANG, Runze HU, Ke GU, Guangtao ZHAI, and Junyu DONG. Underwater image quality assessment: Benchmark database and objective method. *IEEE Transactions on Multimedia*, 26:7734–7747, 2024. 22, 23, 42

[41] Jim HOLDSWORTH and Marcos SCAPICCHIO. O que é deep learning?, jun 2024. Acessado em 02 de dezembro de 2024. 22

[42] A. Murat TEKALP. Deep learning for image/video restoration and super-resolution. *Foundations and Trends® in Computer Graphics and Vision*, 13(1):1–110, 2022. 22, 23

[43] Domonkos VARGA. No-reference image quality assessment with convolutional neural networks and decision fusion. *Applied Sciences*, 12(1), 2022. 22, 42

[44] Licheng Jiao and Jin Zhao. A survey on the new generation of deep learning in image processing. *IEEE Access*, 7:172231–172263, 2019. 23

[45] Abdulhamit Subasi. Chapter 3 - machine learning techniques. In *Practical Machine Learning for Data Analysis Using Python*, pages 91–202. Academic Press, 2020. 23

[46] Pavan C. MADHUSUDANA, Neil BIRKBECK, Yilin WANG, Balu ADSUMILLI, and Alan C. BOVIK. Image quality assessment using contrastive learning. *IEEE Transactions on Image Processing*, 31:4149–4161, 2022. 23, 42

[47] Jongyoo Kim, Anh-Duc Nguyen, and Sanghoon Lee. Deep cnn-based blind image quality predictor. *IEEE Transactions on Neural Networks and Learning Systems*, 30(1):11–24, 2019.

[48] Manni LIU, Jiabin HUANG, Delu ZENG, Xinghao DING, and John PAISLEY. A multiscale approach to deep blind image quality assessment. *IEEE Transactions on Image Processing*, 32:1656–1667, 2023.

[49] Zihan Zhou, Jing Li, Dexiang Zhong, Yong Xu, and Patrick Le Callet. Deep blind image quality assessment using dynamic neural model with dual-order statistics. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7):6279–6290, 2024.

[50] Xingran Liao, Xuekai Wei, Mingliang Zhou, Zhengguo Li, and Sam Kwong. Image quality assessment: Measuring perceptual degradation via distribution measures in deep feature spaces. *IEEE Transactions on Image Processing*, 33:4044–4059, 2024.

[51] Zihan Zhou, Yong Xu, Yuhui Quan, Yun Liang, Jing Li, and Patrick Le Callet. Enhancing cnn-based blind image quality assessment via deep cross-layer pattern encoding. *IEEE Transactions on Multimedia*, 27:9712–9722, 2025. 23, 42

[52] Bo YAN, Bahetiyaer BARE, Chenxi MA, Ke LI, and Weimin TAN. Deep objective quality assessment driven single image super-resolution. *IEEE Transactions on Multimedia*, 21(11):2957–2971, 2019. 23

[53] Chao MA, Chih-Yuan YANG, Xiaokang YANG, and Ming-Hsuan YANG. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017.

[54] Mobeen Ur REHMAN, Imran Fareed NIZAMI, Muhammad MAJID, Farman ULLAH, Irfan HUSSAIN, and Kil To CHONG. Cn-bsriqa: Cascaded network - blind super-resolution image quality assessment. *Alexandria Engineering Journal*, 91:580–591, 2024. 23

[55] Mengyao LI, Yufei LIN, Liquan SHEN, Zheyin WANG, Kun WANG, and Zhengyong WANG. Human perceptual quality driven underwater image enhancement framework. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. 23

[56] Yan LI, Junsheng LIU, and Xiaofei ZHANG. Fw-gan: Underwater image enhancement using generative adversarial network with multi-scale fusion. *Signal Processing: Image Communication*, 109:116855, 2022. 23

[57] K. OHASHI, Y. NAGATANI, M. YOSHIGOE, T. ANDO, M. KAWAMURA, K. MATSUSHITA, and K. OGUCHI. Applicability evaluation of full-reference image quality assessment methods for computed tomography images. *Journal of Digital Imaging*, 36:2623–2634, 2023. 23

