UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E INFORMÁTICA INDUSTRIAL

WYLLIAN BEZERRA DA SILVA

MÉTODOS SEM REFERÊNCIA BASEADOS EM CARACTERÍSTICAS ESPAÇO-TEMPORAIS PARA AVALIAÇÃO OBJETIVA DE QUALIDADE DE VÍDEO DIGITAL

TESE

CURITIBA
2013

WYLLIAN BEZERRA DA SILVA

MÉTODOS SEM REFERÊNCIA BASEADOS EM CARACTERÍSTICAS ESPAÇO-TEMPORAIS PARA AVALIAÇÃO OBJETIVA DE QUALIDADE DE VÍDEO DIGITAL

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do título de "Doutor em Ciências" – Área de Concentração: Telecomunicações e Redes.

Orientador: Prof. Dr. Alexandre de Almeida Prado Pohl

CURITIBA
2013

Dados Internacionais de Catalogação na Publicação

S586 Silva, Wyllian Bezerra da
Métodos sem referência baseados em características espaço-temporais para avaliação objetiva de qualidade de vídeo digital / Wyllian Bezerra da Silva. — 2013.
189 f. : il. ; 30 cm

Orientador: Alexandre de Almeida Prado Pohl.
Tese (Doutorado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. Curitiba, 2013.
Bibliografia: f. 151-161.

1. Vídeo digital. 2. Controle de qualidade. 3. Levenberg-Marquardt, Método de. 4. Redes neurais (Computação). 5. Algoritmos. 6. Simulação (computadores). 7. Engenharia elétrica – Teses I. Pohl, Alexandre de Almeida Prado, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. III. Título.

CDD (22. ed.) 621.3

Biblioteca Central da UTFPR, Campus Curitiba

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
Câmpus Curitiba

Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial

Tese de Doutorado Nº. 84

"Métodos Sem Referência Baseados em Características Espaço-Temporais para Avaliação Objetiva de Qualidade de Vídeo Digital"

por

Wyllian Bezerra da Silva

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI, da Universidade Tecnológica Federal do Paraná – UTFPR, às 09h do dia 13 de março de 2013, como requisito parcial à obtenção do título de Doutor em CIÊNCIAS – Área de Concentração: Telecomunicações e Redes. O trabalho foi aprovado pela Banca Examinadora composta pelos doutores:

Alexandre de Almeida Prado Pohl, Dr.
(Presidente - UTFPR)

Marcelo Sampaio de Alencar, Dr.
(UFCG)

Eduardo Antônio Barros da Silva, Dr.
(UFRJ/RJ)

Profª. Lúcia Valéria Ramos de Arruda, Dra.
(UTFPR)

Prof. Hugo Vieira Neto, Dr.
(UTFPR)

Visto da Coordenação:

Prof. Ricardo Lüders, Dr.
(Coordenador do CPGEI)

À minha esposa, Simone, com admiração, amor e gratidão pela compreensão, carinho e apoio incondicional.

AGRADECIMENTOS

Pelo amor, carinho, incentivo e apoio agradeço à minha esposa Simone, à minha mãe Fátima, aos meus irmãos Jean e Rodrigo e aos meus sogros Carlos e Lúcia.

Ao Professor Alexandre Pohl pela amizade e incentivo nos momentos mais difíceis desta tese, além de sua primorosa orientação e contribuições.

Aos Professores Keiko, Richard e Valeria pelas valorosas dicas e contribuições.

À CAPES pelo apoio financeiro.

Por fim, agradeço ao Sheldon, Emmerson, Ricardo, Rodolfo, Luiz, Marcelo, Sérgio, Roberta e Walter Duarte pelo incentivo e amizade. Também agradeço ao corpo docente e funcionários da UTFPR que direta ou indiretamente contribuíram nesta etapa de minha vida.

“In the middle of difficulty lies opportunity.” Albert Einstein

RESUMO

SILVA, Wyllian Bezerra da. MÉTODOS SEM REFERÊNCIA BASEADOS EM CARACTERÍSTICAS ESPAÇO-TEMPORAIS PARA AVALIAÇÃO OBJETIVA DE QUALIDADE DE VÍDEO DIGITAL. 189 f. Tese – Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2013.

O desenvolvimento de métodos sem referência para avaliação de qualidade de vídeo é um assunto incipiente na literatura e desafiador, no sentido de que os resultados obtidos pelo método proposto devem apresentar a melhor correlação possível com a percepção do Sistema Visual Humano. Esta tese apresenta três propostas para avaliação objetiva de qualidade de vídeo sem referência baseadas em características espaço-temporais. A primeira abordagem segue um modelo analítico sigmoidal com solução de mínimos quadrados que usa o método Levenberg-Marquardt e a segunda e terceira abordagens utilizam uma rede neural artificial *Single-Hidden Layer Feedforward Neural Network* com aprendizado baseado no algoritmo *Extreme Learning Machine*. Além disso, foi desenvolvida uma versão estendida desse algoritmo que busca os melhores parâmetros da rede neural artificial de forma iterativa, segundo um simples critério de parada, cujo objetivo é aumentar a correlação entre os escores objetivos e subjetivos. Os resultados experimentais, que usam técnicas de validação cruzada, indicam que os escores dos métodos propostos apresentam alta correlação com as escores do Sistema Visual Humano. Logo, eles são adequados para o monitoramento de qualidade de vídeo em sistemas de radiodifusão e em redes IP, bem como podem ser implementados em dispositivos como decodificadores, *ultrabooks*, *tablets*, *smartphones* e em equipamentos *Wireless Display* (WiDi).

Palavras-chave: *Extreme Learning Machine*, Levenberg-Marquardt, Métricas Objetivas, Qualidade de Vídeo Sem Referência, Redes Neurais

ABSTRACT

SILVA, Wyllian Bezerra da. NO-REFERENCE OBJECTIVE VIDEO QUALITY ASSESSMENT METHOD BASED ON SPATIO-TEMPORAL FEATURES. 189 f. Tese – Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2013.

The development of no-reference video quality assessment methods is an incipient topic in the literature and it is challenging in the sense that the results obtained by the proposed method should provide the best possible correlation with the evaluations of the Human Visual System. This thesis presents three proposals for objective no-reference video quality evaluation based on spatio-temporal features. The first approach uses a sigmoidal analytical model with least-squares solution using the Levenberg-Marquardt method. The second and third approaches use a Single-Hidden Layer Feedforward Neural Network with learning based on the Extreme Learning Machine algorithm. Furthermore, an extended version of Extreme Learning Machine algorithm was developed which looks for the best parameters of the artificial neural network iteratively, according to a simple termination criteria, whose goal is to increase the correlation between the objective and subjective scores. The experimental results using cross-validation techniques indicate that the proposed methods are correlated to the Human Visual System scores. Therefore, they are suitable for the monitoring of video quality in broadcasting systems and over IP networks, and can be implemented in devices such as set-top boxes, ultrabooks, tablets, smartphones and Wireless Display (WiDi) devices.

Keywords: Extreme Learning Machine, Levenberg-Marquardt, Neural Networks, No-Reference Video Quality, Objective Metrics

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

4CIF	*Four times Common Intermediate Format*
AAC	*Advanced Audio Coding*
AC	*Alternating Coefficient*
ACR	*Absolute Category Rating*
ADALINE	*ADAptive LINear Element*
AVC	*Advanced Video Coding*
BFGS	Broyden-Fletcher-Goldfarb-Shanno *method*
B	*Bidirectionally-predictive-coded frames*
BP	*Back-Propagation*
CBP	*Circular Back-Propagation*
cd	candela
CIF	*Common Intermediate Format*
dB	decibel
DC	*Direct Coefficient*
DCR	*Degradation Category Rating*
DCT	*Discrete Cosine Transform*
DFP	Davidon-Fletcher-Powell *method*
DMOS	*Differential Mean Opinion Score*
DMOSn	*normalized DMOS*
DMOSp	*predicted* DMOS
DSCQS	*Double Stimulus Continuous Quality scale*
DSIS	*Double Stimulus Impairment Scale*
ELM	*Extreme Learning Machine*
EPFL	*École Polytechnique Fédérale de Lausanne*
fps	*frames per second*
FR	*Full-Reference*
FSIM	*Feature Similarity index*
GHz	gigahertz
GNU	*GNU is Not Unix*
GoP	*Group of Pictures*
G*byte*	*gigabyte*
HC	*High Complexity*
HD	*High Definition*
HDi	*HD interlaced mode*
HDp	*HD progressive mode*
HRC	*Hypothetical Reference Circuits*
iCDF	*inverse of the Cumulative Distribution Function*
I	*Intra-coded frames*
IP	*Internet Protocol*
IPTV	*Internet Protocol TV*
IQA	*Image Quality Assessment*

IRCCyN	*Institut de Recherche en Communications et Cybernétique de Nantes*
ISPs	*Internet Service Providers*
IST	Instituto Superior Técnico de Lisboa
ITU	*International Telecommunication Union*
ITU-R	*ITU Radiocommunication sector*
ITU-T	*ITU Telecommunication standardization sector*
IVC	*Image and Video-Communication*
IVP	*Image and Video Processing Laboratory*
JPEG-2000	*Joint Photographic Expert Groups-2000*
JPEG	*Joint Photographic Expert Groups*
JPEG-NR	JPEG *No-Reference*
k*bit*/s	*kilobit per second*
k*byte*	*kilobyte*
LC	*Low Complexity*
LD	*Low Definition*
LIVE	*Laboratory for Image & Video Engineering*
LM	Levenberg-Marquardt *method*
MAD	*Mean Absolute Difference*
MADw	*Mean Absolute Difference weight*
MAE	*Mean Absolute Error*
MLP	*Multi-Layer Perceptron*
MOS	*Mean Opinion Score*
MOSp	*predicted* MOS
MOVIE	*MOtion-based Video Integrity Evaluation index*
MPEG-2	*Motion Picture Experts Group-2*
MP	Moore-Penrose *method*
MSE	*Mean Square Error*
ms	milissegundo
MS-SSIM	*Multi-Scale Structural SIMilarity index*
M*bit*/s	*megabit per second*
NLSA	*Nonlinear Least-Squares Approximation*
NR	*No-Reference*
NRVQA-ELM	*No-Reference Video Quality Assessment based on ELM algorithm*
NRVQA-ELMtc	*NRVQA-ELM with termination criteria*
NRVQA	*No-Reference Video Quality Assessment metric*
NRVQA-LM	*No-Reference Video Quality Assessment using LM method*
OR	*Outlier Ratio*
OS-ELM	*Online Sequencial Extreme Learning Machine*
P	*Predictive-coded frames*
PLCC	*Pearson Linear Correlation Coefficient*
PLSR	*Partial Least Squares Regression*
PoliMi	*Politecnico di Milano*
pPSNR	*perceptual PSNR*
PSNR	*Peak Signal-to-Noise Ratio*
PVS	*Processed Video Sequences*

QEs	*Quality Estimators*
QP0	*QP for the base*
QP1	*QP for the enhancement layer*
QP	*Quantization Parameter*
QVGA	*Quarter Video Graphics Array*
RBF	*Radial Basis Function*
RMS	*Root Mean Square*
RNA	Rede Neural Artificial
RoI	*Regions of Interest*
RR	*Reduced-Reference*
RRVQA	*Reduced-Reference Video Quality Assessment*
SD	*Standard Definition*
SDSCE	*Simultaneous Double Stimulus for Continuous Evaluation*
SI	*Spatial perceptual Information*
SLFN	*Single-Hidden Layer Feedforward Neural Network*
SO	Sistema Operacional
SROCC	*Spearman Rank Order Correlation Coefficient*
SSCQE	*Single Stimulus Continuous Quality Evaluation*
SSE	*Sum of Square Errors*
SSIM	*Structural SIMilarity index*
SVC4QoE	*SVC for Quality of Experience*
SVC	*Scalable Video Coding*
SVD	*Singular Value Decomposition*
SVH	Sistema Visual Humano
SVM	*Support Vector Machine*
SVR	*Support Vector Regression*
TI	*Temporal perceptual Information*
TS	*Transport Stream*
TSS	*Total Square Sum*
TUM	*Technische Universität München*
TV	*Television*
UTFPR	Universidade Tecnológica Federal do Paraná
VGA	*Video Graphics Array*
VQAS	*Video Quality Assessment Scores*
VQEG	*Video Quality Experts Group*
VQM	*Video Quality Metric*
WiDi	*Wireless Display*

SUMÁRIO

1 INTRODUÇÃO

O crescimento das telecomunicações e o avanço da tecnologia voltada para a produção e distribuição de conteúdo visual têm sido expressivos durante os últimos anos. Neste contexto, merece destaque o vídeo digital, transmitido pela Internet ou via radiodifusão para os mais diversos tipos de dispositivos. A TV Digital começou a ser transmitida no Brasil em 2 de dezembro de 2007, substituindo o sistema existente com a possibilidade de interatividade e oferecendo uma melhor definição de som e vídeo, com os formatos AAC (*Advanced Audio Coding*) e AVC (*Advanced Video Coding*), respectivamente. Contudo, vídeos digitais estão sujeitos a distorções introduzidas na digitalização e na transmissão. Neste cenário, a avaliação objetiva da qualidade de vídeo oferece suporte à produção e distribuição de conteúdo visual, com destaque para as métricas objetivas sem referência (NR – *No-Reference*), sendo um tópico que desperta interesse no campo da avaliação de qualidade de imagem – IQA (*Image Quality Assessment*) (OELBAUM *et al.*, 2009). As métricas NR podem ser aplicadas em sistemas finais, pois não utilizam o vídeo ou imagem original no cálculo do seu escore de qualidade, conforme será discutido no próximo capítulo.

Assim, avaliar a qualidade de vídeos submetidos a condições de ruído ou distorções constitui um desafio no desenvolvimento de métodos para avaliação objetiva sem referência, cujo propósito é estabelecer uma adequada correlação entre as métricas objetivas propostas e a percepção do Sistema Visual Humano (SVH).

Nesta tese são propostos três métodos de avaliação objetiva de qualidade de vídeo sem referência que utilizam seis características espaço-temporais. As características espaciais são compostas por um descritor de artefato de blocagem e dois descritores de artefato de borramento. Além disso, são utilizadas três características temporais aplicadas às sequências de vídeo. O primeiro método proposto apresenta uma modelagem analítica com solução baseada na minimização do erro quadrático pelo método iterativo de Levenberg-Marquardt (LM) com a denominação de NRVQA-LM (*No-Reference Video Quality Assessment using LM method*) e o segundo e terceiro métodos propostos utilizam uma rede neural artificial (RNA), cujo aprendizado é baseado em máquina de aprendizado extremo – ELM (*Extreme Learning Machine*) com

a denominação de NRVQA-ELM (*No-Reference Video Quality Assessment based on ELM algorithm*) e a versão estendida do algoritmo ELM denominada como NRVQA-ELMtc (*NRVQA-ELM with termination criteria*). Em ambas as propostas há um mapeamento entre as entradas (características espaço-temporais) e as saídas desejadas (escores subjetivos).

As seções seguintes apresentam os problemas relacionados à avaliação objetiva de qualidade de vídeo digital, a motivação e objetivos acerca deste estudo, bem como o estado da arte, contribuições e a organização desta tese.

1.1 PROBLEMAS EM AVALIAÇÃO OBJETIVA DE QUALIDADE DE VÍDEO

Diversos são os problemas encontrados no desenvolvimento e validação de métodos para avaliação objetiva de qualidade de vídeo, sobretudo aqueles sem referência. A seguir são enumerados alguns problemas e dificuldades encontradas.

1. Problemas:
   (a) Limitação na compreensão de aspectos cognitivos do SVH (WANG *et al.*, 2003a);
   (b) Métodos sem referência para avaliação de qualidade de vídeo não preveem quaisquer informações acerca da qualidade do vídeo de referência. Logo, eles devem oferecer uma estimativa de qualidade tanto para vídeos distorcidos, quanto para originais (referência);
   (c) Métodos sem referência devem apresentar uma correlação adequada com as medidas das métricas subjetivas, sendo que o ideal é que eles apresentem uma correlação de Pearson (predição da acurácia) e de Spearman (predição da monotonicidade) mais próxima possível da unidade (HEMAMI; REIBMAN, 2010; LIN; KUO, 2011).
2. Dificuldades:
   (a) Validação:
      i. Encontrar resultados comparáveis na literatura, tanto na função de mapeamento entre escores objetivos e subjetivos, quanto nas bases de dados de vídeos utilizadas;
      ii. Processo de validação com alto custo computacional no cálculo de métricas de referência completa – FR (*Full-Reference*) em vídeos de alta definição HD (*High Definition*), *e.g.*, a métrica FR MOVIE (*MOtion-based Video Integrity Evaluation index*) opera nos domínios espacial, temporal e espaço-temporal, exigindo alto custo computacional (SESHADRINATHAN; BOVIK, 2010).

(b) Manipulação, armazenamento e processamento de bases de dados de vídeo. O armazenamento representa um dos maiores desafios no desenvolvimento de sistemas de processamento de imagens (MARQUES FILHO; VIEIRA NETO, 1999). Este desafio é intensificado quando aplicado a sequências de vídeo, *e.g.*, um vídeo com duração de dez segundos, resolução HD (1920×1080 *pixels*), contendo 50 quadros por segundo – fps (*frames per second*) e formato de arquivo YUV com subamostragem 422p (4:2:2 no modo progressivo) requer 2,025 G*bytes* de espaço.

(c) Disponibilidade de bases de dados de vídeo que contemplem diversas características, tais como:

   i. Tipos de codificação;
   ii. Taxas de compressão;
   iii. Resoluções (baixa, padrão e alta definição);
   iv. Conteúdo de vídeo (filme, animação, telejornal etc.);
   v. Tipo de movimentação das cenas (câmera em movimento, objetos em movimento, ambos em movimento ou ambos estáticos);
   vi. Velocidade de movimentação das cenas ou objetos (baixa, moderada e alta).

(d) Identificação e extração de características correlacionadas com a percepção do SVH;

(e) Escolha da técnica de mapeamento entre as características espaço-temporais extraídas dos conjuntos de treinamento e os escores subjetivos. Qual modelagem deve ser adotada?

   i. Modelagem analítica (método iterativo ou não iterativo, modelo matemático sigmoidal, exponencial etc.);
   ii. Técnica baseada em inteligência artificial (RNAs, lógica *fuzzy* etc.).

(f) Técnica de mapeamento entre as características espaço-temporais e os escores subjetivos baseada em RNAs:

   i. Qual a arquitetura de RNA deve ser escolhida?
   ii. Deve-se evitar a perda da capacidade de generalização da RNA pela adequada escolha do número de neurônios na camada oculta, de modo a evitar a perda da capacidade de generalização relacionada à memorização dos dados de treinamento ou um subajustamento (*underfitting*) associado à incapacidade de convergência da RNA;
   iii. Escolha de um critério de parada apropriado ao método NR proposto.

1.2 MOTIVAÇÃO

A avaliação objetiva de qualidade de vídeo digital com métricas sem referência representa um desafio, uma vez que esta deve ser realizada sem o auxílio de qualquer informação oriunda da fonte. Este assunto, embora incipiente na literatura, desperta grande interesse tanto industrial quanto acadêmico, pois essas métricas são adequadas para aplicações do mundo real, no qual os usuários tipicamente assistem a conteúdos audiovisuais transmitidos pela Internet ou por radiodifusão (WANG; BOVIK, 2006; LIU; HEYNDERICKX, 2009; LIU *et al.*, 2010; CHOI; LEE, 2011; KEIMEL *et al.*, 2011a, 2011b; LIU *et al.*, 2011; LIAO; CHEN, 2011; DOERMANN, 2012).

As métricas FR e RR (*Reduced-Reference*) podem não ser adequadas em transmissões de vídeo em canal ruidoso, no qual há perdas de pacotes, pois nem sempre é possível obter informações acerca do vídeo original. Nestes casos, as métricas NR podem oferecer uma predição de qualidade de vídeos submetidos em condições de distorções (artefatos), *e.g.*, a Figura 1 ilustra esse cenário, em que artefatos de travamento (*frame freezing* ou *jerkiness*) surgem quando ocorrem atrasos ou perdas de pacotes (VENKATARAMAN *et al.*, 2012), como ilustra a comparação entre os quadros originais nas Figuras 1-a e 1-b e os quadros transmitidos, contendo artefatos de compressão e de compressão e travamento, conforme mostram as Figuras 1-c e 1-d, respectivamente. Tipicamente, uma métrica FR ou RR consideraria o quadro 225 da Figura 1-b como original e o quadro 225 da Figura 1-d como distorcido no cálculo de seu escore. Conforme inspeção visual da Figura 1, observa-se um erro, devido à dessincronia entre o quadro original da Figura 1-b e o quadro distorcido pelo travamento da Figura 1-d. Logo, uma métrica NR calcularia o seu escore de qualidade quadro a quadro, sem recorrer ao vídeo original, evitando o erro cometido nas métricas FR e RR.

Métricas NR podem ser aplicadas em sistemas de visão computacional ou em robótica, bem como podem oferecer suporte ao monitoramento da qualidade de sistemas de radiodifusão ou de vídeo sob demanda, *e.g.*, em emissoras de TV e em Provedores de Serviços de Internet (ISPs – *Internet Service Providers*). Dado o requisito de correlação entre a métrica NR e o SVH, ela poderia ser aplicada na correção ou redução de distorções (artefatos) em vídeo digital, tanto antes da transmissão, após algum processo de compressão, quanto na recepção do sinal de vídeo, em que o algoritmo NR, embarcado em *hardware*, identificaria as distorções e, segundo um limiar de qualidade, poderia aplicar correções locais antes de exibir o conteúdo ao usuário.

Assim, esta tese apresenta o desenvolvimento de métodos NR com uma abordagem analítica sigmoidal, *i.e.*, segundo uma modelagem matemática na forma $\frac{1}{1+e^{x}}$, em que x é

(a) (b)

(c) (d)

Figura 1: Exemplo de artefato de travamento (*frame freezing* ou *jerkiness*). Sequência de vídeo original entre (a) 7 s (quadro 175) e (b) 9 s (quadro 225), transmissão da sequência de vídeo contendo artefatos de compressão entre (c) 7 s (quadro 175) e (d) 9 s (quadro 225) com o artefato de travamento.

Fonte: Adaptado de Venkataraman *et al.* (2012).

definido segundo uma equação que incorpora as características espaço-temporais, ponderadas por parâmetros (coeficientes) otimizados durante a fase de treinamento. Além disso, outra abordagem é apresentada com o emprego de uma RNA e o algoritmo de aprendizado ELM.

1.3 OBJETIVOS

A seguir são descritos, de maneira sucinta, os objetivos geral e específicos que nortearam a tese.

1.3.1 OBJETIVO GERAL

Propor e validar métricas sem referência baseadas em modelos analíticos e em RNAs que ofereçam suporte à avaliação objetiva de qualidade de vídeo e que possuam correlação com o SVH.

1.3.2 OBJETIVOS ESPECÍFICOS

1. Caracterizar e manipular os dados de diversas bases de vídeos disponíveis na Internet;
2. Aplicar a extração de características espaço-temporais a essas bases de vídeos;
3. Calcular os escores de qualidade de métricas FR e NR nas bases de dados utilizadas;
4. Formar grupos de treinamento e teste, segundo procedimentos de validação cruzada;
5. Empregar os métodos propostos (NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc) no mapeamento entre as características espaço-temporais e escores subjetivos das bases de dados empregadas;
6. Avaliar o desempenho dos métodos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc segundo as recomendações do VQEG (*Video Quality Experts Group*);
7. Comparar o desempenho do método NRVQA-ELM e NRVQA-ELMtc com as funções de ativação sigmóide, hardlim, seno e RBF (*Radial Basis Function*);
8. Comparar os resultados obtidos pelos métodos propostos com métricas FR e NR disponíveis na literatura.
9. Discutir os resultados obtidos e apontar possíveis extensões.

1.4 ESTADO DA ARTE

O campo de avaliação objetiva de qualidade de vídeo com o emprego de métricas sem referência tem se mostrado bastante ativo durante os últimos anos. Contudo, o desenvolvimento de algoritmos baseados em métricas sem referência enfrenta limitações devido à dificuldade de compreensão do SVH (WANG *et al.*, 2003a). Entretanto, uma variedade de trabalhos tem sido elaborada com o escopo de realizar avaliações objetivas de qualidade visual com métricas sem referência. Há duas abordagens para o desenvolvimento de métricas objetivas sem referência que visam a avaliação da qualidade de vídeo: (i) extração de parâmetros sobre o fluxo de *bits* do

vídeo, tais como taxa de compressão, informações relativas ao GoP (*Group of Pictures*) e ao parâmetro de quantização (QP – *Quantization Parameter*) (SLANINA *et al.*, 2007; SUGIMOTO *et al.*, 2009; OELBAUM *et al.*, 2009; STAELENS *et al.*, 2010; YANG *et al.*, 2010); (ii) extração de características espaço-temporais, a fim de que sejam detectadas distorções ou artefatos nos quadros que impactam a qualidade do vídeo (KAWAYOKE; HORITA, 2008; KEIMEL *et al.*, 2009; YAO *et al.*, 2009; WANG *et al.*, 2009).

A seguir são relacionados alguns trabalhos que tratam de avaliação objetiva de qualidade de imagens com o uso de métodos sem referência.

Wang *et al.* (2002) propuseram uma das primeiras métricas sem referência para avaliação de imagens JPEG (*Joint Photographic Expert Groups*), baseada em um modelo não linear contendo características espaciais relacionadas ao SVH, denominada JPEG-NR (*JPEG No-Reference*). A métrica proposta tem como objetivo identificar, principalmente, artefatos de borramento e blocagem. Os autores desenvolveram três técnicas para detectar artefatos de compressão JPEG: duas técnicas de detecção de artefatos de borramento, denominadas "A" e "Z", e a técnica "B" que detecta artefatos de blocagem. Esse trabalho utilizou um conjunto de imagens comprimidas em JPEG e em JPEG-2000 e ajustou os parâmetros do modelo não linear com o método iterativo de Levenberg-Marquardt com a posterior comparação entre escores objetivos e subjetivos (MOS – *Mean Opinion Score*).

Sazzad *et al.* (2008) desenvolveram uma métrica NR baseada em uma modelagem sigmoidal para avaliação de qualidade de imagens em JPEG-2000. Esta proposta explora características espaciais, tais como a distorção dos *pixels* e a medida de informação de borda, ambas calculadas considerando apenas a componente de luminância. A estimação da distorção de *pixel* usa duas características: o valor médio do desvio-padrão ao longo das bordas dos blocos e a diferença absoluta entre os *pixels* centrais de blocos adjacentes, ambos aplicados em blocos de 5×5 *pixels*. A medida de informação de borda é similar à de detecção de borramento "Z", proposta por Wang *et al.* (2002). A combinação de todas as características espaciais propostas gera descritores dos artefatos de borramento e *ringing*, segundo um modelo matemático sigmoidal contendo nove coeficientes. O treinamento do modelo foi realizado com a divisão das bases de dados de imagens em duas partes distintas; uma destinada ao treinamento e outra ao teste. Ambos os grupos foram gerados aleatoriamente em dois experimentos, no primeiro os autores criaram a sua própria base de imagens e no segundo utilizaram a base de imagens LIVE (*Laboratory for Image & Video Engineering*) (SHEIKH *et al.*, 2003).

Song e Yang (2009) desenvolveram uma métrica sem referência para avaliação de qualidade de imagens com artefatos de blocagem. Ela aplica máscaras baseadas no SVH que con-

sideram o brilho e a complexidade espacial de imagens que apresentam distorções oriundas de artefatos de blocagem. O modelo proposto detecta irregularidades ao longo das bordas de blocos adjacentes com tamanho típico de 8×8 *pixels*. Os autores utilizaram descritores para os artefatos de blocagem com uma metodologia similar ao modelo proposto por Wang *et al.* (2002). Entretanto, com uma abordagem contendo mais manipulações matemáticas, tais como o desvio-padrão entre o brilho de fundo da borda dos blocos e a média do brilho em toda imagem.

A utilização de RNAs no campo de avaliação objetiva de qualidade de imagem é incipiente na literatura. A seguir são apresentados alguns trabalhos que abordam o problema de avaliação objetiva de qualidade de imagem com o uso de métodos sem referência baseados em RNAs.

Gastaldo *et al.* (2005) desenvolveram um algoritmo para avaliação de qualidade de imagens, com uma métrica sem referência, mediante a utilização de uma RNA *perceptron* multicamada (MLP – *Multi-Layer Perceptron*), contendo duas camadas ocultas. O algoritmo de aprendizado empregou o algoritmo de retropropagação circular (CBP – *Circular Back-Propagation*). Os autores empregaram as seguintes técnicas de extração de características como entradas da RNA: distribuição de luminância, orientação espacial e distribuição de energia no domínio da frequência. No entanto, antes de serem extraídas as características da imagem aplicou-se um filtro de realce nela, para que os artefatos ficassem mais evidentes. A validação dos resultados mostrou um bom desempenho da resposta da rede neural, em comparação com a MOS, tanto na fase de treinamento quanto na fase de teste. Contudo, a questão do custo computacional associado ao algoritmo CBP não foi mencionada no trabalho. A literatura (HUANG *et al.*, 2004; HUANG; SIEW, 2004; HUANG *et al.*, 2006; HUANG, 2013) afirma que técnicas que utilizam algoritmos de treinamento tradicionais por meio de métodos iterativos, como o BP ou variantes, geralmente apresentam maior custo computacional na fase de treinamento do que em propostas mais recentes, tais como o ELM.

Babu *et al.* (2007) propuseram uma métrica sem referência para a avaliação de imagens codificadas no formato JPEG baseada em fatores-chave do SVH, tais como amplitude e largura de bordas, atividade e luminosidade do plano de fundo (*background*). Os autores utilizaram uma RNA MLP com três camadas ocultas e conduziram testes subjetivos. Além disso, compararam seus resultados com a métrica sem referência, proposta por Wang *et al.* (2002). O algoritmo de retropropagação (BP – *Back-Propagation*) foi utilizado durante a fase de aprendizado, a fim de determinar os pesos e polarizações da RNA por meio de treinamento supervisionado.

Suresh *et al.* (2009) apresentaram a primeira aplicação do algoritmo de aprendizado

ELM, voltado para avaliação objetiva de qualidade de imagens sem referência. O trabalho propôs o uso do algoritmo ELM para avaliar a qualidade de imagens codificadas em JPEG, cuja percepção da qualidade é conseguida pela extração de características baseadas na sensibilidade do SVH. Estas características foram utilizadas como entradas da RNA, sendo elas a amplitude e comprimento de borda, atividade e luminância de fundo. No treinamento da RNA foram consideradas as relações entre as características presentes no SVH e os escores subjetivos. O diferencial do algoritmo ELM em relação aos métodos tradicionais, dentre os quais, o método do gradiente descendente (WIDROW; HOFF, 1960), está relacionado ao treinamento da RNA. No ELM, os pesos e polarizações da camada oculta são ajustados aleatoriamente e os da camada de saída são determinados analiticamente. Além disso, o algoritmo ELM é voltado para redes com apenas uma camada oculta, o que, à primeira vista, parece ser uma limitação. Entretanto, a literatura (HUANG *et al.*, 2004; HUANG; SIEW, 2004; HUANG *et al.*, 2006) mostra que redes neurais contendo apenas uma única camada oculta podem realizar aproximação de qualquer função contínua, além de suportar tarefas de regressão e classificação. Os autores deste trabalho mostram a eficiência do algoritmo, frente a problemas de avaliação de qualidade visual, pois o algoritmo mostrou-se eficaz na predição de qualidade de imagens em comparação com os escores subjetivos (MOS). Além disso, os resultados obtidos foram comparados com a métrica objetiva de referência completa SSIM (*Structural SIMilarity index*) e com a métrica sem referência, desenvolvida por Wang *et al.* (2002).

Ciancio *et al.* (2011) desenvolveram uma métrica para avaliação objetiva sem referência de imagens com artefatos de borramento, a partir de um grande banco de imagens (SILVA, 2011). As imagens foram degradadas artificialmente com diferentes intensidades de artefatos de borramento gaussiano e de movimento. Estas imagens foram submetidas a um sistema de avaliação objetiva sem referência baseado em uma rede neural *feed-forward*, contendo um conjunto de características obtidas a partir de um banco de imagens degradadas. A camada de saída da RNA classificou as imagens de acordo com critérios semelhantes à avaliação subjetiva em excelente, bom, regular, ruim e péssimo. Os resultados foram divididos em dois grupos, sendo que em cada grupo foram empregadas três categorias de borramento. No primeiro avaliou-se o desempenho da proposta com a simulação dos artefatos de borramento gaussiano, movimento e combinado. Para tanto, os autores utilizaram os coeficientes de correlação de Pearson e de Spearman, considerando como entradas os escores das métricas objetivas e subjetivas. Foram comparadas três configurações de redes neurais propostas com oito métricas objetivas. Na maioria dos resultados, a utilização da RNA apresenta melhor desempenho do que as métricas objetivas, com coeficientes de correlação maiores do que 0,80. De maneira análoga, o segundo grupo foi comparado com artefatos reais de borramento gaussiano, movimento e combinado.

A avaliação objetiva de qualidade sem referência apresenta maior complexidade quando aplicada a sequências de vídeo do que a imagens, devido à variação temporal dos quadros, conteúdo das cenas e à dificuldade de compreensão das características do SVH (WANG *et al.*, 2003a). O desenvolvimento de métodos sem referência para avaliação de qualidade de vídeo também é um assunto incipiente na literatura. A seguir são apresentados alguns trabalhos que exploram características espaço-temporais na estimação de qualidade de vídeo.

Yang *et al.* (2005) desenvolveram uma métrica NR para avaliação de qualidade de vídeo, que considera a dependência temporal entre quadros adjacentes e as características do SVH. Esse trabalho considera distorções espaciais com base na estimação de movimento e na determinação de regiões de translação que contêm alta complexidade espacial, determinada pela consistência dos vetores de movimento de *pixels* adjacentes. Também são consideradas distorções espaciais dos quadros e a influência da atividade temporal do vídeo. O escore de qualidade de vídeo é obtido pela média dos pesos das distorções espaciais que consideram a estimação de movimento e as regiões de translação de alta complexidade espacial, ambos calculados em toda a sequência de vídeo.

Eden (2008) propôs uma outra métrica sem referência para avaliação de qualidade de sequências de vídeo. Essa métrica é baseada no PSNR (*Peak Signal-to-Noise Ratio*) perceptual ou pPSNR que mede a atividade espacial dos quadros. Esta proposta leva em consideração o gradiente da imagem que pode ser interpretado como uma medida de resistência (*strength*) estrutural em um quadro de vídeo. A atividade espacial é determinada pela medida RMS (*Root Mean Square*) do gradiente que indica as características do quadro avaliado. O autor realizou testes subjetivos com sequências de vídeo em HDp (*HD progressive mode*) codificadas em H.264/AVC com taxas de 4, 8, 12 e 16 M*bit*/s. Além disso, os resultados experimentais indicaram um coeficiente de correlação de 0,88 para a métrica proposta (pPSNR) e de 0,72 para o PSNR.

Kawano *et al.* (2010) propuseram uma métrica NR segundo uma modelagem sigmoidal, baseada em características espaço-temporais, que leva em consideração o impacto do processo de compressão que pode produzir artefatos como blocagem e borramento, que degradam a qualidade do vídeo. O modelo proposto utiliza parâmetros de detecção dos artefatos e explora diferenças entre quadros com o parâmetro de informação perceptual temporal – TI (*Temporal perceptual Information*), definido pela ITU (*International Telecommunication Union*), conforme a recomendação ITU-T (*ITU Telecommunication standardization sector*) P.910 (ITU-T P.910, 1999). O método realiza a identificação de bordas com o algoritmo de Canny (CANNY, 1986) e um detector de ruído ao longo das bordas de blocos com tamanho de 7×7 *pixels*. O

parâmetro de detecção de borramento "Z", desenvolvido por Wang *et al.* (2002) também foi incorporado à modelagem sigmoidal proposta. Além da componente de luminância, as componentes de crominância também foram usadas, exceto para as características "TI" e "Z" que consideram somente a componente de luminância. Os coeficientes do modelo sigmoidal, quatro parâmetros não lineares, foram otimizados por meio do método de mínimos quadrados para minimizar a diferença entre os escores subjetivos e escores objetivos deste modelo. Os autores conduziram quatro experimentos de avaliação subjetiva com vídeos em HDi (*HD interlaced mode*) com 1440×1080 *pixels* no modo entrelaçado, codificados em H.264 e MPEG-2 (*Motion Picture Experts Group-2*), com os quais validaram e avaliaram o desempenho do modelo proposto, cujo coeficiente de correlação de Pearson – PLCC (*Pearson Linear Correlation Coefficient*) apresentou valores entre 0,74 e 0,90. Estes resultados foram comparados com a métrica PSNR, que apresentou valores de PLCC entre 0,52 e 0,84.

Além da abordagem espaço-temporal no desenvolvimento de métodos sem referência para avaliação de qualidade de vídeo, também é possível utilizar as informações contidas no fluxo de *bits*, tais como tamanho e tipo de GoP, taxa de compressão, parâmetro de quantização, tipo de *codec*, informações acerca dos estimadores de movimento, perfil e nível de codificação. A seguir são apresentados alguns trabalhos que exploram características extraídas do fluxo de *bits* para obter uma estimação da qualidade do vídeo.

Hemami e Reibman (2010) fizeram uma revisão da teoria relativa à avaliação objetiva sem referência e de estimadores de qualidade (QEs) para imagens e vídeo. Nesse trabalho são descritos três estágios para os estimadores de qualidade em avaliação objetiva sem referência. Os autores elaboraram um levantamento dos métodos que dependem de informações do fluxo de *bits*, das informações espaciais (*pixels*) ou de ambos os métodos. Além de apresentar um *framework* para a estimação da qualidade para a avaliação objetiva sem referência, o trabalho também trata dos artefatos que são incorporados durante a cadeia de processamento, seja durante a aquisição, transmissão ou armazenamento, decodificação e exibição do vídeo; os artefatos podem ser introduzidos em qualquer destes estágios. Com relação ao desenvolvimento de métricas sem referência, os autores sugerem que elas devem incorporar informações sobre a percepção visual humana com inclusão de (i) modelos psicovisuais que possam ser usados para estimar o parâmetro distorção visível ($d_{visible}$); (ii) medidas de sensibilidade a artefatos específicos, que podem ser usados para estimar o parâmetro $d_{visible}$; (iii) preferências conhecidas para "conteúdos perceptualmente agradáveis", relativos à nitidez da cor, contraste, graus de borramento, adição de ruído. Os autores ainda sugerem a inclusão de índices de qualidade subjetiva, associados a um banco de dados de conteúdos para treinamento.

Keimel *et al.* (2011a, 2011b) propuseram uma métrica sem referência, baseada em modelagem sigmoidal para avaliação de qualidade de vídeos codificados em H.264/AVC. As características são extraídas do fluxo de *bits*, entre as quais a taxa de compressão, o parâmetro de quantização, a estimação de movimento e informações relativas ao GoP. A otimização de parâmetros não lineares (coeficientes) é feita com o método de regressão por mínimos quadrados parciais – PLSR (*Partial Least Squares Regression*). Os autores conduziram um experimento com avaliação subjetiva, conforme a recomendação ITU-R (*ITU Radiocommunication sector*) (ITU-R, 2004), contendo apenas quatro vídeos com resolução HD no modo progressivo com taxas variando de 5,4 a 30 M*bit*/s e perfis de codificação LC (*Low Complexity*) e HC (*High Complexity*). Os resultados experimentais da métrica proposta apresentaram PLCC entre 0,91 e 0,94.

Yamagishi *et al.* (2012) propuseram um modelo NR, baseado no método de aproximação por mínimos quadrados não lineares ou NLSA (*Nonlinear Least-Squares Approximation*), para monitorar a qualidade de vídeo H.264/AVC em serviços de IPTV (*Internet Protocol TV*) com a extração de características derivadas do cabeçalho do TS (*Transport Stream*). O modelo proposto gerou um banco de características do fluxo de *bits*, composto pela média de *bits* do quadro do tipo I, número de quadros danificados pela perda de pacotes, estimativa de erro sobre os quadros do GoP (I, B e P), perfil, nível e tipo de codificação, formato de vídeo, tamanho do GoP e a taxa de de quadros por segundo. Os autores conduziram uma avaliação subjetiva com 16 vídeos em HDi $(1440\times1080i)$, codificados em H.264/AVC no modo entrelaçado com o perfil HC e nível 4 com fps $= 30$, GoP na configuração $M = 3$ e $N = 15$ e taxa de compressão entre 2 e 18 M*bit*/s. Os resultados experimentais do modelo proposto indicaram valores de PLCC entre 0,89 e 0,93 na fase de treinamento, enquanto na fase de teste, com padrões não contidos no treinamento, apresentaram valores de PLCC entre 0,90 e 0,94. Aponta-se aqui uma falha na avaliação de desempenho do método proposto quanto à ausência de comparações com outras métricas, sejam elas pertencentes à categoria NR ou à FR.

Métodos para avaliação objetiva de qualidade de vídeo sempre buscam melhorar a correlação entre os escores objetivos e subjetivos. A utilização de RNAs tem o intuito de emular a percepção do SVH frente a padrões de teste identificados pelo treinamento. O uso de RNAs na estimação de qualidade de sequências de vídeo é incipiente e há poucos trabalhos na literatura que tratam desta abordagem. A seguir são apresentados alguns trabalhos que empregam RNAs na estimação de qualidade de sequências de vídeo.