[58] Sergey KASTRYULIN, Jamil ZAKIROV, Nicola PEZZOTTI, and Dmitry V. DYLOV. Image quality assessment for magnetic resonance imaging. *IEEE Access*, 11:14154–14168, 2023. 23

[59] Yuzhen NIU, Yini ZHONG, Wenzhong GUO, Yiqing SHI, and Peikun CHEN. 2d and 3d image quality assessment: A survey of metrics and challenges. *IEEE Access*, 7:782–801, 2019. 23

[60] Margaret H. PINSON. Why no reference metrics for image and video quality lack accuracy and reproducibility. *IEEE Transactions on Broadcasting*, 69(1):97–117, 2023. 23

[61] Saeed ANWAR, Salman KHAN, and Nick BARNES. A deep journey into super-resolution: A survey. *ACM Comput. Surv.*, 53(3), may 2020. 23

[62] Yingjie TIAN, Duo SU, Stanislao LAURIA, and Xiaohui LIU. Recent advances on loss functions in deep learning for computer vision. *Neurocomputing*, 497:129–158, 2022. 24, 113

[63] Zhihao WANG, Jian CHEN, and Steven C. H. HOI. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3365–3387, 2021. 113

[64] Garas GENDY, Guanghui HE, and Nabil SABOR. Lightweight image super-resolution based on deep learning: State-of-the-art and future directions. *Information Fusion*, 94:284–310, 2023. 113

[65] Hongying LIU, Zekun LI, Fanhua SHANG, Yuanyuan LIU, Liang WAN, Wei FENG, and Radu TIMOFTE. Arbitrary-scale super-resolution via deep learning: A comprehensive survey. *Information Fusion*, 102:102015, 2024. 113

[66] Nour ABURAED, Mohammed Q. ALKHATIB, Stephen MARSHALL, Jaime ZABALZA, and Hussain Al AHMAD. A comparative study of loss functions for hyperspectral sisr. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 484–487, 2022. 113

[67] Zafran Hussain SHAH, Marcel MÜLLER, Barbara HAMMER, Thomas HUSER, and Wolfram SCHENCK. Impact of different loss functions on denoising of microscopic images. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10, 2022. 113

[68] Boyang WANG and Jafar SANIIE. Massive ultrasonic data compression using wavelet packet transformation optimized by convolutional autoencoders. *IEEE Transactions on Neural Networks and Learning Systems*, 34(3):1395–1405, 2023. 113

[69] Shutong YE, Shengyu ZHAO, Yaocong HU, and Chao XIE. Single-image super-resolution challenges: A brief review. *Electronics*, 12(13), 2023. 113

[70] Sabina UMIRZAKOVA, Shabir AHMAD, Latif U. KHAN, and Taegkeun WHANGBO. Medical image super-resolution for smart healthcare applications: A comprehensive survey. *Information Fusion*, 103:102075, 2024. 24, 113

[71] W. XUE, L. ZHANG, X. MOU, and A. C. BOVIK. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2014. 24, 43, 49, 50, 71, 75, 91, 113

[72] Huizhen JIA, Lu ZHANG, and Tonghan WANG. Contrast and visual saliency similarity-induced index for assessing image quality. *IEEE Access*, 6:65885–65893, 2018. 24, 49, 51, 71, 91, 113

[73] Guangyi YANG, Deshi LI, Fan LU, Yue LIAO, and Wen YANG. RVSIM: a feature similarity method for full-reference image quality assessment. *EURASIP Journal on Image and Video Processing*, 6:1687–5281, 2018. 25, 49, 53, 54, 71, 73, 79, 87, 91, 105, 113

[74] Z. SHI, J. ZHANG, Q. CAO, K. PANG, and T. LUO. Full-reference image quality assessment based on image segmentation with edge feature. *Signal Processing*, 145:99–105, April 2018. 49, 55, 56, 71, 73, 79, 87, 91, 113