Jiang *et al.* (2008) apresentaram um novo modelo de métrica sem referência para avaliação perceptual de qualidade de vídeos em HD. O modelo é baseado em uma rede neural que

usa o algoritmo de aprendizado BP com a extração de seis características espaço-temporais, baseadas no conceito de regiões de interesse ou RoI (*Regions of Interest*). Para fins de validação, comparou-se o desempenho da rede neural com métricas subjetivas. A questão do custo computacional associado ao tipo de arquitetura de rede empregada não foi discutida.

Choe *et al.* (2009) desenvolveram uma métrica NR com o uso de uma RNA *feed-forward* com três camadas ocultas para avaliação de qualidade de vídeo no formato H.264/AVC. A RNA recebe as características extraídas do fluxo de *bits* do TS, entre as quais a razão entre o número de blocos do tipo I e o número total de quadros, QP, taxa de quadros por segundo, primeiro e segundo momentos de quantização dos coeficientes da transformada DCT (*Discrete Cosine Transform*). Os autores conduziram experimentos com avaliação subjetiva de vídeo no formato CIF (*Common Intermediate Format*), cuja resolução é de 352×288 *pixels* e taxas de compressão variando entre 128 e 720 k*bit*/s, segundo o método ACR (*Absolute Category Rating*) que inclui o vídeo de referência. A fase de treinamento da RNA foi realizada com amostras de vídeos diferentes do padrão de teste e o desempenho desta fase foi comparado com a métrica PSNR, em termos de acurácia, com PLCC variando entre 0,67 e 0,89 para a métrica proposta e PLCC entre 0,21 e 0,84 para o PSNR.

Yao *et al.* (2012) desenvolveram uma métrica sem referência para avaliação de qualidade de vídeo, baseada em estatísticas das características ao longo da trajetória temporal. Os autores utilizaram a base de imagens LIVE (SHEIKH *et al.*, 2003) com imagens distorcidas por (1) compressão JPEG, (2) compressão JPEG-2000, (3) ruído branco, (4) borramento gaussiano e (5) desvanecimento rápido (*fast fading*) *wavelet* para formar um conjunto de padrões de distorções. Assim, as estimativas de distorções foram obtidas com a comparação entre os quadros do vídeo avaliado e o conjunto de padrões de distorções, segundo uma probabilidade p_i $(i = 1,2,3,4,5)$ usada na classificação, conforme um escore de qualidade q_i $(i = 1,2,3,4,5)$ para cada quadro de vídeo, em que i representou as cinco possíveis distorções. O escore de qualidade ponderado (Q_j^w) foi obtido pela soma do produto entre p_i e q_i, para o j-ésimo quadro no domínio *wavelet* Daubechies 9/7 (DAUBECHIES, 1992). Dessa forma, cada quadro distorcido foi classificado neste domínio, considerando três orientações espaciais (horizontal, vertical e diagonal) com três escalas, formando um vetor de características que foi classificado em uma das cinco categorias de distorções, por meio do treinamento multiclasses por SVM (*Support Vector Machine*) com uma RBF. O índice de qualidade espacial (Q_j^w) do quadro foi obtido a partir do vetor de características por SVR (*Support Vector Regression*). A componente temporal da avaliação (Q_j^t) foi obtida por meio da estimação de movimento com os canais de luminância e crominância (YCbCr) e da métrica MS-SSIM (*Multi-Scale Structural SIMilarity index*) em uma janela temporal entre 200 e 350 ms. A estimação de qualidade do vídeo (Q) foi gerada

a partir da soma do produto entre Q_j^w e Q_j^t para $j = 1,\ldots,N$ quadros. Os preditores de qualidade dos vídeos (Q_p) foram obtidos com o mapeamento entre os escores subjetivos da DMOS (*Differential Mean Opinion Score*) e de Q, com a utilização de uma função de mapeamento logística, conforme uma recomendação já ultrapassada do VQEG (VQEG, 2000, 2003), pois as recomendações atuais deste sugerem o uso de uma função de mapeamento cúbica (VQEG, 2008, 2009, 2010). As amostras de vídeo pertencem à base de vídeos LIVE (SESHADRINATHAN *et al.*, 2010) e foram avaliadas pelo método proposto e comparadas com as métricas PSNR e SSIM. Os coeficientes de correlação de Spearman – SROCC (*Spearman Rank Order Correlation Coefficient*) e Pearson (PLCC) foram as medidas de desempenho utilizadas. A métrica proposta apresentou PLCC igual a 0,52, enquanto as métricas PSNR e SSIM apresentaram valores de 0,40 e 0,54, respectivamente.

1.5 CONTRIBUIÇÕES

As principais contribuições desta tese são:

1. Desenvolvimento da métrica objetiva sem referência NRVQA-LM que obedece a uma modelagem sigmoidal, cujos coeficientes da equação são determinados analiticamente pelo método iterativo de Levenberg-Marquardt.

2. Desenvolvimento das métricas objetivas sem referência NRVQA-ELM e NRVQA-ELMtc baseadas em uma RNA que promove o mapeamento entre os escores subjetivos e as características espaço-temporais.

    - NRVQA-ELMtc: método que implementa uma versão estendida do algoritmo ELM que busca iterativamente os melhores parâmetros da RNA, cujo propósito é obter a melhor correlação possível entre a predição objetiva de qualidade e a percepção do SVH.

3. Método de análise dos experimentos com validação cruzada, segundo uma distribuição estatística (*box-plot*) das medidas de desempenho dos métodos propostos, conforme as recomendações do VQEG.

1.6 ORGANIZAÇÃO DA TESE

O restante desta tese está estruturada da seguinte forma:

O Capítulo 2 apresenta a fundamentação teórica relativa à avaliação objetiva e subjetiva de qualidade de vídeo.

O Capítulo 3 descreve as bases de vídeos utilizadas nesta tese e as recomendações do VQEG quanto ao mapeamento entre os escores objetivos e subjetivos. Esse capítulo também apresenta a metodologia para aferir o desempenho dos métodos propostos, bem como as características espaço-temporais e a metodologia acerca dos métodos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc.

O Capítulo 4 apresenta as propostas dos métodos objetivos sem referência para avaliação de qualidade de vídeo; a primeira utiliza uma modelagem sigmoidal, por meio do método iterativo de Levenberg-Marquardt e a segunda e terceira propostas realizam um mapeamento entre as características espaço-temporais e os escores subjetivos com a utilização de uma RNA, cujo aprendizado é baseado no algoritmo ELM.

O Capítulo 5 apresenta a descrição e discussão dos resultados. E, finalmente, o Capítulo 6 apresenta as conclusões desta tese e suas possíveis extensões.

1.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo introduziu o assunto de avaliação objetiva de qualidade de vídeo digital e discutiu os problemas relacionados ao tema, bem como apresentou a motivação, objetivos geral e específico, o estado da arte e as contribuições da tese. O próximo capítulo apresenta uma breve fundamentação teórica que trata das métricas subjetivas e objetivas.

2 AVALIAÇÃO DE QUALIDADE DE VÍDEO

Com o advento das tecnologias de vídeo digital surgiu a necessidade do controle da qualidade do conteúdo visual produzido ou transmitido pelos meios de comunicação. A avaliação de qualidade de vídeo pode ser realizada de forma objetiva ou subjetiva. A primeira faz o uso de algoritmos para estimar a qualidade do conteúdo visual, baseada em modelos estatísticos ou matemáticos, cujo objetivo é obter uma estimativa de qualidade mais próxima possível da percepção visual humana. A segunda estima a qualidade visual com avaliadores humanos e apesar de ser considerada a mais apropriada para avaliar a qualidade de um vídeo, requer muitos recursos humanos, tornando-a impraticável para provedores de conteúdo (WANG *et al.*, 2002, 2003a, 2004a, 2004b; ENGELKE; ZEPERNICK, 2007b).

As seções a seguir discutem as principais características da avaliação subjetiva e objetiva, e as técnicas FR, RR e NR são brevemente apresentadas.

2.1 AVALIAÇÃO SUBJETIVA

O SVH pode adaptar-se a uma larga faixa de níveis de intensidade luminosa com uma ordem de 10^{10} cd e uma relação aproximadamente logarítmica entre estes níveis na cena e a percepção subjetiva de brilho (MARQUES FILHO; VIEIRA NETO, 1999). Schalkoff (1989) enumera três características do processo de percepção do SVH: (i) uma base de dados muito rica, (ii) alta velocidade de processamento e (iii) capacidade de trabalhar sob condições variadas. Assim, a percepção da qualidade do SVH tem papel fundamental na orientação e validação de projetos de sistemas de conteúdo visual (PAPPAS; SAFRANEK, 2000). A métrica de qualidade obtida com o SVH pode ser representada por um escore, denominado média de escore das opiniões (MOS) (ITU-R, 2004). Contudo, o emprego de métricas subjetivas apresenta elevado custo de implementação, devido à utilização de muitos recursos humanos (ESKICIOGLU; FISHER, 1995).

A avaliação subjetiva de qualidade de vídeo emprega uma metodologia para quantificar o desempenho de sistemas de vídeo, mediante a aplicação de medidas que descrevem o grau

de satisfação do usuário frente ao conteúdo visual. A avaliação subjetiva possui uma divisão em duas categorias: a que propõe a avaliar o desempenho de sistemas de vídeo em condições ideais (em termos qualitativos) e aquela que se compromete a manter a qualidade sob condições ótimas, designada como avaliação de imparidade. Aconselha-se que a realização dos testes subjetivos sejam conduzidos segundo as recomendações do ITU (ITU-R, 2004), cujos métodos são resumidos na Tabela 1.

Tabela 1: Métodos para métricas subjetivas

Métodos	Aplicação
DSIS – *Double Stimulus Impairment Scale*: escala de imparidades com duplo estímulo	Medidas de robustez do sistema, ou seja, características de falha
DSCQS – *Double Stimulus Continuous Quality scale*: escala contínua de qualidade com duplo estímulo	Medidas de qualidade de sistemas referenciados para medição da qualidade de codificação de imagens
SSCQE – *Single Stimulus Continuous Quality Evaluation*: avaliação contínua de qualidade com estímulo simples	Medidas de qualidade de sistemas sem referência
SDSCE – *Simultaneous Double Stimulus for Continuous Evaluation*: avaliação contínua com duplo estímulo simultâneo	Medidas de fidelidade entre duas sequências de vídeos previamente avaliados, comparação entre ferramentas distintas de recuperação de erros

Fonte: Norma ITU-R BT-500.11 (ITU-R, 2004).

A MOS é indicada para validação de métricas NR, enquanto a diferença da média de escore das opiniões (DMOS) é uma medida subjetiva indicada para validar métricas FR e RR (VQEG, 2008, 2009, 2010). As recomendações do VQEG definem a DMOS conforme a expressão a seguir.

$$\text{DMOS} = \text{MOS(PVS)} - \text{MOS(ref)} + 5, \tag{1}$$

em que PVS (*Processed Video Sequences*) é a sequência de vídeo processada e MOS(ref) é a MOS do vídeo de referência. Segundo as recomendações do VQEG, quanto maior o valor da DMOS, melhor é a qualidade do vídeo avaliado.

2.2 AVALIAÇÃO OBJETIVA DE QUALIDADE DE VÍDEO

A avaliação objetiva é compreendida como uma modelagem matemática que possibilita avaliar qual o grau de degradação do vídeo, após algum processo de distorção. A degradação do vídeo é perceptível quando surgem artefatos em seu conteúdo, tais como blocagem (*tiling* ou *blocking*), borramento (*blurring*), distorção localizada em áreas pouco nítidas do quadro e travamentos (WU *et al.*, 2007). Dessa forma, estas degradações interferem diretamente na qualidade

do vídeo. A métrica de avaliação de vídeo melhor é aquela que apresenta maior correlação com a métrica subjetiva. As métricas objetivas apresentam vantagens, pois são imparciais, reprodutíveis, confiáveis e apresentam baixo custo. Nos últimos anos, tem se intensificado a procura por métodos objetivos de avaliação de qualidade de imagem e vídeo que apresentem maior correlação com o SVH. Nesse sentido, o VQEG tem mostrado empenho internacional visando a padronização de métricas objetivas para conteúdos visuais (BRUNNSTROM *et al.*, 2009).

As métricas objetivas de avaliação de qualidade de vídeo podem ser subdivididas em três categorias: métricas com referência completa, reduzida e sem referência.

2.2.1 MÉTRICAS DE REFERÊNCIA COMPLETA (FR)

As métricas FR são extensamente utilizadas no processo de validação de propostas RR ou NR, por meio da comparação de desempenho entre os escores objetivos e subjetivos. A seguir são apresentados os métodos FR: PSNR, SSIM e MS-SSIM que são amplamente utilizados na literatura, seja na validação de propostas RR/NR ou na avaliação de qualidade de vídeos processados ou transmitidos (WANG *et al.*, 2003b, 2004a; WU *et al.*, 2007).

A PSNR é uma métrica FR classicamente relatada na literatura, cuja unidade e o dB. Porém, é conhecido que sua medida nem sempre se correlaciona bem com a qualidade de vídeos digitais, pois artefatos, como a blocagem e o borramento, são efeitos de natureza altamente estruturada, devido ao processo de compressão, diferentemente dos artefatos presentes em vídeos analógicos. Dessa forma, o PSNR não reflete adequadamente a percepção do SVH na ocorrência de artefatos ou distorções provenientes do processamento com a transformada, *e.g.*, DCT ou *wavelet* (WU *et al.*, 2007; SALOMON, 2007). Sua medida é processada sobre o erro quadro a quadro, *i.e.*, entre o quadro original x e o quadro degradado y. Logo, a expressão a seguir determina o PSNR para um conjunto de quadros F.

$$\mathrm{PSNR} = \frac{1}{F}\sum_{f=1}^{F} 20\log_{10}\left(\frac{v_f}{\sqrt{\mathrm{MSE}_f}}\right), \tag{2}$$

em que,

$$\mathrm{MSE}_f = \frac{1}{NM}\sum_{i=0}^{N}\sum_{j=0}^{M}\left[x(f,i,j) - y(f,i,j)\right]^2, \tag{3}$$

em que o termo v_f é igual a $(2^k - 1)$, cujo k é o número de *bits* por *pixel* (componente de luminância) do quadro f. Os termos $x(f,i,j)$ e $y(f,i,j)$ são os valores de luminância do quadro de origem e degradado, respectivamente. As componentes bidimensionais são representadas pelo número de colunas N e pelo número de linhas M. O termo MSE_f é definido como erro

quadrático médio ou MSE (*Mean Square Error*) do quadro f que pode assumir valores no intervalo $[0,\infty)$. Assim, a qualidade do vídeo melhora à medida que o MSE se aproxima de zero, por conseguinte, o valor do PSNR tende a aumentar.

Por outro lado, a métrica SSIM é baseada na hipótese de que o SVH é fortemente adaptado para extrair informações acerca das características estruturais de um quadro ou imagem. Assim, uma medida de similaridade estrutural (ou distorção) pode prover boa aproximação para a qualidade perceptual de um quadro de vídeo (WANG; BOVIK, 2002; WANG *et al.*, 2004a; WANG; BOVIK, 2006; SHI *et al.*, 2009). Sejam x e y dois sinais não negativos, em que y é o sinal degradado e x é o sinal original (sem perda de qualidade), a similaridade é a medida quantitativa da qualidade do sinal degradado, ou seja, x é tomado como referência para medir a qualidade de y.

Wang *et al.* (2004a) propuseram a métrica SSIM, conforme ilustra a Figura 2, que descreve a técnica de predição de qualidade que usa a combinação de três componentes, cujo escore de qualidade representa uma medida global de similaridade. As componentes de luminância, contraste e estrutura dos quadros de origem x e distorcido y são comparadas e combinadas para gerar esse escore.

No cálculo da predição de qualidade de vídeo, cuja sequência é representada por f, as componentes de luminância, contraste e estrutura são definidas conforme as Equações (4-6), respectivamente (WANG *et al.*, 2004a).

$$l(f,x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{4}$$

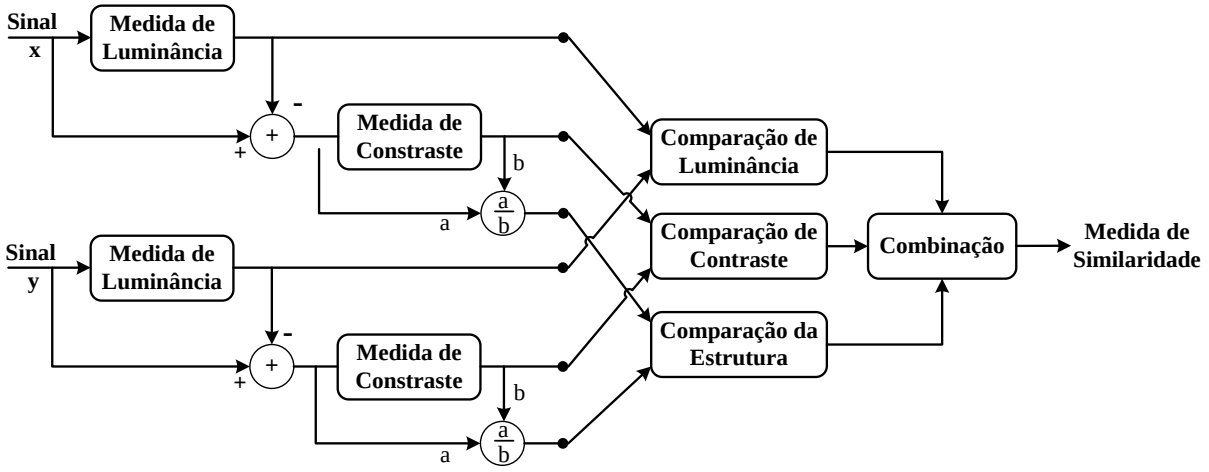

Figura 2: Diagrama do sistema de medida da similaridade estrutural (SSIM).

Fonte: Adaptado de Wang *et al.* (2004a).

$$c(f,x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{5}$$

$$s(f,x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{6}$$

em que μ_x, μ_y, σ_x^2, σ_y^2 e σ_{xy} são a média de x e y, a variância de x e y e a covariância cruzada de x e y, respectivamente. Os termos C_1, C_2 e C_3 são constantes de baixa magnitude. Wang *et al.* (2004a) definem $C_1 = (K_1L)^2$, em que L é um valor dinâmico de *pixel* (255 para quadros em tom cinza de 8 *bits*) e $K_1 \ll 1$; $C_2 = (K_2L)^2$ e $K_2 \ll 1$; $C_3 = \frac{C_2}{2}$.

A forma geral da expressão do índice SSIM relaciona x e y com as componentes de luminância, contraste e estrutura para um conjunto de quadros F, conforme (WANG *et al.*, 2004a).

$$\text{SSIM} = \frac{1}{F}\sum_{f=1}^{F}[l(f,x,y)]^{\alpha}[c(f,x,y)]^{\beta}[s(f,x,y)]^{\gamma}, \tag{7}$$

em que, os parâmetros α, β e γ são usados para ajustar as três componentes do SSIM. Wang *et al.* (2004a) simplificaram a Equação (7), tornando os parâmetros α, β e γ positivos e iguais a 1.

A métrica SSIM apresenta diversas extensões, entre as quais a versão MS-SSIM, proposta por Wang *et al.* (2003b) que assume uma abordagem multiescalar, por um método iterativo (Φ iterações) em que se emprega um filtro passa-baixas e uma subamostragem (*downsampling*) no quadro ou imagem. A expressão a seguir resume o método.

$$\text{MS-SSIM} = \frac{1}{F}\sum_{f=1}^{F}[l_{\Phi}(f,x,y)]^{\alpha_{\Phi}}\prod_{k=1}^{\Phi}[c_k(f,x,y)]^{\beta_k}[s_k(f,x,y)]^{\gamma_k}, \tag{8}$$

em que $\sum_{k=1}^{\Phi}\alpha_k + \beta_k + \gamma_k = 1$, $\Phi = 5$ e $\forall k \in [1,\ldots,\Phi]$; $0 \leq \alpha_k \leq 1$; $0 \leq \beta_k \leq 1$; $0 \leq \gamma_k \leq 1$. Analogamente ao que foi abordado na Equação (7), os expoentes α_{Φ}, β_k e γ_k são utilizados para ajustar as diferentes componentes da expressão (8). A componente de luminância do quadro f, denotada por $[l_{\Phi}(f,x,y)]^{\alpha_{\Phi}}$ é calculada somente na escala Φ.

2.2.2 MÉTRICAS DE REFERÊNCIA REDUZIDA (RR)

Métricas de referência reduzida (RR) extraem um certo número de características do vídeo original, em função do movimento ou de detalhes espaciais. Esse método é utilizado no monitoramento de transmissões em rede. Neste cenário, o vídeo é transmitido com uma sequência de informações codificadas (*overhead*) e no lado receptor ocorre a sua decodificação, seguido pelo cálculo do índice de qualidade, que é obtido pela comparação entre a representação reduzida de informação nos pares fonte e destino (CALLET *et al.*, 2006). Tipicamente, métricas

RR são implementadas por funções, divididas em duas etapas (CALLET; BARBA, 2001):

1. Cálculo do erro entre os vídeos original e degradado, pela diferença de suas características para compor a representação reduzida da informação;
2. Função que agrupa os erros ou diferenças para obter um índice de qualidade global.

Algumas categorias de métricas RR exploram propriedades dos artefatos, pela extração de características e um modelo de parametrização (MIYAHARA *et al.*, 1998), com foco em alguns tipos particulares de artefatos, *e.g.*, blocagem, borramento e *ringing* (KARUNASEKERA; KINGSBURY, 1995; WU; YUEN, 1997). Esta abordagem pode ser encontrada em Silva *et al.* (2013), com a proposta do método RRVQA (*Reduced-Reference Video Quality Assessment*) para avaliação de qualidade de vídeo baseada na diferença de atividade dos coeficientes DCT (RAO; YIP, 1990). As componentes de alta frequência, *i.e.*, os coeficientes AC (*Alternating Coefficient*), são responsáveis pelos detalhes em um quadro e são menos perceptíveis pelo SVH. Por esse motivo, os sistemas de compressão (*e.g.*, MPEG-2 e H.264) reduzem a informação contida nesses coeficientes, tendo como consequência a produção de artefatos típicos do processo de compressão, tais como blocagem e borramento. Assim, o método RRVQA opera sobre a diferença entre os coeficientes AC e DC (*Direct Coefficient*) em um macrobloco com resolução $\tau \times \tau$, em que $\tau = 16$. Os coeficientes em cada macrobloco são representados por coef_p (coeficientes AC e DC) e a média dos coeficientes DC é representada por $\overline{\text{DC}}$. Logo, a diferença absoluta de atividade para um macrobloco j com resolução $\tau \times \tau$ é definida pela expressão a seguir.

$$\text{Actf}_j = \frac{1}{\tau \times \tau} \sum_{p=1}^{\tau \times \tau} \left| \text{coef}_p - \overline{\text{DC}} \right|, \tag{9}$$

em que $\overline{\text{DC}}$ é definido como

$$\overline{\text{DC}} = \frac{64}{\tau \times \tau} \sum_{k=1}^{\frac{\tau \times \tau}{64}} \left| \text{DC}_k \right|. \tag{10}$$

O parâmetro Actf_j é calculado para cada macrobloco dos quadros do vídeo de origem ou de referência e encapsulado em um TS, que é transmitido no canal de comunicação, *e.g.*, TV digital ou rede IP (*Internet Protocol*). Este método gera um *overhead* entre 19 e 21 *bits* em cada macrobloco. No lado receptor, calcula-se a variável Actf_j, conforme Equação (9) e o erro quadrático entre as diferenças absolutas de atividades dos lados emissor (ActfS_j) e receptor (ActfR_j), conforme a expressão a seguir.

$$\text{SEf}_j = \left(\text{ActfS}_j - \text{ActfR}_j \right)^2. \tag{11}$$

Esta abordagem requer o cálculo do MSE no domínio da frequência, *i.e.*, a média da Equação (11) que considera todos os macroblocos (M) de um quadro i.

$$\text{MSEf}_i = \frac{1}{M}\sum_{j=1}^{M}\text{SEf}_j, \tag{12}$$

Logo, o método RRVQA tem uma abordagem similar à métrica PSNR, contudo, no domínio da transformada DCT, cuja definição é dada pela expressão a seguir.

$$\text{RRVQA} = \frac{1}{N}\sum_{i=1}^{N} 10 \times \log_{10}\frac{[\max(cs_i, cr_i)]^2}{\text{MSEf}_i}, \tag{13}$$

em que N é o número de quadros em um vídeo, cs_i e cr_i são os coeficientes DCT do lado emissor e receptor, respectivamente. Somente as variáveis Actf_j e cs_i são encapsuladas no TS.

2.2.3 MÉTRICAS SEM REFERÊNCIA (NR)

Diferentemente das métricas FR ou RR, métodos NR sempre tentarão realizar suposições acerca do conteúdo ou das degradações de um determinado vídeo, baseadas na relação entre suas características e o SVH (WU *et al.*, 2007). Uma métrica objetiva de qualidade de vídeo sem referência ou NRVQA (*No-Reference Video Quality Assessment metric*) realiza medidas diretamente em um vídeo, sem recorrer ao seu conteúdo original. Para tanto, são sugeridos três princípios, cuja combinação é um razoável indicador da qualidade geral de um quadro que se aproxima da avaliação do SVH (WU *et al.*, 2007):

1. Utilizar vídeos que apresentam degradação durante a transmissão, compressão e processamento. Essas situações impõem um efeito monotônico e não-contínuo na qualidade. A melhor qualidade é aquela que esteja livre de artefatos;

2. Empregar vídeos com atributos melhorados, tais como nitidez e contraste, bem como vídeos com artefatos atenuados (menor intensidade de degradação);

3. O vídeo original é isento de degradações, definido como uma referência virtual.

A seguir é apresentada a métrica JPEG-NR proposta por Wang *et al.* (2002) que emprega a detecção de artefatos de blocagem e borramento, cuja componente de luminância é representada por $y(f,i,j)$ com $i \in [1,M]$ e $j \in [1,N]$, em que M é o número de linhas e N é o número de colunas em uma imagem ou quadro f. As diferenças ao longo das direções horizontal

(colunas) e vertical (linhas) são determinadas conforme as fórmulas a seguir.

$$d_h(f,i,j) = y(f,i,j+1) - y(f,i,j), \quad j \in [1, N-1], \tag{14}$$

$$d_v(f,i,j) = y(f,i+1,j) - y(f,i,j), \quad i \in [1, M-1]. \tag{15}$$

O efeito de blocagem pode ser estimado pela média das diferenças entre as bordas dos blocos DCT nas direções horizontal e vertical, conforme as Equações (16-17), com tamanho típico de 8×8 *pixels* no bloco DCT, *i.e.*, $\tau = 8$.

$$B_h = \frac{1}{M\left(\lfloor \frac{N}{\tau} \rfloor - 1\right)} \sum_{i=1}^{M} \sum_{j=1}^{\lfloor \frac{N}{\tau} \rfloor - 1} \left| d_h(f,i,\tau j) \right|, \tag{16}$$

$$B_v = \frac{1}{N\left(\lfloor \frac{M}{\tau} \rfloor - 1\right)} \sum_{i=1}^{\lfloor \frac{M}{\tau} \rfloor - 1} \sum_{j=1}^{N} \left| d_v(f,\tau i,j) \right|. \tag{17}$$

A combinação entre B_h e B_v resulta no descritor de blocagem B_f.

$$B_f = \frac{B_h + B_v}{2}. \tag{18}$$

A medida do artefato de borramento é obtida pela redução de atividade espacial nas direções horizontal e vertical, devido ao processo de quantização no domínio da DCT. As Equações (19-20) expressam as médias das diferenças da redução de atividade espacial nas direções horizontal e vertical, respectivamente.

$$A_h = \frac{1}{\tau - 1} \left[\frac{\tau}{M(N-1)} \sum_{i=1}^{M} \sum_{j=1}^{N-1} \left| d_h(f,i,j) \right| - B_h \right], \tag{19}$$

$$A_v = \frac{1}{\tau - 1} \left[\frac{\tau}{N(M-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N} \left| d_v(f,i,j) \right| - B_v \right]. \tag{20}$$

A combinação entre A_h e A_v gera o descritor A_f para detecção de artefatos de borramento.

$$A_f = \frac{A_h + A_v}{2}. \tag{21}$$

O segundo fator que contribui para a detecção de artefatos de borramento é a taxa de cruzamento por zero ou *zero-crossing* (ZC) nas direções horizontal e vertical.

$$Z_h = \frac{1}{M(N-2)} \sum_{i=1}^{M} \sum_{j=1}^{N-2} z_h(f,i,j), \tag{22}$$

$$Z_v = \frac{1}{N(M-2)} \sum_{i=1}^{M-2} \sum_{j=1}^{N} z_v(f,i,j), \tag{23}$$

em que z_h e z_v são expressos como

$$z_h(f,i,j) = \begin{cases} 1, & \text{se existe } ZC \text{ na direção horizontal} \\ 0, & \text{caso contrário} \end{cases}, \tag{24}$$

$$z_v(f,i,j) = \begin{cases} 1, & \text{se existe } ZC \text{ na direção vertical} \\ 0, & \text{caso contrário} \end{cases}. \tag{25}$$

A combinação entre Z_k e Z_v gera o descritor de borramento Z_f.

$$Z_f = \frac{Z_h + Z_v}{2}. \tag{26}$$

Finalmente, o modelo NR proposto por Wang *et al.* (2002) combina os descritores A_f, B_f e Z_f para um dado número de quadros F, conforme a expressão

$$JPEG\text{-}NR = \frac{1}{F} \sum_{f=1}^{F} \alpha + \beta B_f^{\gamma_1} A_f^{\gamma_2} Z_f^{\gamma_3}, \tag{27}$$

em que Wang *et al.* (2002) obtiveram os parâmetros $\alpha = -245,9$, $\beta = 261,9$, $\gamma_1 = -0,0240$, $\gamma_2 = 0,0160$ e $\gamma_3 = 0,0064$ por meio de uma regressão não linear com o método iterativo de Levenberg-Marquardt (LEVENBERG, 1944; MARQUARDT, 1963; MORÉ, 1977) que realiza um ajuste não linear entre as características espaciais, incorporadas no modelo da Equação (27) e os escores subjetivos (DMOS) das imagens comprimidas em JPEG e JPEG-2000.

2.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo apresentou a fundamentação teórica que trata da avaliação objetiva de qualidade de vídeo com uma breve explanação sobre as métricas subjetivas e objetivas, dentre as quais as técnicas FR, RR e NR. O próximo capítulo apresenta as ferramentas e conceitos que dão suporte aos métodos propostos.

3 METODOLOGIA

Neste capítulo são descritos os materiais e os conceitos que dão suporte aos métodos propostos para avaliação objetiva de qualidade de vídeo sem referência.

Além de gráficos bidimensionais, esta tese também utilizou uma forma de representação que sintetiza estatisticamente o comportamento de uma série de dados com um diagrama de caixa ou *box-plot* (TUKEY, 1977). A Figura 3 apresenta um exemplo com escores de qualidade distribuídos segundo o diagrama de caixa à direita. Os valores extremos ou discrepantes são representados pelo símbolo +, enquanto os valores não discrepantes inferiores e superiores estão confinados entre o símbolo $\perp$ e o primeiro quartil (Q_1) e entre o terceiro quartil (Q_3) e o símbolo $\top$, respectivamente. A mediana (segundo quartil ou $Q2$), média e desvio-padrão são representados pelos símbolos —, ● e ■, respectivamente.

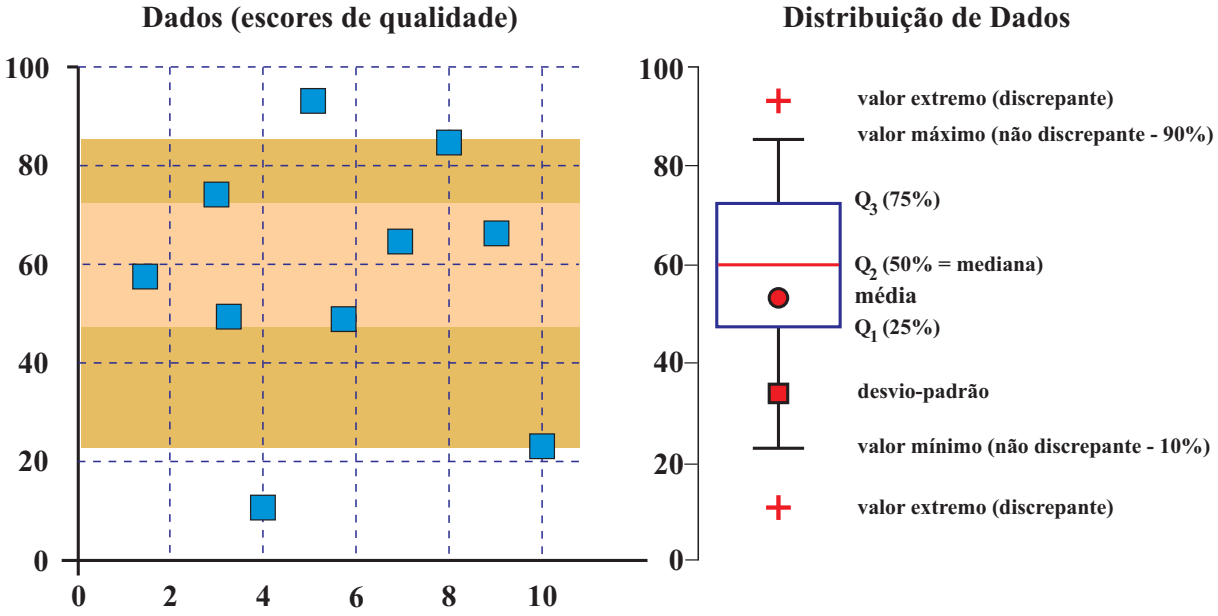

Figura 3: Diagrama de caixa (*box-plot*).

Fonte: Adaptado de http://support2.dundas.com/onlinedocumentation/winchart2005/BoxPlotChart.html.

O desvio-quartil (amplitude interquartílica) e a mediana ($Q2$) são medidas de dispersão que não são influenciadas pelos valores extremos, como ocorre com a média e desvio-padrão. A amplitude interquartílica corresponde à diferença entre o terceiro e o primeiro quartil ($Q3-Q1$) e contém 50% dos dados da distribuição. O primeiro (Q_1), segundo (Q_2) e terceiro $(Q3)$ quartis indicam que os valores da distribuição são menores que 25%, 50% e 75%, respectivamente.

As seções seguintes apresentam as características das bases de vídeos utilizadas nesta tese, o processamento entre os escores objetivos e subjetivos por um mapeamento não linear, bem como as medidas estatísticas de desempenho entre esses escores. Além disso, este capítulo também apresenta as características espaço-temporais e a metodologia dos algoritmos de Levenberg-Marquardt e ELM.

3.1 BASES DE VÍDEOS

Esta tese foi validada com duas bases de dados em escala DMOS e com dezessete bases de dados em escala MOS, que consistem em diversos conteúdos e resoluções. A seguir há uma descrição sucinta dessas bases de dados, acompanhada dos descritores de informação temporal (TI) e espacial (SI – *Spatial perceptual Information*), definidos pela recomendação ITU-T P.910 (ITU-T P.910, 1999), conforme Equações (28) e (29), para uma sequência de quadros F na faixa $[2,F]$ e $[1,F]$, respectivamente.

$$\mathrm{TI}_F = \max_F\{\sigma_s[m(f,i,j)]\}, \tag{28}$$

$$\mathrm{SI} = \max_F\{\sigma_s[\mathrm{Sobel}(y(f,i,j))]\}, \tag{29}$$

em que $m(f,i,j) = y(f,i,j) - y(f-1,i,j)$ é a diferença de movimento (*i.e.*, diferença de luminância) de quadros adjacentes que considera a diferença de movimento dos quadros f e $f-1$, cujos valores de luminância de *pixels* estão localizados em uma mesma região espacial $y(f,i,j)$, com i representando a linha e j a coluna do *pixel* y, $\sigma_s[m(f,i,j)]$ é o desvio-padrão de $m(f,i,j)$ com $f>1$ e $\sigma_s[\mathrm{Sobel}(y(f,i,j))]$ é o desvio-padrão do quadro f processado pelo filtro de Sobel (PARKER, 1997; MARQUES FILHO; VIEIRA NETO, 1999). Os gráficos TI *vs.* SI foram obtidos conforme a recomendação ITU-T P.910 (ITU-T P.910, 1999) que considera os valores de pico $(\max_F)$.

3.1.1 BASE DE DADOS LIVE

A base de dados LIVE (SESHADRINATHAN *et al.*, 2010) inclui 150 amostras de vídeos com dez conteúdos que contemplam distorções por compressão MPEG-2 e H.264, trans-

missão em canal ruidoso sem fio e perdas de pacotes em rede IP. Nesta base de dados há sete sequências de vídeo com 25 fps e as três restantes com 50 fps. Os arquivos não contêm cabeçalho, possuem o formato 4:2:0 progressivo e tamanho de 768×432 *pixels*, cujos escores subjetivos estão em escala DMOS. A Figura 4 descreve o comportamento espacial (Figura 4-a), temporal (Figura 4-b) e temporal *vs.* espacial (Figura 4-c) para os vídeos de referência dessa base de dados. O eixo das abscissas nas Figuras 4-a e 4-b correspondem às amostras de vídeos originais associadas aos pontos da Figura 4-c.

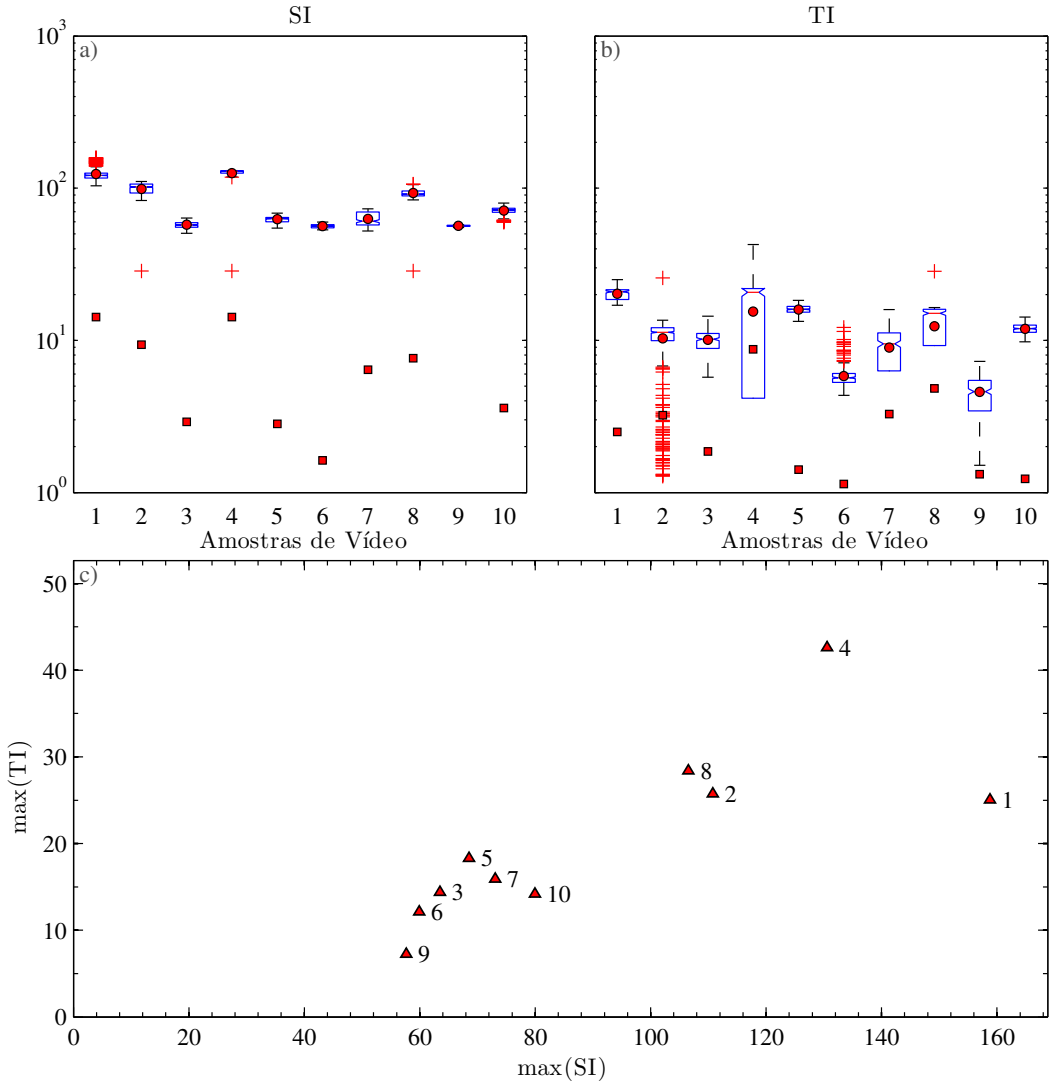

Figura 4: Diagrama TI *vs.* SI da base de dados LIVE com *box-plot* para (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

3.1.2 BASE DE DADOS IVP

A base de dados IVP (*Image and Video Processing Laboratory*) (LI; MA, 2012) é composta por 128 vídeos com dez sequências de referência em alta definição (1920×1088

pixels) no modo progressivo (HDp) com arquivos sem cabeçalho, 25 fps e formato 4:2:0, cujos escores subjetivos estão em DMOS. Além disso, esta base de dados contém quatro tipos de distorções: compressão em Dirac *wavelet*, H.264 e MPEG-2, bem como distorções do tipo perda de pacotes H.264 sob transmissão em rede IP. A Figura 5 expressa o comportamento espacial (Figura 5-a), temporal (Figura 5-b) e temporal *vs.* espacial (Figura 5-c) para os vídeos de referência dessa base de dados.