[75] W. SUN, Q. LIAO, J. XUE, and F. ZHOU. SPSIM: A superpixel-based similarity index for full-reference image quality assessment. *IEEE Transactions on Image Processing*, 27(9):4232–4244, September 2018. 25, 49, 55, 57, 64, 71, 88, 91, 105, 113

[76] A. MANSOURI and A. MAHMOUDI-AZNAVEH. SSVD: Structural svd-based image quality assessment. *Signal Processing: Image Communication*, 74:54–63, May 2019. 49, 56, 57, 71, 88, 91, 105, 107, 113

[77] C. SHI and Y. LIN. Full reference image quality assessment based on visual salience with color appearance and gradient similarity. *IEEE Access*, 8:97310–97320, 2020. 25, 43, 47, 49, 56, 58, 71, 73, 79, 91, 105, 113

[78] Jiehang DENG, Haomin CHEN, Zhongming YUAN, Guosheng Susheng GU, Shihe XU, Shaowei WENG, and Hao WANG. An enhanced image quality assessment by synergizing superpixels and visual saliency. *Journal of Visual Communication and Image Representation*, 88:103610, 2022. 25, 50, 59, 60, 71, 87, 88, 91, 113

[79] Sergio A. C. BEZERRA, Sérgio A. C. BEZERRA JÚNIOR, José L. de S. PIO, José R. H. CARVALHO, and Keiko V. O. FONSECA. Perceptual error logarithm: An efficient and effective analytical method for full-reference image quality assessment. *IEEE Access*, 13:68587–68606, 2025. 26, 74, 76

[80] Hélio PEDRINI and William R. SCHWARTZ. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. Thomson Learning, São Paulo, 2008. 28, 29, 31, 39

[81] Mark F. BEAR, Barry W. CONNORS, and Michael A. PARADISO. *Neurociências: Desvendando o Sistema Nervoso*. Artmed, Porto Alegre, 2008. 28, 29, 30, 31

[82] Helga KOLB, Eduardo FERNANDEZ, and Ralph NELSON. *The Organization of the Retina and Visual System*. Moderna, University of UTAH, February 2001. 28, 29

[83] Josef BIGUN. *Vision with Direction: A Systematic Introduction to Image Processing and the Geometry of Early Vision*. Springer, Berlin, Heidelberg, 2006. 28, 29

[84] Robert SNOWDEN, Peter THOMPSON, and Tom TROSCIANKO. *Basic Vision: An Introduction to Visual Perception*. Oxford University Press, Oxford, UK, 2nd edition, 2012. 28

[85] Roberto LENT. *Cem bilhões de neurônios: conceitos fundamentais de neurociência*. Atheneu, São Paulo, 2005. 28, 30

[86] M. S. GAZZANIGA and T. F. HEATHERRTON. *Psychological science: mind, brain and behavior*. W. W. Norton & Company, New York, 2003. 30

[87] Peter G. J. BARTEN. *Contrast sensitivity of the human eye and its effects on image quality*. HV Press, Knegsel, 1999. 29, 87

[88] Francisco RAMALHO JÚNIOR, N. G. FERRARO, and P. A. de Toledo SOARES. *Os fundamentos da física*. Moderna, São Paulo, 2003. 29, 30

[89] Rafael C. GONZALEZ and Richard E. WOODS. *Processamento de Imagens Digitais*. Editora Blucher, São Paulo, 2000. 30

[90] Stefan WINKLER. *Digital Video Quality: Vision Models and Metrics*. Wiley, England, 2005. 30, 31, 34

[91] Jae Jeong HWANG, Hong Ren WU, and K. R. RAO. Digital picture compression and coding structure. In H. R. Wu and K. R. Rao, editors, *Digital Video Image Quality and Perceptual Coding*, pages 3–43. CRC Press, USA, 2006. 31

[92] A. A. MICHELSON. *Studies in Optics*. University of Chicago Press, USA, 1927. 31, 87

[93] ITU-T. *ITU-T Recommendation P.930: Principles of a reference impairment system for video*. ITU-T, August 1996. 32, 33, 34

[94] Michael YUEN. Digital video image quality and perceptual coding. In H. R. Wu and K. R. Rao, editors, *Digital Video Image Quality and Perceptual Coding*, page 87–122. CRC Press, USA, 2006. 32, 33, 34

[95] Lawrence HARTE. *Introduction to IP Video: Digitalization, Compression and Transmission*. Althos, USA, 2007. 32

[96] ITU-T. H.261: Video codec for audiovisual services at p x 384 kbit/s - recommendation h.261 (11/88). Recommendation, International Telecommunication Union - Telecommunication Standardization Sector, 1988. Retrieved 2010-10-21. 32

[97] Iain E. G. RICHARDSON. *H.264 and MPEG-4 Compression*. Wiley, England, 2003. 32, 39, 45

[98] Y. YANG, N. P. GALATSANOS, and A. K. KATSAGGELOS. Recovery methods for postprocessing of compressed images. In Al Bovik, editor, *Handbook of image & Video Processing*, pages 761–774. Elsevier, USA, 2 edition, 2005. 32

[99] Geoff DOUGHERTY. *Digital Image Processing for Medical Applications*. Cambridge University Press, Cambridge, may 2009. 34

[100] Keith JACK. *Video Demystified: A Handbook for the Digital Engineer*. Newnes (Elsevier), Burlington, 4 edition, 2005. 34

[101] Chris SOLOMON and Toby BRECKON. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley-Blackwell, 2011. 34

[102] ITU-T. *ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications*. ITU-T, 2008. 35, 36, 37, 38

[103] ITU-T. *ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications*. ITU-T, 2023. 35

[104] ITU-R. *ITU-R Recommendation BT.500-11: Methodology for the subjective assessment of the quality of television pictures*. ITU-R, June 2002. 35, 36, 37, 38

[105] ITU-R. *Methodologies for the subjective assessment of the quality of television images*. ITU-R, May 2023. 35, 36

[106] Hans HOFFMANN, Takebumi ITAGAKI, David WOOD, Tobias HINZ, and Thomas WIEGAND. A novel method for subjective picture quality assessment and further studies of hdtv formats. *IEEE Transactions on Broadcasting*, 54(1):1–13, March 2008. 35, 36, 38

[107] Pedro CUENCA, Luis O. BARBOSA, and F. GARRIDO, A.; QUILES. Study of video quality metrics for mpeg-2 based video communications. page 280–283. Communications, Computers and Signal Processing, 1999 IEEE Pacific Rim Conference on, 1999. 35, 36

[108] Elisa D. GELASCA. *Full-Reference Objective Quality Metrics for Video Watermarking, Vídeo Segmentation and 3D Model Watermarking*. Università di Trieste, Lausanne, 2005. 38

[109] Zhaoqing PAN, Feng YUAN, Xu WANG, Long XU, Xiao SHAO, and Sam KWONG. No-reference image quality assessment via multibranch convolutional neural networks. *IEEE Transactions on Artificial Intelligence*, 4(1):148–160, 2023. 42

[110] Anmin LIU, Weisi LIN, and Manish NARWARIA. Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, 21(4):1500–1512, 2012. 43, 49, 50, 75

[111] Zhou WANG and Qiang LI. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198, 2011. 44, 47, 75

[112] Songnan LI, Fan ZHANG, Lin MA, and King Ngi NGAN. Image quality assessment by separately evaluating detail losses and additive impairments. *IEEE Transactions on Multimedia*, 13(5):935–949, 2011.

[113] Hua-Wen CHANG, Hua YANG, Yong GAN, and Ming-Hui WANG. Sparse feature fidelity for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 22(10):4007–4018, 2013. 44, 74, 75

[114] Mohammad Hossein KHOSRAVI and Hamid HASSANPOUR. Image quality assessment using a novel region smoothness measure. *Journal of Visual Communication and Image Representation*, 60:217–228, 2019. 44, 47

[115] Keyan DING, Rijin ZHONG, Zhihua WANG, Yang YU, and Yuming FANG. Adaptive structure and texture similarity metric for image quality assessment and optimization. *IEEE Transactions on Multimedia*, 26:5398–5409, 2024.