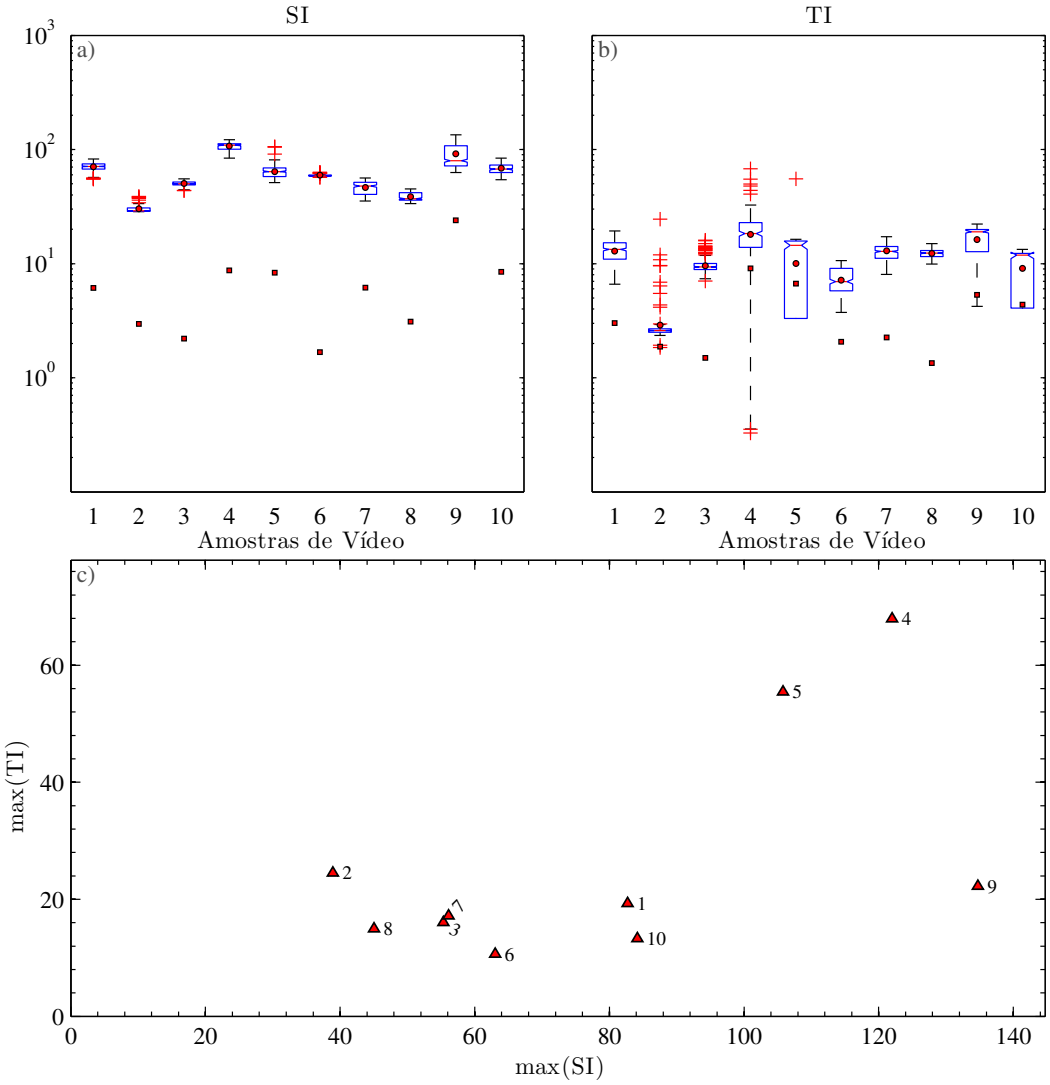

Figura 5: Diagrama TI *vs.* SI da base de dados IVP com *box-plot* para (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

3.1.3 SUPERCONJUNTO *S*

A Figura 6 expressa o comportamento temporal *vs.* espacial das sequências de referência das demais bases de dados utilizadas nesta tese. Denominada superconjunto *S* que comporta 17 bases de dados de vídeo em escala MOS com 2.627 amostras que contemplam diversos

conteúdos e distorções. Agrupando as amostras do superconjunto *S* por resolução, há 204 em LD (*Low Definition*); 1.288 em SD (*Standard Definition*); 520 em HDp e 615 em HDi. Pela inspeção visual da Figura 6 observa-se uma grande variação espaço-temporal dos vídeos do superconjunto *S*, que contém 216 vídeos de referência, cujas características espaciais e temporais tornam-se atrativas do ponto de vista do desenvolvimento de métodos objetivos para avaliação de qualidade de vídeo.

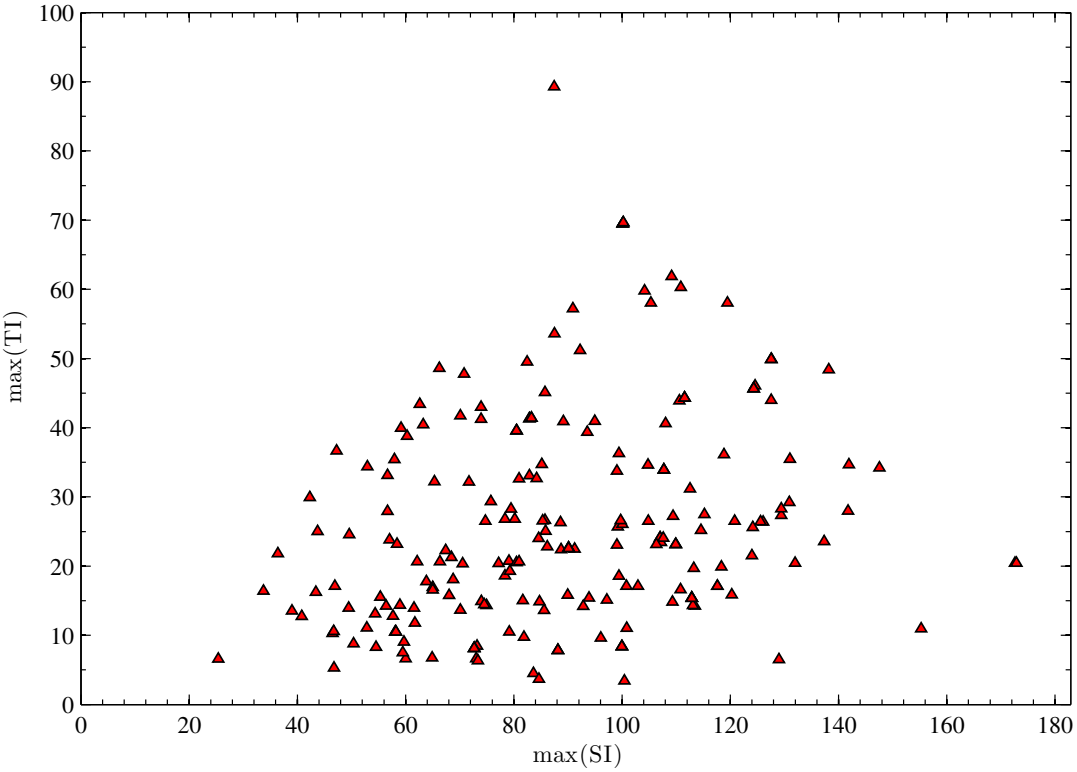

Figura 6: Diagrama max(TI) *vs.* max(SI) do superconjunto *S* com 216 vídeos de referência.

A seguir são descritas as principais características do superconjunto *S* agrupado por resolução. Detalhes sobre o comportamento espacial e temporal dos vídeos de referência de cada base de dados podem ser encontrados no Apêndice A.

Baixa Resolução (LD):

1. EPFL/PoliMi (*École Polytechnique Fédérale de Lausanne/Politecnico di Milano*) (SIMONE *et al.*, 2009, 2010): 78 fluxos de vídeo (*streams*) codificados e comprimidos em H.264/AVC com adição de ruído por simulação com perda de pacotes em transmissão de rede IP. Esta base de dados inclui vídeos em resolução CIF (352×288 *pixels*). Os escores subjetivos foram obtidos em duas instituições acadêmicas: PoliMi (Itália) e EPFL (Suíça).

2. IRCCyN/IVC H264 AVC *vs.* SVC (*Scalable Video Coding*) VGA (*Video Graphics Array*) *Video database* (PITREY *et al.*, 2010a): 28 sequências de vídeo com quatro diferen-

tes referências com resolução QVGA (*Quarter Video Graphics Array*) de 320×240 *pixels* e configurações de teste ou HRC (*Hypothetical Reference Circuits*) que são versões dos vídeos originais contendo algum tipo de distorção (artefato), as quais são apresentadas aos avaliadores (VQEG, 2008, 2009, 2010). Os vídeos processados (PVS) por HRC apresentam variação na taxa de compressão dos padrões de codificação H.264 e H.264/SVC sem a adição de erros de transmissão.

3. IST (Instituto Superior Técnico de Lisboa) (BRANDÃO; QUELUZ, 2010): 98 amostras de vídeo com resolução CIF (352×288 *pixels*) e degradação DCR (*Degradation Category Rating*), contendo doze referências distintas. A metodologia adotada obedeceu à recomendação ITU-T P.910 (1999). As sequências de vídeos foram comprimidas com diferentes taxas de codificação H.264 e MPEG-2. Somente os artefatos provenientes de compressão foram considerados.

Resolução Padrão (SD):

4. EPFL/PoliMi (SIMONE *et al.*, 2009, 2010): 78 amostras em resolução 4CIF (*Four times Common Intermediate Format*) com 704×576 *pixels*, cujo conteúdo e descrição são os mesmos do item 1.

5. IRCCyN/IVC H.264 AVC *vs.* SVC VGA *Video Database* (PITREY *et al.*, 2010a): 28 sequências de vídeo com quatro diferentes referências com resolução VGA (640×480 *pixels*) e formato HRC baseado na variação da taxa de compressão dos padrões de codificação H.264 e H.264/SVC sem a adição de erros de transmissão.

6. IRCCyN/IVC *Influence Content Video* VGA *Database* (PITREY *et al.*, 2012): 300 amostras com resolução VGA (640×480 *pixels*), sendo 240 distorcidas por variações durante a compressão em HRC no padrão de codificação H.264/SVC, contendo 60 sequências diferentes de referência. Dessa forma, em cada conjunto de 20 amostras foram geradas quatro degradações distintas e aleatórias. O HRC foi baseado em codificação H.264/SVC sem erros de transmissão com diversas partições do parâmetro QP entre a camada base (*base layer*) e a melhorada (*enhancement layer*).

7. IRCCyN/IVC SVC4QoE (*SVC for Quality of Experience*) QP0 (*QP for the base layer*) QP1 (*QP for the enhancement layer*) *Video* VGA *Database* (PITREY *et al.*, 2011a); 324 amostras de vídeo com resolução VGA (640×480 *pixels*): 11 sequências de referência diferentes e 313 distorcidas por codificação H.264 e H.264/SVC, mediante a aplicação de 29 diferentes HRC, tomando diversas partições QP entre as camadas base e melhorada com a metodologia ACR.

8. IRCCyN/IVC SVC4QoE *Replace Slice Video* VGA *Database* (PITREY *et al.*, 2011a): 135 amostras com resolução VGA (640×480 *pixels*), sendo nove sequências de referência distintas e 126 distorcidas por codificação H.264 e H.264/SVC com algumas amostras contendo erros de transmissão. As sequências de vídeo contendo degradações foram obtidas por meio da simulação de erros de transmissão com a aplicação de 14 parâmetros HRC diferentes.

9. IRCCyN/IVC SVC4QoE *Temporal Switch Video* VGA *Database* (PITREY *et al.*, 2011b): 423 amostras com resolução VGA (640×480 *pixels*), sendo onze vídeos de referência distintos e das 412 sequências restantes, algumas foram distorcidas por variação na compressão dos padrões de codificação H.264 e H.264/SVC, pela manipulação do parâmetro QP entre as camadas base e melhorada do SVC. A degradação dos vídeos distorcidos foi realizada com 36 HRC diferentes.

Alta Resolução no Modo Entrelaçado (HDi):

10. HDTV Phase I VQEGHD-4 (VQEG, 2010): 168 amostras em definição HD (1920×1080 *pixels*) no modo entrelaçado. As versões degradadas são variações de HRC nos padrões de codificação H.264 e MPEG-2 sem erros de transmissão.

11. IRCCyN/IVC 1080i *Database* (PÉCHARD *et al.*, 2008): 192 amostras de vídeo com definição HD (1920×1080 *pixels*) no modo entrelaçado e 50 fps. Esta base de dados é composta por 24 sequências diferentes de referência e 168 amostras distorcidas por compressão no padrão de codificação H.264. A distorção foi gerada a partir de 24 vídeos de referência com sete variações na taxa de compressão H.264.

12. IRCCyN/IVC H.264 HD *vs. Upscaling and Interlacing Video Database* (PITREY *et al.*, 2010b): 87 amostras em HD (1920×1080 *pixels*) no modo entrelaçado e 50 fps. Esta base de dados possui três sequências originais, das quais 84 são versões distorcidas por 28 variações de degradação (HRC) no padrão de codificação H.264 sem erros de transmissão. Além disso, foram empregadas seis resoluções no processo *upscaling*: 1920×1080p, 1920×1080i, 1280×720p, 1280×1080i, 1280×1080p e SD (720×576p), bem como três taxas de compressão: 3 M*bits*/s, 6 M*bits*/s e 9 M*bits*/s. Todas as amostras processadas (PVS) foram exibidas aos avaliadores em definição HD (1920×1080) *pixels* no modo progressivo. Além disso, durante a exibição foi aplicado o desentrelaçamento nas sequências que estavam no modo entrelaçado.

13. VQEG *Pool2* 1080i *Video Database* (BARKOWSKY *et al.*, 2010): 168 amostras em HD (1920×1080 *pixels*) no modo entrelaçado, também é denominada como HDTV Phase I VQEGHD-2 que também está referenciada em (VQEG, 2010). Esta base de dados contém nove sequências originais, das quais 159 são versões distorcidas por quinze variações de degradação

(HRC) nos padrões de codificação H.264 e MPEG-2 sem erros de transmissão.

Alta Resolução no Modo Progressivo (HDp):

14. HDTV Phase I VQEGHD-1 (VQEG, 2010): 168 amostras em definição HD (1920×1080 *pixels*) no modo progressivo. As versões degradadas são variações de HRC nos padrões de codificação H.264 e MPEG-2 sem erros de transmissão.

15. HDTV Phase I VQEGHD-3 (VQEG, 2010): 168 amostras e mesma descrição do item 14. Entretanto, esta base de dados apresenta conteúdos diferentes.

16. HDTV Phase I VQEGHD-5 (VQEG, 2010): 136 amostras e mesma descrição do item 14. Todavia, esta base de dados possui conteúdos distintos.

17. TUM (*Technische Universität München*) 1080p25 *Data Set* (KEIMEL *et al.*, 2010): 48 amostras de vídeo em definição HD (1920×1080 *pixels*) no modo progressivo e 25 fps, sendo quatro sequências de referência com dezesseis versões codificadas em Dirac *wavelet* com quatro taxas de compressão entre 5 e 30 M*bits*/s. Além disso, esta base de dados também contém 16 amostras em H.264 codificadas com baixa complexidade (LC) e dezesseis sequências codificadas com alta complexidade (HC).

Na maioria das bases de dados do superconjunto S não é possível identificar de maneira explícita qual o tipo de degradação nas amostras de vídeo. Assim, para efeito de validação dos métodos propostos, o superconjunto S foi agrupado em subcategorias, segundo a definição LD, SD, HDi e HDp ou todas as suas amostras foram consideradas no processo de validação.

3.2 PROCESSAMENTO DOS ESCORES OBJETIVOS E SUBJETIVOS

O mapeamento entre os escores objetivos e subjetivos e o processo de validação dos métodos propostos nesta tese seguiram as recomendações disponíveis nas versões mais atuais do Grupo de Especialistas em Qualidade de Vídeo ou VQEG (VQEG, 2008, 2009, 2010) que é um órgão internacional voltado à padronização de métricas subjetivas e objetivas de qualidade de vídeo. A Figura 7 resume as etapas de mapeamento entre os escores objetivos (VQAS – *Video Quality Assessment Scores*) e subjetivos (MOS/DMOS), bem como o processo de validação de métodos objetivos de qualidade de vídeo. Além disso, o VQEG (VQEG, 2000, 2003, 2008, 2009, 2010) recomenda que a MOS seja aplicada em métricas NR, enquanto que a DMOS seja utilizada em métricas FR/RR, cujas amostras de referência sejam ocultadas no processo de avaliação. Nesta tese, utilizou-se uma função polinomial cúbica no mapeamento não linear entre VQAS (x) e a MOS/DMOS, conforme recomendações mais atuais do VQEG (VQEG,

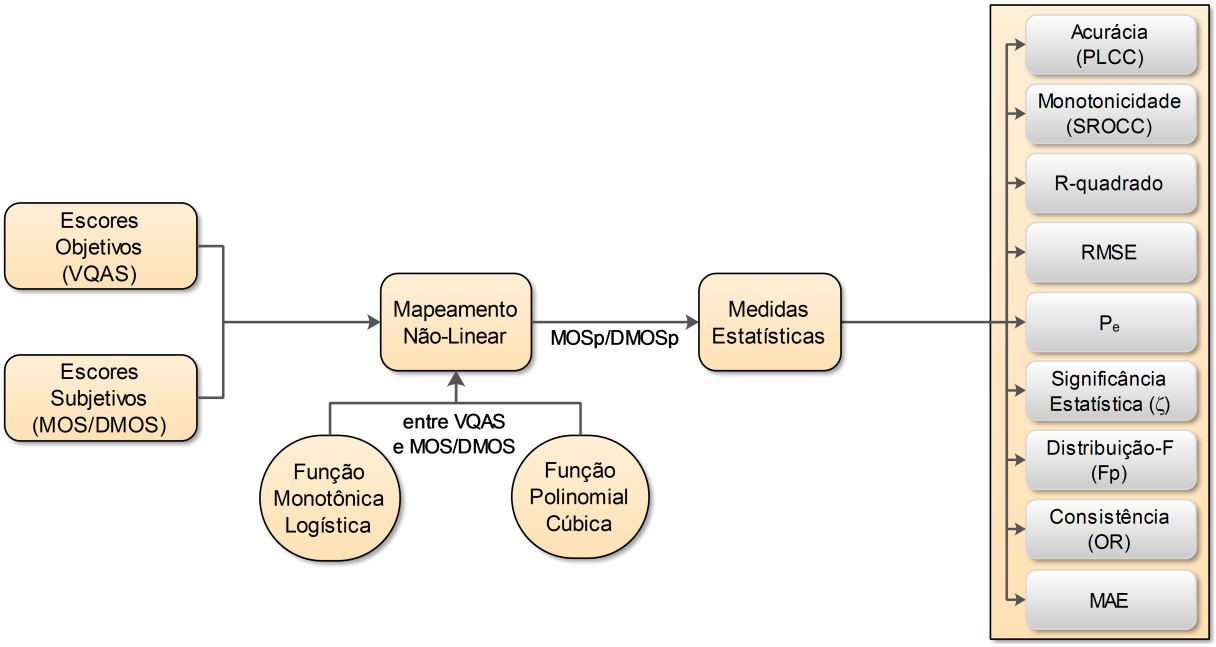

Figura 7: Pós-processamento dos escores objetivos e medidas estatísticas de desempenho. Fonte: Autoria própria.

2008, 2009, 2010) que além de ser uma predição mais simples, não causa sobreposição de dados como pode ocorrer com a função monotônica logística (ENGELKE *et al.*, 2009). A função polinomial cúbica é definida como

$$\text{MOSp} = ax^3 + bx^2 + cx + d, \tag{30}$$

em que os coeficientes a, b, c e d são obtidos por um ajuste (*fitting*) cúbico entre os escores subjetivos e objetivos. Com o cálculo da Equação (30), os escores objetivos (x) são mapeados para a escala MOS e recebem a denominação de MOSp/DMOSp (*predicted* MOS/DMOS), conforme inspeção visual da Figura 7. Procedimento análogo ao da Equação (30) também pode ser aplicado à DMOSp. Logo, com o objetivo de simplificação, deste ponto em diante, serão usadas as notações MOS e MOSp.

Seshadrinathan *et al.* (2010) empregam uma versão baseada em recomendações já ultrapassadas do VQEG que sugerem a utilização de uma função de mapeamento monotônica logística, conforme expressão a seguir (VQEG, 2000, 2003).

$$\text{MOS}_p^l = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-\left(\frac{x-\beta_3}{|\beta_4|}\right)}}, \tag{31}$$

em que β_1 a β_4 são parâmetros otimizados com a estimação de mínimos quadrados pelo método

iterativo LM (LEVENBERG, 1944; MARQUARDT, 1963; MORÉ, 1977).