[116] Shujun LANG, Xu LIU, Mingliang ZHOU, Jun LUO, Huayan PU, Xu ZHUANG, Jason WANG, Xuekai WEI, Taiping ZHANG, Yong FENG, and Zhaowei SHANG. A full-reference image quality assessment method via deep meta-learning and conformer. *IEEE Transactions on Broadcasting*, 70(1):316–324, 2024. 43

[117] M. G. KENDALL. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 06 1938. 44

[118] Mo SHEN, Rongrong SUN, and Wen YE. Sparse representation-based ldct image quality assessment using the jnd model. *IEEE Access*, 13:10422–10431, 2025. 47

[119] Sanaz NAMI, Farhad PAKDAMAN, Mahmoud Reza HASHEMI, Shervin SHIRMOHAMMADI, and Moncef GABBOUJ. Jntd: Toward just noticeable frame rate-based temporal difference for perceptual video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(3):3422–3437, 2026. 47

[120] L. ZHANG, Y. SHEN, and LI=H. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014. 47, 49, 50, 52, 71, 87, 91, 113

[121] Yuuki SAWABE, Satoshi IKEHATA, and Kiyoharu AIZAWA. Saliency-based multiple region of interest detection from a single 360° image. *IEEE Access*, 10:89124–89133, 2022. 47

[122] Ulrich ENGELKE, Anthony MAEDER, and Hans-Jurgen ZEPERNICK. Visual attention modelling for subjective image quality databases. In *2009 IEEE International Workshop on Multimedia Signal Processing*, pages 1–6, 2009. 47, 81, 82, 83

[123] Patrick LE CALLET and Florent AUTRUSSEAU. Subjective quality assessment IRCCyN/IVC database, 2005. Subjective database. 47, 64, 65, 66, 91, 113

[124] Y. HORITA, K. SHIBATA, Y. KAWAYOKE, and Z. M. P. SAZZAD. MICT image quality evaluation database, 2011. http://mict.eng.u-toyama.ac.jp/mictdb.html. 47, 64, 68, 91, 113

[125] Ke GU, Leida LI, Hong LU, Xiongkuo MIN, and Weisi LIN. A fast reliable image quality predictor by fusing micro- and macro-structures. *IEEE Transactions on Industrial Electronics*, 64(5):3903–3912, 2017. 49, 51, 52, 71, 91, 113

[126] S. BAE and M. KIM. A novel image quality assessment with globally and locally consilient visual quality perception. *IEEE Transactions on Image Processing*, 25(5):2392–2406, 2016. 49, 52, 53, 64, 71, 87, 91, 105, 113

[127] A. AHAR, A. BARRI, and P. SCHELKENS. From sparse coding significance to perceptual quality: A new approach for image quality assessment. *IEEE Transactions on Image Processing*, 27(2):879–893, 2018. 49, 54, 71, 91, 113

[128] Rafael REISENHOFER, Sebastian BOSSE, Gitta TAKAMIYA, Thomas WIEGAND, and Gitta KUTYNIOK. A haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, 61:33–43, 2018. 49, 53, 71, 91, 113

[129] Junfeng YANG, Wei ZHANG, Xiaolong LI, Ting ZHOU, and Bo OU. Full reference image quality assessment by considering intra-block structure and inter-block texture. *IEEE Access*, 8:179702–179715, 2020. 49, 58, 59, 71, 91, 113

[130] Mohammad ESMAEILPOUR, Azadeh MANSOURI, and Ahmad MAHMOUDI-AZNAVEH. A new svd-based image quality assessment. In *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, pages 370–374, 2013. 56

[131] Shuigen WANG, Chenwei DENG, Weisi LIN, Baojun ZHAO, and Jie CHEN. A novel svd-based image quality assessment metric. In *2013 IEEE International Conference on Image Processing*, pages 423–426, 2013. 56

[132] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik. Objective quality assessment of multiply distorted images. In *Proc. Conf. Rec. 46th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, pages 1693–1697, Nov. 2012. 59

[133] Dogancan TEMEL and Ghassan ALREGIB. Perceptual image quality assessment through spectral analysis of error representations. *Signal Processing: Image Communication*, 70:37–46, 2019. 61, 71, 91, 113

[134] E. C. LARSON and D. M. CHANDLER. Most apparent distortion:full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006.1–011006.21, January 2010. 64, 67, 91, 113

[135] A. ZARIC, N. TATALOVIC, N. BRAJKOVIC, H. HLEVNJAK, M. LONCARIC, E. DUMIC, and S. GRGIC. VCL@FER image quality assessment database. In *AUTOMATIKA*, volume 53, page 344–354, 2012. 64, 68, 69, 91, 113

[136] N. PONOMARENKO, O. IEREMEIEV, V. LUKIN, K. EGIAZARIAN, L. JIN, J. ASTOLA, B. VOZEL, K. CHEHDI, M. CARLI, F. BATTISTI, and C. J. KUO. Color image database tid2013: Peculiarities and preliminary results. In *European Workshop on Visual Information Processing (EUVIP)*, pages 106–111, June 2013. 64, 69, 91, 109, 113

[137] Xinwei LIU, Marius PEDERSEN, and Jon Yngve HARDEBERG. CID:IQ - a new image quality database. In *International Conference on Image and Signal Processing (ICISP 2014)*, volume 8509, Jul. 2014. http://www.colourlab.no/cid. 64, 70, 91, 113

[138] Hanhe LIN, Vlad HOSU, and Dietmar SAUPE. KADID-10K: A large-scale artificially distorted iqa database. In *2019 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2019. 64, 70, 71, 91, 113

[139] Michaela SPITZER, Jan WILDENHAIN, Juergen RAPPELSBERGER, and Mike TYERS. Boxplotr: A web tool for generation of box plots. *Nature Methods*, 11(2):121–122, Feb. 2014. 74

[140] Kristin C. POTTER. Methods for presenting statistical information: The box plot. In *Visualization of Large and Unstructured Data Sets*, 2006. 74

[141] Chinh Thi My CHU, Hans-Jürgen ZEPERNICK, and Veronica SUNDSTEDT. Analysis of variance of opinion scores for mpeg-4 scalable and advanced video coding. In *2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pages 1–10, 2018.

[142] Xiangxu YU, Christos G. BAMPIS, Praful GUPTA, and Alan Conrad BOVIK. Predicting the quality of images compressed after distortion in two steps. *IEEE Transactions on Image Processing*, 28(12):5757–5770, 2019.

[143] Kathryn A. SZABAT, David M. LEVINE, and David F. STEPHAN. *Statistics for Managers Using Microsoft Excel*. Pearson, USA, 2020. 74

[144] Zhou WANG. SSIM index for image quality assessment, 2003. http://live.ece.utexas.edu/research/quality/ssim/. 78

[145] J.-M. GEUSEBROEK, R. VAN DEN BOOMGAARD, A. W. M. SMEULDERS, and H. GEERTS. Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1338–1350, 2001. 78

[146] Ying TONG and Jin CHEN. Multi-focus image fusion algorithm in sensor networks. *IEEE Access*, 6:46794–46800, 2018. 79

[147] Linru ZHENG, Liquan SHEN, Jianan CHEN, Ping AN, and Jun LUO. No-reference quality assessment for screen content images based on hybrid region features fusion. *IEEE Transactions on Multimedia*, 21(8):2057–2070, 2019. 79

[148] K. L. CHAN and K. MCCARTY. Aspects of the statistical texture analysis of medical ultrasound images. In *IEE Colloquium on Ultrasound Instrumentation*, pages 3/1–3/3, 1990. 79

[149] Yuanyuan KANG, Bin LI, Lianfang TIAN, and Zongyuan MA. Multi-modal medical image fusion based on wavelet transform and texture measure. In *2007 Chinese Control Conference*, pages 697–700, 2007. 79