A Figura 8 compara as funções de mapeamento cúbica (primeira coluna) e logística (segunda coluna) usando as métricas PSNR, SSIM e MS-SSIM, a partir da base de dados LIVE (SESHADRINATHAN *et al.*, 2010) que contém 150 vídeos. Nesta comparação, observa-se que ambas as funções de mapeamento apresentam desempenho equivalente para a medida de acurácia (PLCC).

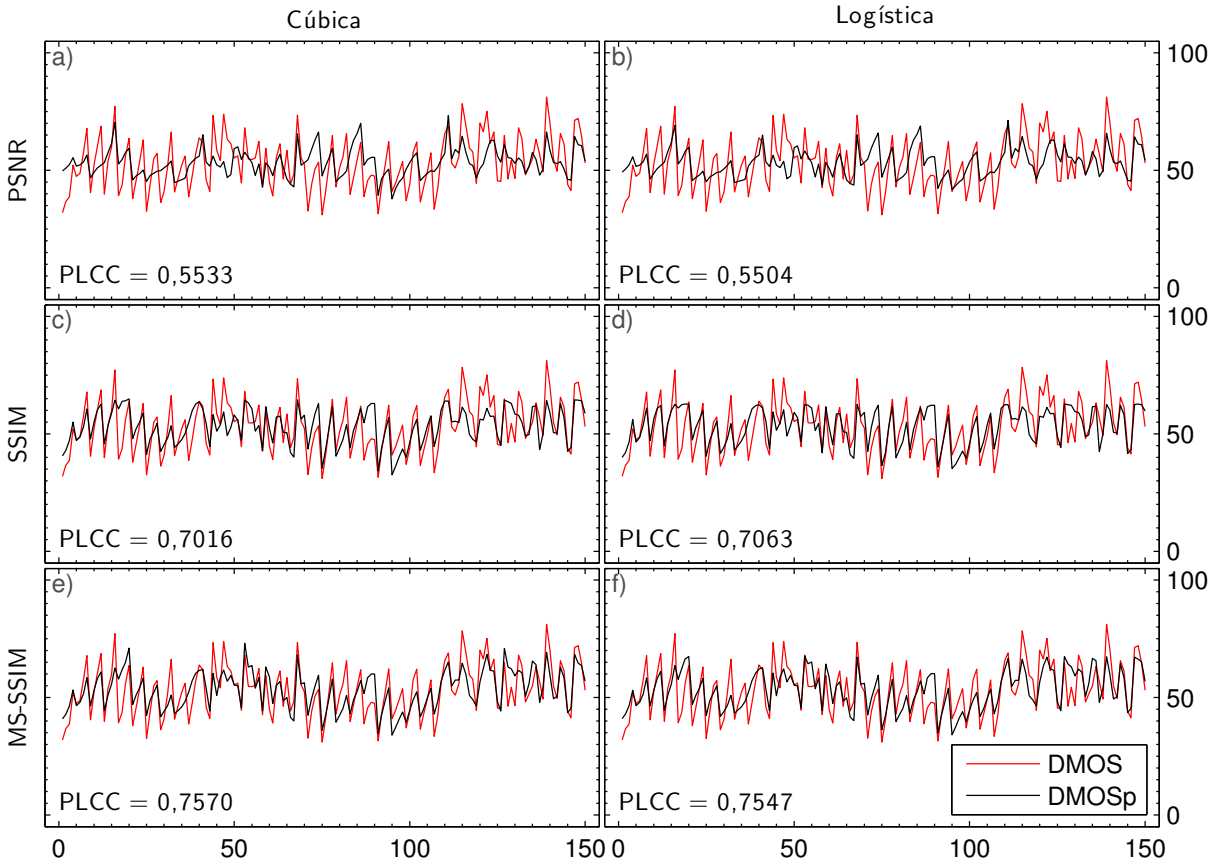

Figura 8: Comparação entre as funções de mapeamento cúbica e logística usando os vídeos da base de dados LIVE.

A comparação entre funções de mapeamento apresenta diferenças mais significativas quando é aplicada em imagens, *e.g.*, a Figura 9 compara as funções de mapeamento cúbica e logística para 808 imagens da base de dados LIVE (SHEIKH *et al.*, 2003). Esta base de dados contém artefatos de compressão JPEG-2000, JPEG, borramento gaussiano e desvanecimento rápido ou *fast fading* que é uma simulação de canal ruidoso com imagens comprimidas em JPEG-2000. A inspeção visual da Figura 9-a mostra que a função de mapeamento cúbica, quando aplicada à métrica PSNR, apresenta uma diferença significativa da acurácia (PLCC = 0,8384) em comparação com a função logística (Figura 9-b), cujo PLCC é igual a 0,7218. O mapeamento com a função cúbica acompanha as transições dos valores de PSNR na escala da DMOS, enquanto a função de mapeamento logística apresenta uma tendência de sobreposição

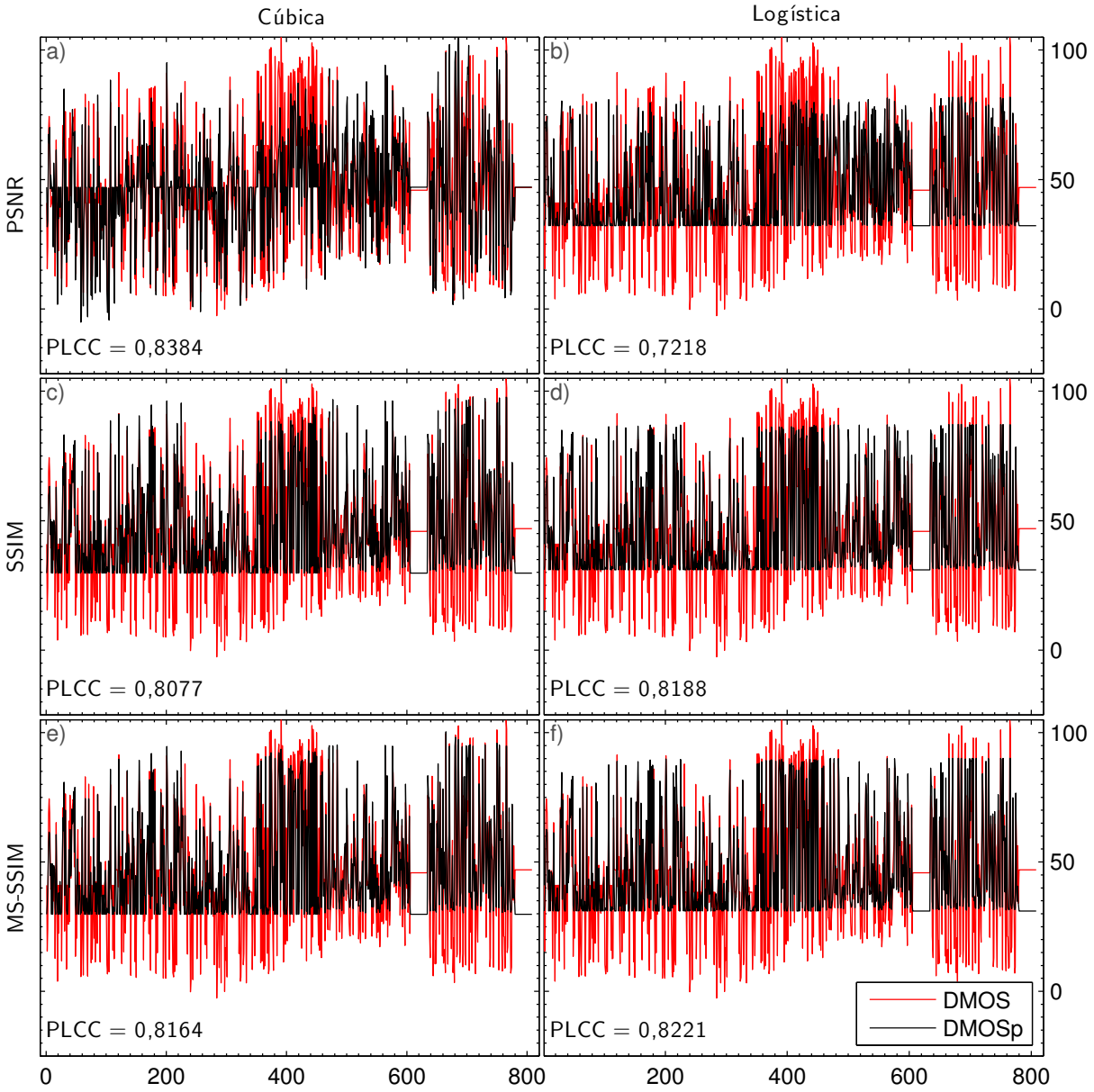

Figura 9: Comparação entre as funções de mapeamento cúbica e logística usando 808 imagens da base de dados LIVE (SHEIKH *et al.*, 2003).

de dados em torno de 32 dB, conforme a Figura 9-b. Entretanto, para as métricas SSIM e MS-SSIM, ambas as funções de mapeamento apresentam desempenho equivalente.

Após o mapeamento não linear entre os escores objetivo e subjetivo, ocorre o processo de validação entre a MOS e a MOSp com as medidas estatísticas de desempenho, dentre as quais o coeficiente de correlação linear de Pearson (PLCC) e o coeficiente de correlação de postos de Spearman (SROCC), R-quadrado (R^2), RMSE, P_e (erro entre os escores subjetivos e objetivos), significância estatística (ζ), distribuição F percentual (F_p) com quatro graus de liberdade, OR (*Outlier Ratio*) ou consistência e a média do erro absoluto ou MAE (*Mean Absolute Error*).

3.2.1 MEDIDAS ESTATÍSTICAS DE DESEMPENHO

Esta subseção descreve as medidas estatísticas de desempenho recomendadas pelo VQEG para validação de métricas objetivas de qualidade de vídeo, quando comparadas com métricas subjetivas ou outras métricas objetivas, sejam elas pertencentes às categorias RR, FR ou NR.

O VQEG (VQEG, 2000, 2003, 2008, 2009, 2010) recomenda que as amostras de referência sejam excluídas do processo de avaliação ou validação em métricas FR/RR, *i.e.*, $\xi = F - R$, enquanto que todas as amostras participem da avaliação ou validação em métricas NR, *i.e.*, $\xi = F$; sendo F o número total de amostras e R a quantidade de amostras de referência. Assim, como já fora mencionado, segundo o VQEG, métricas FR e RR devem empregar a DMOS e métricas NR devem utilizar a MOS.

A predição da acurácia de uma medida objetiva de qualidade de vídeo é determinada pelo coeficiente de correlação linear de Pearson ou PLCC. Quanto mais próximo de 1 ou -1, maior é a correlação entre a medida objetiva e a subjetiva, *i.e.*, mais forte é associação (acurácia) da medida objetiva em relação à percepção do SVH. O coeficiente PLCC é calculado tomando um conjunto de escores de vídeo ξ, associados aos escores subjetivos μ_k e objetivos ν_k, conforme a expressão (SPIEGEL; STEPHENS, 1998; VQEG, 2000, 2003, 2008, 2009, 2010):

$$\text{PLCC} = \frac{\sum_{k=1}^{\xi} (\mu_k - \overline{\mu})(\nu_k - \overline{\nu})}{\sqrt{\sum_{k=1}^{\xi} (\mu_k - \overline{\mu})^2} \sqrt{\sum_{k=1}^{\xi} (\nu_k - \overline{\nu})^2}}, \tag{32}$$

em que $\overline{\mu}$ e $\overline{\nu}$ são as médias do conjunto de escores subjetivos e objetivos, respectivamente. Além disso, a medida PLCC está confinada ao intervalo $[-1, 1]$.

A monotonicidade representada pelo coeficiente de correlação de postos de Spearman (SROCC) quantifica as mudanças em uma medida, *i.e.*, detecta seu acréscimo ou decréscimo, bem como acompanha as alterações de magnitude de uma medida em relação à outra, ou seja, mede a variação dos escores objetivos em relação aos subjetivos. Além disso, quanto mais sua medida se aproximar de $|1|$, maior é a correlação de postos de Spearman entre os escores objetivos e subjetivos. A monotonicidade é expressa pelo coeficiente SROCC, conforme a seguinte expressão (SPIEGEL; STEPHENS, 1998; VQEG, 2000, 2003)

$$\text{SROCC} = \frac{\sum_{k=1}^{\xi} (\rho_k - \overline{\rho})(\gamma_k - \overline{\gamma})}{\sqrt{\sum_{k=1}^{\xi} (\rho_k - \overline{\rho})^2} \sqrt{\sum_{k=1}^{\xi} (\gamma_k - \overline{\gamma})^2}}, \tag{33}$$

em que ρ_k e γ_k são os postos da medida subjetiva e objetiva de uma amostra de vídeo k, respecti-

vamente. Além disso, $\overline{\rho}$ e $\overline{\gamma}$ são as médias de um conjunto de postos (ξ) subjetivos e objetivos, respectivamente. Da mesma forma que o coeficiente PLCC, a medida SROCC também está confinada ao intervalo $[-1,1]$.

O coeficiente de determinação, representado pelo R-quadrado (R^2) (SPIEGEL; STEPHENS, 1998) indica o percentual de determinação das variáveis independentes em relação às variáveis dependentes, *i.e.*, mede a componente de regressão decorrente da variação simultânea entre os escores subjetivos e objetivos. O coeficiente de determinação, também conhecido como coeficiente de explicação, é definido conforme as fórmulas a seguir.

$$R^2 = 1 - \frac{SSE}{TSS}, \tag{34}$$

ou

$$R^2 = 1 - \frac{\sqrt{\frac{\sum_{k=1}^{\xi}(\mathrm{MOS}_k - \mathrm{MOSp}_k)^2}{\xi - d}}}{\sum_{k=1}^{\xi}\left(\mathrm{MOS}_k - \overline{\mathrm{MOS}}\right)^2}, \tag{35}$$

em que MOS_k e MOSp_k são a MOS e a MOS predita da k-ésima amostra de vídeo, respectivamente; $\overline{\mathrm{MOS}}$ é a média da MOS sobre as amostras de vídeo ξ e d é o número de graus de liberdade, no caso de uma regressão linear, $d = 2$. A soma de erros quadráticos ou SSE (*Sum of Square Errors*) é o numerador da fração na Fórmula (35) e a soma total dos quadrados ou TSS (*Total Square Sum*) é definida pelo denominador da fração da Fórmula (35). A medida R^2 assume valores no intervalo $[0,1]$ e indica o aumento do desempenho da predição de qualidade quanto mais próximo for de 1.

O RMSE do erro absoluto de predição P_e entre os escores subjetivos e objetivos é calculado conforme (VQEG, 2008, 2009, 2010):

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{\xi - d}\right)\sum_{k=1}^{\xi} P_e(k)^2}, \tag{36}$$

em que ξ é o número amostras de vídeo consideradas na análise. O número de graus de liberdade d vem da função de mapeamento da Equação (30) que possui quatro graus de liberdade (VQEG, 2008, 2009, 2010), *i.e.*, $d = 4$. O erro absoluto de predição da k-ésima amostra de vídeo é representado conforme expressão a seguir (VQEG, 2008, 2009, 2010).

$$P_e(k) = MOS(k) - MOSp(k). \tag{37}$$

Quanto mais próximo de 0 for o valor do RMSE, melhor é o desempenho da medida de qualidade de vídeo. Entretanto, Okamoto *et al.* (2006) realizaram um estudo experimental

com vídeos codificados em diferentes taxas de compressão que relata uma acurácia na predição, envolvendo valores de RMSE menores do que 7,24. O intervalo de confiança de 95% para o RMSE, segundo uma distribuição qui-quadrado, $\chi^2(\xi - d)$, está definido no intervalo (VQEG, 2008, 2009, 2010)

$$\frac{\text{RMSE}\sqrt{\xi - d}}{\sqrt{\chi^2_{0.025}(\xi - d)}} < \text{RMSE} < \frac{\text{RMSE}\sqrt{\xi - d}}{\sqrt{\chi^2_{0.975}(\xi - d)}}. \tag{38}$$

A significância estatística ζ entre o RMSE de uma métrica qualquer e o método proposto é calculada conforme a fórmula a seguir (VQEG, 2008, 2009, 2010).

$$\zeta = \frac{(\text{RMSE}_{\text{max}})^2}{(\text{RMSE}_{\text{min}})^2}, \tag{39}$$

em que são considerados os valores de RMSE da métrica comparada e do método proposto. Entretanto, a significância estatística ζ, segundo uma distribuição F de Snedecor (VEERARAJAN, 2008) com $F(0,05; \xi - d_1; \xi - d_2)$ é definida pela inversa da função de distribuição acumulada ou iCDF (*inverse of the Cumulative Distribution Function*) (ABRAMOWITZ; STEGUN, 1964), a qual garante um nível de significância de 95%, com d_1 e d_2 graus de liberdade para o RMSE_{max} e RMSE_{min}, respectivamente (VQEG, 2008, 2009, 2010). Se ζ for maior do que $F(0,05; \xi - d_1; \xi - d_2)$ haverá uma diferença significativa entre os valores de RMSE. Alternativamente, quando o RMSE do método proposto é igual a RMSE_{min}, a Fórmula (39) pode ser representada em termos de uma distribuição F percentual (F_p), a qual é expressa por

$$F_p = (\zeta - 1) \cdot 100. \tag{40}$$

Considerando que d_1 e d_2 possuem quatro graus de liberdade, conforme a Equação (36), em que $d = 4$, tanto para RMSE_{max} quanto para RMSE_{min}, haverá uma diferença significativa entre o método proposto e a métrica comparada quando $\zeta > F(0,05; v_1; v_2)$, conforme a iCDF (ABRAMOWITZ; STEGUN, 1964):

$$\zeta = F^{-1}(p \mid v_1, v_2) = \{\zeta : F(\zeta \mid v_1, v_2) = p\}, \tag{41}$$

em que p é a função densidade de probabilidade, expressa por

$$p = F(\zeta \mid v_1, v_2) = \int_0^{\zeta} \frac{\Gamma\left[\frac{(v_1 + v_2)}{2}\right]}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}} \frac{t^{\frac{v_1 - 2}{2}}}{\left[1 + \left(\frac{v_1}{v_2}\right)t\right]^{\frac{v_1 + v_2}{2}}}\, dt, \tag{42}$$

com $v_1 = \xi - d_1$ e $v_2 = \xi - d_2$.

Assim, se $\xi = 150$ e $d = 4$, então $\xi - d = 146$ e $\zeta = 1,3141$. Logo, substituindo $\zeta = 1,3141$ na Fórmula (40) haverá uma diferença significativa positiva, *i.e.*, o método proposto apresenta desempenho superior à métrica comparada, se $F_p > 31,41\%$. Quando o método proposto apresenta desempenho inferior à métrica comparada, *i.e.*, o RMSE do método proposto é igual a $\text{RMSE}_{\max}$, a Fórmula (40) é reescrita como

$$F_p = (\zeta + 1) \cdot 100, \tag{43}$$

em que a Fórmula (41) é reescrita com o sinal negativo:

$$\zeta = -F^{-1}(p \mid v_1, v_2) = -\{\zeta : F(\zeta \mid v_1, v_2) = p\}. \tag{44}$$

Neste caso, considerando $\xi - d = 146$ e $\zeta = -1,3141$, haverá uma diferença significativa negativa entre a métrica comparada e o método proposto, se $F_p < -31,41\%$.

A predição da consistência (OR) representa a razão entre o número de pontos discrepantes Θ e a quantidade de amostras ξ. Esta medida é relatada em (VQEG, 2000, 2003, 2008, 2009) e simplificada em (VQEG, 2003; ENGELKE *et al.*, 2009)

$$OR = \frac{\Theta}{\xi}, \tag{45}$$

em que Θ é determinado pelo número de ocorrências da condição (46), pela comparação entre o k-ésimo erro absoluto de predição $P_e(k)$ e o desvio-padrão $\sigma_{MOS(k)}$ de cada amostra de vídeo k.

$$P_e(k) > 2\sigma_{MOS(k)}, \quad k = 1, ..., \xi. \tag{46}$$

Além disso, a predição da média do erro absoluto (MAE) entre a MOS e a MOSp é calculada conforme (SPIEGEL; STEPHENS, 1998)

$$\text{MAE} = \frac{1}{\xi} \sum_{k=1}^{\xi} P_e(k). \tag{47}$$

Ambas as medidas, OR e MAE indicam uma melhor predição quando seus valores se aproximam de 0 (SPIEGEL; STEPHENS, 1998; ENGELKE *et al.*, 2009).