[150] Pengwei SONG, Hongyu SI, Hua ZHOU, Rui YUAN, Enqing CHEN, and Zhendong ZHANG. Feature extraction and target recognition of moving image sequences. *IEEE Access*, 8:147148–147161, 2020. 79

[151] A. GORE, H. KANSAL, and S. GUPTA. Local standard deviation based image quality metrics for JPEG compressed images. *Indones. J. Electr. Eng. Comput. Sci.*, 12(10):7280–7286, 2014. 79

[152] Yi ZHANG, Thien D. PHAN, and Damon M. CHANDLER. Reduced-reference image quality assessment based on distortion families of local perceived sharpness. *Signal Processing: Image Communication*, 55:130–145, 2017. 79

[153] Ramesh JAIN, Rangachar KASTURI, and Brian G. SCHUNCK. *Machine vision*. McGraw-Hill, USA, 1995. 79, 83

[154] Phillip A. MLSNA and Jeffrey J. RODRÍGUEZ. Gradient and laplacian edge detection. In Al Bovik, editor, *Handbook of image and video processing*, page 535–553. Elsevier Academic Press, USA, 2005. 79

[155] Petros MARAGOS. Morphological filtering for image enhancement and feature detection. In Al Bovik, editor, *Handbook of image and video processing*, pages 135–156. Elsevier Academic Press, USA, 2005. 83

[156] Anne HUMEAU-HEURTIER. Texture feature extraction methods: A survey. *IEEE Access*, 7:8975–9000, 2019. 84

[157] H. SUN, X. CHEN, and T. CHIANG. *Digital video transcoding for transmission and storage*. CRC Press, USA, 2005. 87

[158] Maame G. ASANTE-MENSAH, Anh Huy PHAN, Salman AHMADI-ASL, Zaher AL AGHBARI, and Andrzej CICHOCKI. Image reconstruction using superpixel clustering and tensor completion. *Signal Processing*, 212:109158, 2023. 88

[159] Xiaofeng REN and Jitendra MALIK. Learning a classification model for segmentation. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 10–17 vol.1, 2003. 88

[160] Subhashree SUBUDHI, Ram Narayan PATRO, Pradyut Kumar BISWAL, and Fabio DELL'ACQUA. A survey on superpixel segmentation as a preprocessing step in hyperspectral image analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5015–5035, 2021. 88

[161] Junjie YAN, Yinan YU, Xiangyu ZHU, Zhen LEI, and Stan Z. LI. Object detection by labeling superpixels. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5107–5116, 2015. 88

[162] Sinong QUAN, Deliang XIANG, Wei WANG, Boli XIONG, and Gangyao KUANG. Scattering feature-driven superpixel segmentation for polarimetric sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:2173–2183, 2021. 88

[163] Junjun YIN, Tao WANG, Yanlei DU, Xiyun LIU, Liangjiang ZHOU, and Jian YANG. Slic superpixel segmentation for polarimetric sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022. 88

[164] Elineide S. dos SANTOS, Rodrigo de M. S. VERAS, Kelson R. T. AIRES, Helano M. B. F. PORTELA, Geraldo BRAZ JUNIOR, Justino D. SANTOS, and João Manuel R. S. TAVARES. Semi-automatic segmentation of skin lesions based on superpixels and hybrid texture information. *Medical Image Analysis*, 77:102363, 2022. 88

[165] Jiaxin YU, Florian WELLMANN, Simon VIRGO, Marven VON DOMARUS, Mingze JIANG, Joyce SCHMATZ, and Bastian LEIBE. Superpixel segmentations for thin sections: Evaluation of methods to enable the generation of machine learning training data sets. *Computers Geosciences*, 170:105232, 2023. 88

[166] Aiping ZHANG, Wenqi REN, Yi LIU, and Xiaochun CAO. Lightweight image super-resolution with superpixel token interaction. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12682–12691, 2023. 88