3.3 CARACTERÍSTICAS ESPAÇO-TEMPORAIS

O SVH é mais sensível às variações de brilho (luminância) do que às variações de cor (crominância) (MARQUES FILHO; VIEIRA NETO, 1999; WU *et al.*, 2007), por isso, nesta tese foram consideradas apenas as características espaço-temporais relacionadas à componentes

de luminância. Os modelos propostos nesta tese empregam três características espaciais associadas à detecção de artefatos de blocagem com o descritor B e de borramento com os descritores A e Z (WANG *et al.*, 2002). Além disso, são usados três descritores relacionados à variação temporal entre quadros: informação perceptual temporal TI, média da diferença absoluta MAD (*Mean Absolute Difference*) e média da diferença absoluta ponderada MADw (*Mean Absolute Difference weight*).

O descritor de blocagem B para um conjunto de quadros Q é obtido a partir da média da Fórmula (18).

$$B = \frac{1}{Q}\sum_{f=1}^{Q} B_f. \tag{48}$$

Os descritores de borramento A e Z para uma sequência de vídeo com Q quadros são reescritos a partir das médias das Fórmulas (21) e (26).

$$A = \frac{1}{Q}\sum_{f=1}^{Q} A_f, \tag{49}$$

$$Z = \frac{1}{Q}\sum_{f=1}^{Q} Z_f. \tag{50}$$

O descritor TI (ITU-T P.910, 1999) é uma versão da Equação (28) descrita no início desta seção. A Equação (51), diferentemente da Equação (28), não usa o valor de pico (max) no cálculo de sua medida, pois considera a média do desvio-padrão da medida $m(f,i,j)$ para f quadros.

$$\mathrm{TI} = \frac{1}{Q-1}\sum_{f=2}^{Q} \sigma\left[m(f,i,j)\right], \tag{51}$$

em que Q é o número total de quadros de um vídeo e $\sigma\left[m(f,i,j)\right]$ é o desvio-padrão da diferença de movimento entre quadros sucessivos.

O descritor MAD (DING *et al.*, 2008) representa a diferença temporal absoluta entre quadros sucessivos $m(f,i,j)$, cuja definição é

$$\mathrm{MAD}_k = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} \left|m(f,i,j)\right|, \tag{52}$$

em que $f > 1$ e a média de MAD_k tem $Q-1$ amostras, *i.e.*, $k = 1,2,\ldots,Q-1$. Logo, o descritor MAD_k tem a média definida como

$$\overline{\mathrm{MAD}} = \frac{1}{Q-1}\sum_{k=1}^{Q-1} \mathrm{MAD}_k. \tag{53}$$

O descritor MADw (DING *et al.*, 2008) considera a razão entre o valor de MAD do quadro atual k e do anterior $k-1$, conforme a definição a seguir.

$$\text{MADw} = \frac{1}{Q-1}\sum_{k=1}^{Q-1}\frac{\text{MAD}_k}{\text{MAD}_{k-1}}. \tag{54}$$

A Figura 10 exibe o comportamento dos seis descritores A, B, Z, TI, $\overline{\text{MAD}}$ e MADw que caracterizam trinta vídeos codificados em MPEG-2 da base de dados IVP (LI; MA, 2012). Em algumas amostras de vídeo da Figura 10, *e.g.*, no intervalo $[5,10]$ e $[25,30]$, observa-se a mesma tendência tanto nas curvas espaciais (A,B,Z) quanto nas temporais $(\text{TI},\overline{\text{MAD}})$.

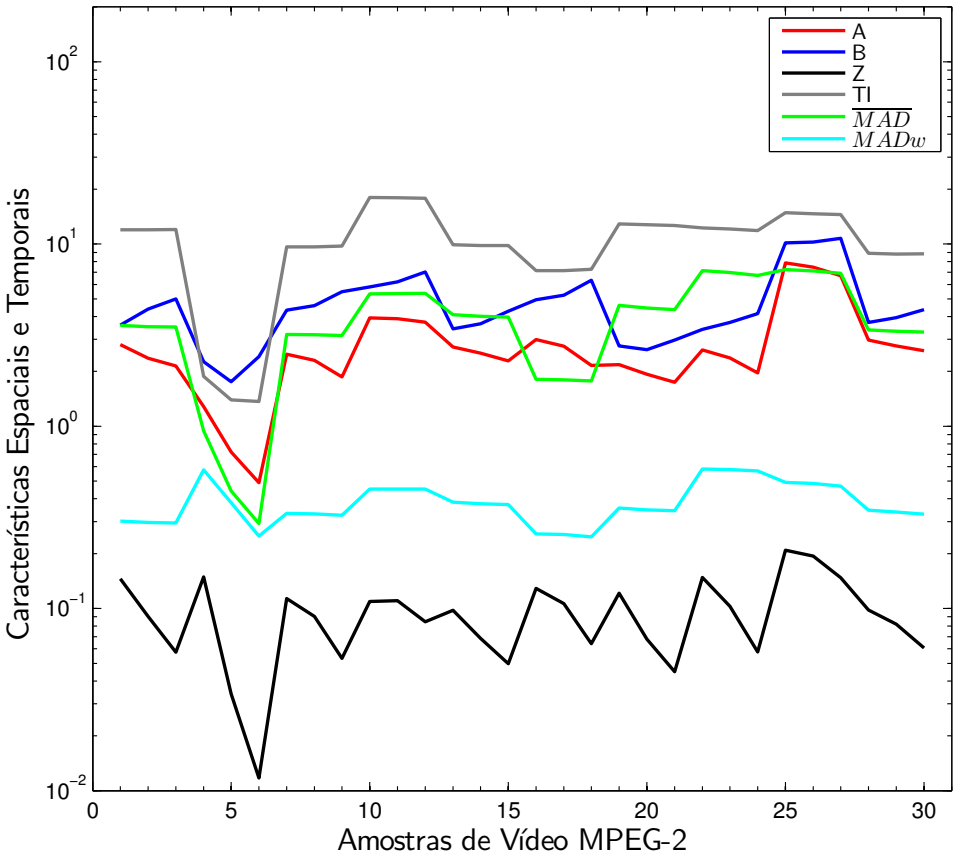

Figura 10: Características espaço-temporais para vídeos em MPEG-2 (base de dados IVP).

3.4 MÉTODO ITERATIVO DE LEVENBERG-MARQUARDT (LM)

O método iterativo LM (LEVENBERG, 1944; MARQUARDT, 1963; MORÉ, 1977) combina a vantagem de aceleração no processo de localização do mínimo de uma função não linear baseada no método do gradiente descendente (WIDROW; HOFF, 1960) e também da aceleração de convergência baseada na localização de mínimos adjacentes pelo método de Gauss-Newton (HOFFMAN, 2001).

As Equações (55) a (63) detalham o método LM, cuja finalidade é minimizar o vetor de erro ($\mathbf{d} = \mathbf{s} - \widehat{\mathbf{s}}$) que considera o vetor de escores de qualidade $\mathbf{s}$ e a saída desejada, *i.e.*, o vetor de escores subjetivos $\widehat{\mathbf{s}}$.

$$\underset{(\beta_1,\beta_2,\ldots,\beta_7)}{\arg\min} \ \mathbf{d}^{\mathrm{T}}\mathbf{d}, \tag{55}$$

em que $\mathbf{d} = \left[\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_\phi\right]$ e $\mathbf{d}_\phi = \left[\widehat{\mathbf{s}}_\phi - f\left(A_\phi, B_\phi, Z_\phi, TI_\phi, \overline{MAD}_\phi, MADw_\phi; \beta_1, \beta_2, \ldots, \beta_7\right)\right]$ com $\widehat{\mathbf{s}}_\phi$ denotando o escore subjetivo da amostra de vídeo ϕ de um determinado subconjunto de uma base de dados. A função f que descreve o produto $\mathbf{d}^{\mathrm{T}}\mathbf{d}$ na Equação (56) é determinada pela relação entre as características espaciais e temporais descritas na Seção 3.3 e os parâmetros β. Assim, f é definida pelo modelo não linear proposto, cujos resultados são os escores objetivos, representados pelo vetor $\mathbf{s}$.

$$\mathbf{d}^{\mathrm{T}}\mathbf{d} = f\left(A_\phi, B_\phi, Z_\phi, TI_\phi, \overline{MAD}_\phi, MADw_\phi; \beta_1, \beta_2, \ldots, \beta_7\right). \tag{56}$$

O método de Gauss-Newton é uma extensão do método de Newton. Este requer que a matriz Hessiana ($\mathbf{H}$) seja inversível e definida positivamente (MORÉ, 1977). Todavia, nem sempre há garantias de que $\mathbf{H}$ obedeça a estas condições, ao passo que o método de Gauss-Newton requer apenas a matriz Jacobiana do vetor de erros. O método de Gauss-Newton é utilizado para minimizar a função f com uma fórmula recorrente que considera os termos $\mathbf{H}$ e o gradiente de uma função (∇f).

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \mathbf{H}^{-1}\nabla f, \tag{57}$$

em que k é o número da iteração, $\mathbf{H}$ é a matriz quadrada das derivadas parciais de segunda ordem da função f que pode ser aproximada conforme a expressão a seguir.

$$\mathbf{H} \approx \mathbf{J}^{\mathrm{T}}\mathbf{J}, \tag{58}$$

e o gradiente ∇f é expresso conforme

$$\nabla f = \mathbf{J}^{\mathrm{T}}\mathbf{d}, \tag{59}$$

em que $\mathbf{J}$ é a matriz Jacobiana que contém as derivadas primeiras do vetor de erros $\mathbf{d}$.

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial d_1}{\partial \beta_1} & \cdots & \dfrac{\partial d_1}{\partial \beta_7} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial d_\phi}{\partial \beta_1} & \cdots & \dfrac{\partial d_\phi}{\partial \beta_7} \end{pmatrix}. \tag{60}$$

O método de Gauss-Newton apesar de rápido, apresenta problemas de estabilidade quanto à inversibilidade da matriz Hessiana aproximada, *i.e.*, $\mathbf{H} \approx \mathbf{J}^{\mathrm{T}}\mathbf{J}$. Levenberg (1944) propôs uma alteração no método de Gauss-Newton com a introdução de uma solução iterativa, envolvendo a matriz identidade $\mathbf{I}$ e o coeficiente de combinação λ (fator de ajuste), cujo valor é alterado a cada iteração.

$$\mathbf{H} \approx \mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda_k \mathbf{I}, \tag{61}$$

em que k é o número de iteração e λ deve ser sempre positivo para que torne $\mathbf{H}$ definida positivamente e inversível.

O método de Levenberg altera o método de Gauss-Newton com uma solução aproximada de $\mathbf{H}$. Logo, a Equação (57) é reescrita em termos da Jacobiana do vetor de erros.

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \left(\mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda_k \mathbf{I}\right)^{-1} \mathbf{J}^{\mathrm{T}}\mathbf{d}. \tag{62}$$

Caso λ seja muito pequeno ($\lambda \to 0$), emprega-se o método de Gauss-Newton. Entretanto, se λ for grande ($\lambda \to \infty$) utiliza-se o método do gradiente descendente com um pequeno passo (LEVENBERG, 1944). Inicialmente, nesta tese adotou-se $\lambda = 10^{-4}$, mas a cada iteração seu valor é alterado.

O método de Levenberg apresenta instabilidade quando λ cresce muito ($\lambda \to \infty$). Marquardt (1963) propôs um aperfeiçoamento no método de Levenberg (62) com a substituição da matriz identidade $\mathbf{I}$ pela matriz diagonal dos elementos de $\mathbf{J}^{\mathrm{T}}\mathbf{J}$.

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \left[\mathbf{J}^{\mathrm{T}}\mathbf{J} + \mathrm{diag}\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\lambda_k\right]^{-1} \mathbf{J}^{\mathrm{T}}\mathbf{d}. \tag{63}$$

Logo, solução da Equação (63) é definida como método de Levenberg-Marquardt (LEVENBERG, 1944; MARQUARDT, 1963; MORÉ, 1977).

3.5 ALGORITMO ELM

O sistema nervoso humano é constituído de células que realizam funções altamente especializadas, conhecidas como neurônios. Eles possuem as mesmas estruturas de uma célula convencional, porém com o adendo de extensões em forma de filamentos, designadas dendritos e axônios, que se ramificam a partir de seu corpo celular. Os dendritos têm a função de receber os impulsos nervosos e os direcionar ao corpo celular, já os axônios transmitem o sinal para os dendritos dos neurônios vizinhos. Mcculloch e Pitts (1943) propuseram o primeiro modelo de neurônio artificial que interpreta as funções do neurônio, analogamente a um circuito binário

básico que combina inúmeras entradas e produz um sinal de saída. Nesse modelo, as entradas de um neurônio representam um vetor $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$, cuja dimensão é n. Cada entrada $\mathbf{x}_i$ possui um peso associado $\mathbf{w}_i$ que imita a concentração de neurotransmissores de uma conexão sináptica. A saída linear $\mathbf{u}$ é composta pela soma ponderada das entradas $\mathbf{x}_i$ e seus pesos correspondentes $\mathbf{w}_i$. A saída depende de uma função de ativação g que produz uma saída de ativação $\mathbf{y}$ de um nerônio, ou seja, $\mathbf{y} = g(u)$. Há várias funções de ativação, dentre as quais: base radial, sigmoidal, identidade, rampa e degrau. A Figura 11 ilustra algumas funções de ativação utilizadas pelo algoritmo ELM (HUANG *et al.*, 2006) com n amostras de treinamento.

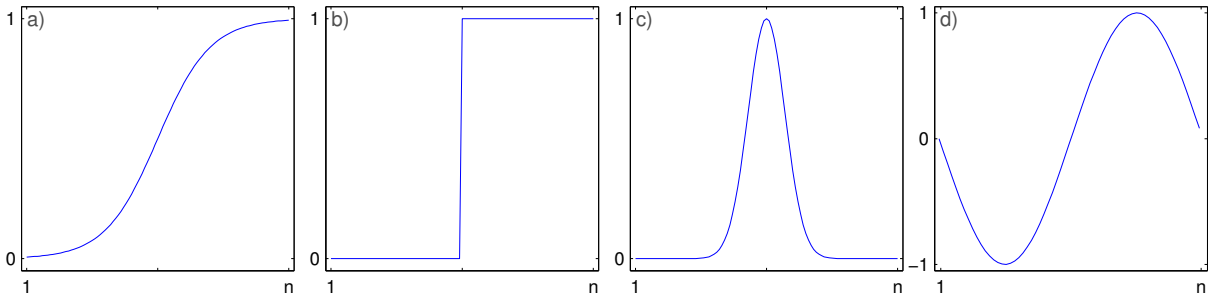

Figura 11: Funções de ativação utilizadas pelo algoritmo ELM: a) sigmoidal (sig), b) degrau (hardlim), c) base radial (radbas) e d) seno (sin).

Fonte: Autoria própria.

A organização das unidades de processamento em uma rede neural é fortemente influenciada pelas características do problema a ser resolvido, bem como a escolha do algoritmo de aprendizado. As arquiteturas de redes neurais disponíveis na literatura possuem três classificações elementares: *feedforward*, *feedback* e auto-organizável (KARAYIANNIS; VENETSANOPOULOS, 1992). A arquitetura de rede *feedforward* pode conter uma ou mais camadas de unidades de processamento, tipicamente, não lineares, cujas conexões com as camadas vizinhas são determinadas por um conjunto de pesos sinápticos. Contudo, para que seja considerada uma rede neural *feedforward* ela deve ter suas saídas conectadas apenas com as unidades de processamento da camada adjacente. As redes *perceptron* (ROSENBLATT, 1962) e ADALINE (*ADAptive LINear Element*) (WIDROW, 1987) foram as primeiras arquiteturas de rede *feedforward* que surgiram na literatura.

A arquitetura de rede neural escolhida nesta tese é a *feedforward* com aprendizado supervisionado pelo algoritmo ELM que basicamente implementa o teorema de Cover sobre a separabilidade de padrões, *i.e.*, o problema de classificação de padrões disposto de forma não linear em um espaço de alta dimensão chamado de espaço de características ou espaço oculto, em que há maior probabilidade de ser linearmente separável do que em um espaço de baixa dimensão (COVER, 1965). Tipicamente, no aprendizado supervisionado, há apresentação

de um padrão de resposta ou alvo (*target*), cujas entradas são comparadas com o alvo por meio da comparação de erros para, quando necessário, executar reajustes nos pesos sinápticos. Entretanto, no algoritmo ELM clássico, esses pesos são atribuídos aleatoriamente com uma iteração apenas.

O método do gradiente descendente ou método da descida mais íngreme (HAYKIN, 1999) é empregado em muitos algoritmos de treinamento em redes neurais artificiais. Há mais de duas décadas, esses algoritmos têm sido empregados no aprendizado de padrões. Entretanto, algoritmos baseados no método do gradiente descendente são tipicamente lentos e convergem para mínimos locais com facilidade. Nesse algoritmo o procedimento de treinamento é feito de forma iterativa até que seja conseguida uma melhor generalização, resultando em longos períodos de treinamento da rede. O algoritmo de treinamento proposto por Huang *et al.* (2006), denominado ELM, foi desenvolvido para minimizar os problemas habituais dos algoritmos de aprendizagem baseados no método do gradiente descendente.

Na etapa de aprendizagem do algoritmo ELM, os parâmetros pesos de entrada e as polarizações (*biases*) da camada oculta são calculados aleatoriamente e os pesos da camada de saída são determinados de forma analítica e aproximada. O algoritmo ELM tem aplicação exclusiva às redes neurais conhecidas como perceptrons de múltiplas camadas (MLP), no entanto, com a diferença de que o ELM tem apenas uma camada oculta (HUANG *et al.*, 2004, 2006). Contudo, esta não constitui uma limitação significativa, pois estudos realizados por Tamura e Tateishi (1997) comprovam que as redes neurais contendo apenas uma única camada oculta podem prover aproximação de qualquer função contínua. Além disso, o algoritmo de treinamento ELM apresenta baixo custo computacional e produz baixos erros de treinamento, bem como provê boa capacidade de generalização. A sua teoria envolve puramente tratamento matricial e o recurso de matriz generalizada inversa de Moore-Penrose (MP) (RAO; MITRA, 1971; SERRE, 2002). As entradas e saídas são mapeadas por meio do produto interno das entradas pelos pesos que as conectam aos nós da camada oculta.

3.5.1 TREINAMENTO DA REDE SLFN COM O ALGORITMO ELM

O treinamento de uma rede neural com única camada oculta (SLFN – *Single-Hidden Layer Feedforward Neural Network*) pelo algoritmo ELM considera N amostras de treinamento $(\mathbf{x}_i, \mathbf{t}_i)$, em que $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{in}]^{\mathrm{T}} \in \mathbb{R}^n$ é a matriz das características espaço-temporais (entradas) e $\mathbf{t}_i = [t_{i1}, t_{i2}, ..., t_{im}]^{\mathrm{T}} \in \mathbb{R}^m$ é o vetor de alvo que representa os escores subjetivos (MOS), conforme Equações (64) e (65), respectivamente.

$$\mathbf{x_i} = \begin{bmatrix} A_1 & B_1 & Z_1 & TI_1 & \overline{MAD_1} & MADw_1 \\ A_2 & B_2 & Z_2 & TI_2 & \overline{MAD_2} & MADw_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_i & B_i & Z_i & TI_i & \overline{MAD_i} & MADw_i \end{bmatrix}_{i\times 6}, \tag{64}$$

$$\mathbf{t_i} = \begin{bmatrix} MOS_1 \\ MOS_2 \\ \vdots \\ MOS_i \end{bmatrix}_{i\times 1}, \tag{65}$$

em que $n = 6$ e $m = 1$, logo $n > m$.

Assim, uma rede neural SLFN, com uma camada oculta e uma função de ativação $g(x)$, é matematicamente modelada conforme expressão a seguir (HUANG *et al.*, 2006).

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i\left(\mathbf{x}_j\right) = \sum_{i=1}^{\tilde{N}} \beta_i g\left(\mathbf{w}_i \cdot \mathbf{x}_j + b_i\right) = \mathbf{o}_j, \tag{66}$$

em que $\tilde{N}$ é o número de neurônios na camada oculta, $j = 1,...,N$ e $\mathbf{w}_i = [w_{i1}, w_{i2}, ..., w_{in}]^T$ é o vetor de peso que conecta o i-ésimo nó oculto com os nós de entrada, $\beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T$ é o vetor de pesos que conecta o i-ésimo nó oculto aos nós de saída, b_i é a polarização do i-ésimo nó oculto, $\mathbf{o}_j$ é o vetor de saída (escore objetivo) e $(\mathbf{w}_i \cdot \mathbf{x}_j)$ é o produto interno entre $\mathbf{w}_i$ e $\mathbf{x}_j$, conforme ilustra a Figura 12.