[167] Yuming FANG, Yan ZENG, Wenhui JIANG, Hanwei ZHU, and Jiebin YAN. Superpixel-based quality assessment of multi-exposure image fusion for both static and dynamic scenes. *IEEE Transactions on Image Processing*, 30:2526–2537, 2021. 88

[168] Wenjin HU, Yuqi YE, Jiahao MENG, and Fuliang ZENG. No reference quality assessment for thangka color image based on superpixel. *Journal of Visual Communication and Image Representation*, 59:407–414, 2019. 88

[169] Zhi ZHENG, Yun LIU, Yun LIU, Baoqing HUANG, and Hongwei YU. No-reference stereoscopic images quality assessment method based on monocular superpixel visual features and binocular visual features. *Journal of Visual Communication and Image Representation*, 71:102848, 2020. 88

[170] Gustav Theodor FECHNER. *Elements of Psychophysics*, volume 1. Holt, Rinehart and Winston, New York, 1966. Original work published 1860. 90

[171] International Telecommunication Union. Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models. Technical report, ITU-T Recommendation P.1401, Geneva, Switzerland, July 2012. 90

[172] MATLAB. *version 8.5.0 (R2015a)*. The MathWorks Inc., Natick, Massachusetts, 2015. 92

[173] John WATKINSON. *The Art of Digital Video*. Focal Press, Oxford, 1990. 92

[174] G. KELLOG and C. WAGNER. Effects of update and refresh rates on flight simulation visual displays. Technical Report 100415, NASA Langley Research Center, February 1988.

[175] A. HANSEN, W. SMITH, and R. RYBACKI. Real-time synthetic vision cockpit display for general aviation. In *Proceedings of SPIE 3691, Enhanced and Synthetic Vision*, pages 231–240, April 1999.

[176] Jon C. LEACHTENAUER. *Electronic Image Display: Quality and Prediction of Receiver Performance*. SPIE Press, Bellingham, Washington, 2004. 92

[177] G. HINES, Z. RAHMAN, D. JOBSON, and G. WOODELL. Dsp implementation of the retinex image enhancement algorithm. In *Visual Information Processing XIII, Proceedings of SPIE*, volume 5438, pages 13–24. SPIE, 2004. 92

[178] T. DIAS, S. LOPEZ, N. ROMA, and L. SOUSA. A flexible architecture for the computation of direct and inverse transforms in h.264/avc video codecs. *IEEE Transactions on Consumer Electronics*, 57(2):936–944, 2011. 92

[179] Ze SONG, Xudong KANG, Xiaohui WEI, Shutao LI, and Haibo LIU. Unified and real-time image geo-localization via fine-grained overlap estimation. *IEEE Transactions on Image Processing*, 33:5060–5072, 2024.

[180] Ruoheng YAO, Shengming ZHOU, Zhiyue GAO, Yangyi ZHANG, Yiwei LUO, Lei CHEN, and Fengwei AN. Live demonstration: A video denoising co-processor with non-local means algorithm for fhd 30fps image sensor. In *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–1, 2024.

[181] Nils KUCZA, Florian PORRMANN, Christian STOLLENWERK, and Jens HAGEMEYER. Efficient edge ai for next generation smart mirror applications. *IEEE Access*, 13:93755–93774, 2025.

[182] Dafei QIN, Hongyang LIN, Qixuan ZHANG, Kaichun QIAO, Longwen ZHANG, Jun SAITO, Zijun ZHAO, Jingyi YU, Lan XU, and Taku KOMURA. Instant gaussian splatting generation for high-quality and real-time facial asset rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–15, 2025.

[183] Haoyou JIANG, Tao LI, Pingdan XIAO, Sichun DU, and Qinghui HONG. Edcssm: Edge detection with convolutional state space model. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 73(1):384–396, 2026. 92

[184] S. A. C. BEZERRA and A. A. P. POHL. Foveated model based on the action potential of ganglion cells to improve objective image quality metrics. In *12th International Conference on Signal Processing and Multimedia Applications*, page 84–91, Colmar, France, 2015. INSTICC, SciTePress. 116