Segundo Huang *et al.* (2006), uma rede SLFN com $\tilde{N}$ neurônios na camada oculta e função de ativação $g(x)$ pode aproximar as N amostras por uma norma euclidiana livre de erros, tal que $\sum_{j=1}^{\tilde{N}} \|\mathbf{o}_j - \mathbf{t}_j\| = 0$, *i.e.*, existem β_i, $\mathbf{w}_i$ e b_i que satisfaçam a Equação (67).

$$\sum_{i=1}^{\tilde{N}} \beta_i g\left(\mathbf{w}_i \cdot \mathbf{x}_j + b_i\right) = \mathbf{t}_j, \quad j = 1, \ldots, N. \tag{67}$$

Entretanto, na prática, isso não ocorre, pois é razoável que sempre exista algum erro, por menor que seja, na predição de qualidade do vídeo, por exemplo. Logo, podem existir parâmetros β_i, $\mathbf{w}_i$ e b_i que conduzam a um erro próximo de zero, *i.e.*, nesta tese o erro da predição de qualidade e a Equação (67) são reinterpretados, respectivamente, como

$$\sum_{j=1}^{\tilde{N}} \|\mathbf{o}_j - \mathbf{t}_j\| \to 0 \tag{68}$$

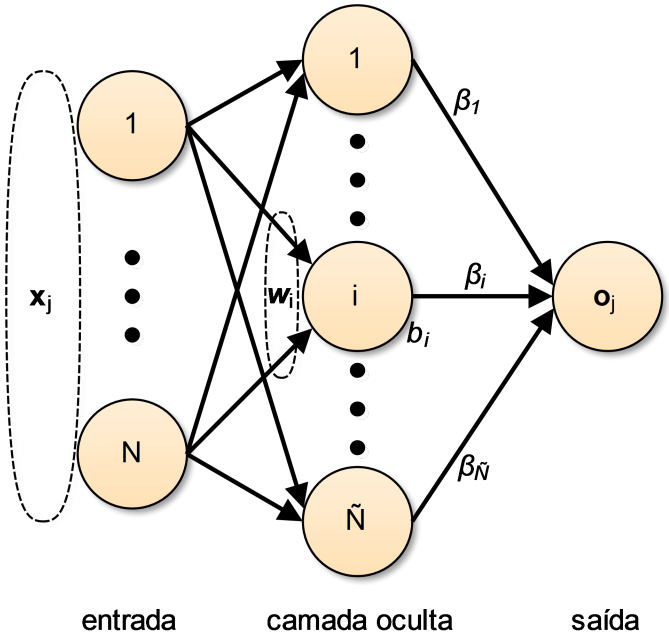

Figura 12: Arquitetura de rede SLFN.

Fonte: Autoria própria.

e

$$\sum_{i=1}^{\tilde{N}} \beta_i g\left(\mathbf{w}_i \cdot \mathbf{x}_j + b_i\right) \rightarrow \mathbf{t}_j, \quad j = 1, \ldots, N, \tag{69}$$

quando há uma forte correlação entre os escores objetivos e subjetivos, *i.e.*, $\mathbf{o}_j \rightarrow \mathbf{t}_j$.

A atribuição aleatória dos pesos dados às entradas ($\mathbf{w}_i$) e às polarizações (b_i) busca a solução do sistema linear da Equação (70), empregando o método dos mínimos quadrados que é baseado no gradiente. As N equações podem ser reescritas compactamente, envolvendo a matriz de saída da camada oculta $\mathbf{H}$ e o vetor de alvo $\mathbf{T}$, *i.e.*, o vetor dos escores subjetivos (MOS) (HUANG; BABRI, 1998; HUANG, 2003; HUANG *et al.*, 2006).

$$\mathbf{H}\beta = \mathbf{T}, \tag{70}$$

em que

$$\mathbf{H}(\mathbf{w}_1, \ldots, \mathbf{w}_{\tilde{N}}, b_1, \ldots, b_{\tilde{N}}, \mathbf{x}_1, \ldots, \mathbf{x}_N) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}, \tag{71}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{\tilde{N}} \end{bmatrix}_{\tilde{N}\times m} \quad \text{e} \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{t}_N^{\mathrm{T}} \end{bmatrix}_{N\times m} . \tag{72}$$

Os pesos $\mathbf{w}_i$ e $\mathbf{b}_i$ são atribuídos aleatoriamente, segundo uma distribuição uniforme (HUANG *et al.*, 2006). Além disso, considerando a solução do sistema linear (70), os pesos da camada de saída são calculados usando mínimos quadrados, *i.e.*,

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}, \tag{73}$$

em que $\mathbf{H}^{\dagger}$ é a matriz inversa generalizada de Moore-Penrose (MP), a qual possui as seguintes propriedades (RAO; MITRA, 1971; SERRE, 2002):

$$\begin{aligned} \mathbf{H}\mathbf{H}^{\dagger}\mathbf{H} &= \mathbf{H}, \\ \mathbf{H}^{\dagger}\mathbf{H}\mathbf{H}^{\dagger} &= \mathbf{H}^{\dagger}, \\ (\mathbf{H}\mathbf{H}^{\dagger})^{\mathrm{T}} &= \mathbf{H}\mathbf{H}^{\dagger}, \\ (\mathbf{H}^{\dagger}\mathbf{H})^{\mathrm{T}} &= \mathbf{H}^{\dagger}\mathbf{H}. \end{aligned} \tag{74}$$

Diversos métodos podem ser utilizados para calcular $\mathbf{H}^{\dagger}$, entre os quais (ORTEGA, 1987): a projeção ortogonal, método de ortogonalização, método iterativo e decomposição de valores singulares ou SVD (*Singular Value Decomposition*). Caso $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ seja não-singular, *i.e.*, $det(\mathbf{H}^{\mathrm{T}}\mathbf{H}) \neq 0$ e $\mathbf{H}^{\dagger} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}$, a solução por projeção ortogonal poderia ser usada (FERRARI; STENGEL, 2005). Entretanto, nem sempre $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ é não-singular, e neste caso pode-se calcular $\mathbf{H}^{\dagger}$ por SVD (HUANG *et al.*, 2006). O algoritmo ELM calcula $\mathbf{H}^{\dagger}$ usando a SVD com o método de Golub-Kahan-Reinsch (GOLUB; KAHAN, 1965; GOLUB; REINSCH, 1970) que compreende duas fases, sendo que na primeira emprega-se uma solução direta e na segunda utiliza-se uma solução iterativa. O método de Golub-Kahan-Reinsch é acurado, mas pode requerer alto custo computacional quando $\mathbf{H}$ é muito grande (LU *et al.*, 2012).

Logo, o algoritmo ELM realiza o treinamento a partir de um conjunto de entrada ou amostras de treinamento $\aleph = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbb{R}^n, \mathbf{t}_i \in \mathbb{R}^m, i = 1, ..., N\}$, uma função de ativação $g(x)$ e o número de nós da camada oculta $\tilde{N}$. O processo de treinamento é resumido conforme as seguintes etapas (HUANG *et al.*, 2006):

1. Geração aleatória dos pesos de entrada ($\mathbf{w}_i$) e polarização (b_i), com $i = 1, ..., N$;
2. Cálculo da matriz de saída da camada oculta, $\mathbf{H}$;
3. Cálculo dos pesos de saída β, conforme Equação (73).

3.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo foram apresentadas as bases teóricas e a metodologia relacionada ao desenvolvimento e validação dos métodos propostos. Em especial, neste capítulo foram descritas as bases de dados de vídeos utilizadas nesta tese, bem como foram apresentados o processamento entre os escores objetivos e subjetivos, as medidas estatísticas de desempenho, as características espaço-temporais e os algoritmos LM e ELM. O capítulo a seguir apresenta os métodos NR propostos para avaliação objetiva de qualidade de vídeo digital.

4 MÉTODOS PROPOSTOS

Este capítulo apresenta os métodos propostos para avaliação objetiva de qualidade de vídeo: NRVQA-LM e NRVQA-ELM/ELMtc. O primeiro é um modelo analítico sigmoidal e o segundo recorre a uma arquitetura de rede neural SLFN com o empego do algoritmo ELM. Ambos utilizam características espaço-temporais de vídeos com o mapeamento dos escores objetivos para a escala subjetiva que está relacionada à percepção do SVH.

A primeira abordagem NRVQA-LM otimiza os coeficientes ou parâmetros de seu modelo matemático, a partir de uma solução de mínimos quadrados usando o método iterativo de Levenberg-Marquardt. A segunda abordagem, com o método NRVQA-ELM e sua versão estendida NRVQA-ELMtc, emprega uma arquitetura de RNA SLFN em que os pesos e a polarização na camada oculta são atribuídos aleatoriamente e os pesos da camada de saída são determinados de forma analítica pelo cálculo da pseudo-inversa de Moore-Penrose.

As seções a seguir descrevem os métodos propostos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc.

4.1 NRVQA-LM

O método NRVQA-LM proposto tem origem em estudos desenvolvidos ao longo desta tese, envolvendo manipulações matemáticas entre as características espaciais e temporais descritas na Seção 3.3 e os escores subjetivos das bases de dados LIVE (SESHADRINATHAN *et al.*, 2010) e IVP (LI; MA, 2012). A métrica NRVQA-LM é determinada por uma abordagem analítica baseada em um modelo sigmoidal que incorpora três características espaciais e três temporais descritas na Seção 3.3. Além disso, funções sigmoidais apresentam comportamento monotônico relatado em diversos trabalhos relacionados a problemas de avaliação de qualidade de imagem e vídeo (SAZZAD *et al.*, 2008; KAWANO *et al.*, 2010; KEIMEL *et al.*, 2011a, 2011b). A fase de treinamento compreende o mapeamento entre as entradas $(A, B, Z, \mathrm{TI}, \overline{\mathrm{MAD}}, \mathrm{MADw})$ e os escores subjetivos (DMOS), a partir de uma solução de mínimos quadrados. As expressões (75), (76) e (77) definem os métodos espacial, temporal e

espaço-temporal proposto, respectivamente.

$$NRVQA\text{-}LMe = \frac{1}{1+e^{(\beta_1 B+\beta_2 Z+\beta_3 A+\beta_7)}}, \tag{75}$$

$$NRVQA\text{-}LMt = \frac{1}{1+e^{(\beta_4 TI+\beta_5 \overline{MAD}+\beta_6 MADw+\beta_7)}}, \tag{76}$$

$$NRVQA\text{-}LM = \frac{1}{1+e^{(\beta_1 B+\beta_2 Z+\beta_3 A+\beta_4 TI+\beta_5 \overline{MAD}+\beta_6 MADw+\beta_7)}}, \tag{77}$$

em que os parâmetros β_1 a β_7 são otimizados pelo método iterativo LM (LEVENBERG, 1944; MARQUARDT, 1963; MORÉ, 1977) descrito na Seção 3.4 e relatado na literatura em diversos trabalhos relacionados à avaliação de qualidade de imagem e vídeo que o empregam na solução de problemas de mínimos quadrados (WANG *et al.*, 2002; RIES *et al.*, 2006; ENGELKE; ZEPERNICK, 2007a; KEIMEL *et al.*, 2009; BRANDÃO; QUELUZ, 2010; SHAHID *et al.*, 2011; CALYAM *et al.*, 2012; KIPLI *et al.*, 2012).

A Figura 13 apresenta o diagrama de blocos do método proposto, conforme Equação (77). A fase de treinamento compreende a seleção de um determinado conjunto de vídeo, *e.g.*, baseado no tipo de artefato ou codificação, bem como os seus escores subjetivos normalizados no intervalo $[0,1]$, os parâmetros β iniciais (tipicamente, na primeira iteração $\beta_{1,\ldots,7} = 0$) e as seis características espaço-temporais $(A, B, Z, \mathrm{TI}, \overline{\mathrm{MAD}}, \mathrm{MADw})$. A fase de teste utiliza o modelo proposto na Equação (77) com os parâmetros β otimizados pelo algoritmo LM na fase de treinamento e, em seguida é gerado um escore de qualidade para a sequência de vídeo selecionada.

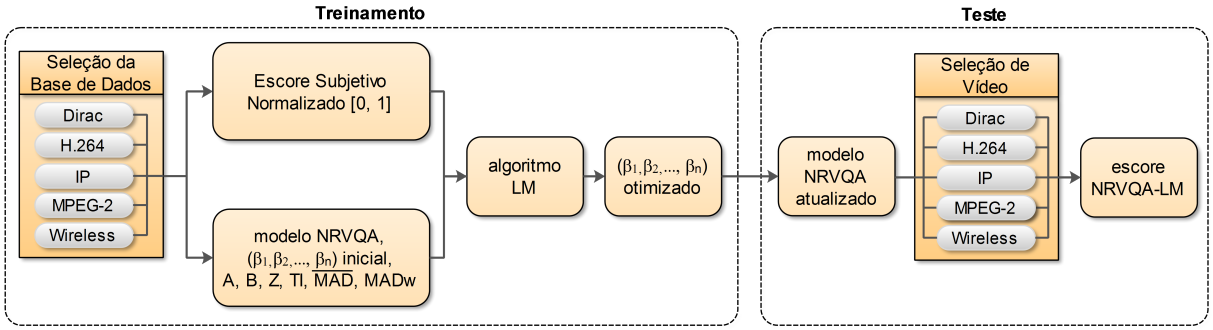

Figura 13: Diagrama de blocos do método NRVQA-LM com o treinamento dos parâmetros β pelo método iterativo LM.

Fonte: Adaptado de Silva e Pohl (2012).

A Figura 14 compara, como exemplo, o desempenho entre as componentes espacial (75), temporal (76) e espaço-temporal do método proposto (77) e a DMOS normalizada (DMOSn) no intervalo $[0,1]$ para o conteúdo em MPEG-2 da base de dados IVP. Conforme

inspeção visual da Figura 14 é possível observar que em algumas sequências de vídeo, *e.g.*, nas amostras 5, 6, 12, 15, 20 e 21, o método proposto na Equação (77) apresenta maior correlação com a medida subjetiva (DMOSn).

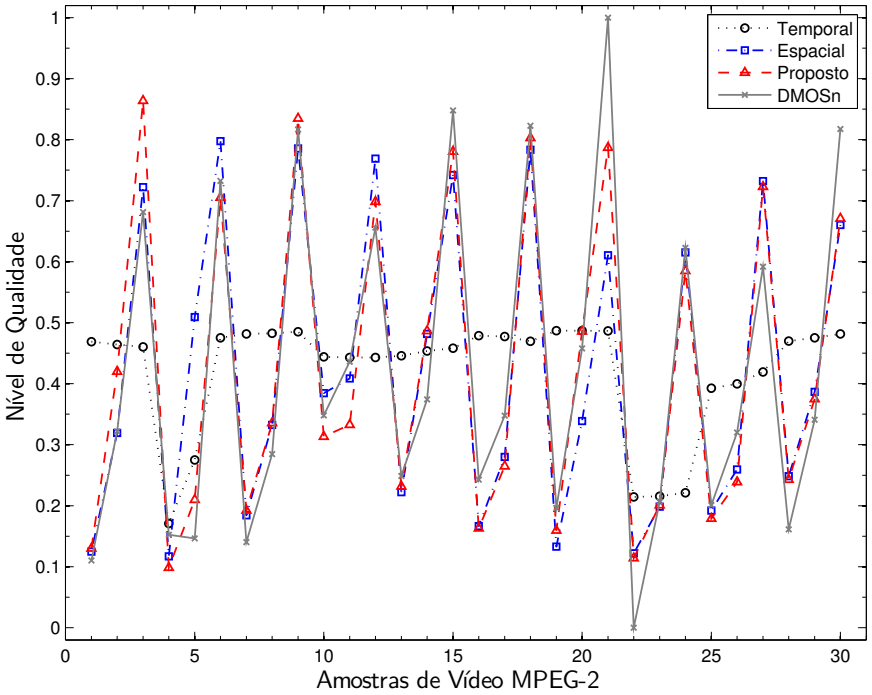

Figura 14: Comparação entre os modelos temporal (76), espacial (75), espaço-temporal proposto (77) e a DMOSn para vídeos em MPEG-2 (base de dados IVP).

A Tabela 2 mostra o desempenho do método proposto em relação à DMOS com uma correlação de Pearson (acurácia) igual a 0,9473, enquanto que as componentes espacial e temporal apresentam uma acurácia de 0,8906 e 0,3544, respectivamente. Além disso, os parâmetros β_1 a β_7 obtidos pelo algoritmo LM para o conteúdo e métodos comparados na Figura 14 também estão descritos na Tabela 2.

A otimização dos coeficientes β_1 e β_2 para os métodos espacial e proposto apresentam valores próximos no exemplo da Figura 14, embora a acurácia (PLCC) do modelo proposto seja

Tabela 2: Comparação da acurácia (PLCC) entre os modelos espacial, temporal e espaço-temporal e os parâmetros β otimizados pelo método LM.

Modelo	PLCC	Parâmetros						
		β_1	β_2	β_3	β_4	β_5	β_6	β_7
Espacial	0,8906	-0,4015	41,6746	-0,3232				-0,5774
Temporal	0,3544				-0,0289	-0,0653	4,6533	-0,8901
Proposto	**0,9473**	-0,3922	41,9226	-0,1441	0,0223	-0,5875	9,1590	-2,4752

maior, conforme mostra a Tabela 2, devido à combinação das características espaciais e temporais. Logo, este exemplo mostra que a combinação destas características é justificada, devido à acurácia do modelo proposto. Além disso, o modelo sigmoidal proposto na Equação (77) é influenciado pelos parâmetros β que estão intrinsecamente relacionados com o conteúdo da base de dados escolhida na fase de treinamento.

4.2 NRVQA-ELM

A escolha do algoritmo ELM para solução de problemas de avaliação de qualidade de vídeo é motivada tanto pelo desempenho quanto pelo tempo de processamento relatados na literatura em diversos tipos de problemas (HUANG *et al.*, 2004; HUANG; SIEW, 2004; LI *et al.*, 2005; ZHU *et al.*, 2005; HUANG *et al.*, 2006; LIANG *et al.*, 2006a, 2006b; HUANG; CHEN, 2008; HUANG *et al.*, 2008, 2010), dentre os quais a avaliação objetiva de qualidade de imagens (SURESH *et al.*, 2009). Além disso, até o momento da escrita da tese não foi encontrado na literatura nenhum trabalho que tenha citado o uso do algoritmo ELM em problemas de avaliação objetiva de qualidade de vídeo sem referência.

O algoritmo ELM em sua versão original não garante que os parâmetros β, $\mathbf{w}$ e b conduzam ao menor erro possível entre o escore objetivo $(\mathbf{o}_j)$ e o alvo $(\mathbf{t}_j)$ com $\tilde{N}$ neurônios na camada oculta e $j = 1,\ldots,N$, em que N é o número de amostras de vídeo. Assim, a tese traz uma contribuição neste sentido. Para tanto, além de propor o uso do algoritmo ELM descrito na Seção 3.5 para problemas NRVQA, a seguir também é proposta uma versão iterativa do algoritmo ELM que implementa um simples critério de parada, a fim de que seja encontrado o menor erro quadrático possível entre $\mathbf{o}_j$ e $\mathbf{t}_j$ para um conjunto de parâmetros escolhidos. Logo, dentre k iterações, o método proposto a seguir encontra os melhores parâmetros da RNA para que a correlação entre os escores objetivos e subjetivos seja a maior possível.

4.2.1 NRVQA-ELMtc

O algoritmo ELM quando opera com um número de amostras de treinamento maior do que o número de neurônios na camada oculta, *i.e.*, $N > \tilde{N}$, apresenta apenas uma etapa no cálculo dos pesos da camada de saída (HUANG *et al.*, 2006). Dessa forma, a solução da pseudo-inversa de MP na Equação (73) pode ser resolvida por SVD com o método de Golub-Kahan-Reinsch (GOLUB; KAHAN, 1965; GOLUB; REINSCH, 1970). Assim, como já fora mencionado na Seção 3.5.1, este método embora seja bastante acurado, pode exigir alto custo computacional quando aplicado em matrizes de alta dimensionalidade (LU *et al.*,

2012). Portanto, nesta tese adota-se uma representação reduzida do vetor de características espaço-temporais, pois para cada amostra de vídeo i são utilizadas apenas seis variáveis, *i.e.*, $\mathbf{x_i} = \{\mathbf{A_i}, \mathbf{B_i}, \mathbf{Z_i}, \mathbf{TI_i}, \overline{\mathbf{MAD_i}}, \mathbf{MADw_i}\}$.

O método NRVQA-ELM com critério de parada ou NRVQA-ELMtc seleciona os melhores parâmetros da RNA, *i.e.*, $\mathbf{w}$, b e β, que conduzam ao menor RMSE entre a MOS e MOSp em k iterações. Esta versão estendida do algoritmo ELM busca uma minimização do erro quadrático entre os escores objetivos e subjetivos, *i.e.*, $\sum_{j=1}^{\tilde{N}} \|\mathbf{o}_j - \mathbf{t}_j\| \to 0$, para satisfazer a Equação (69).

O Algoritmo 1 descreve o método proposto com um vetor de entrada ou *input* (contendo j amostras) que compreende as características espaço-temporais $\mathbf{x_j}$ (descritas na Seção 3.3) e os alvos ($\mathbf{t_j}$) ou escores subjetivos (MOS) normalizados no intervalo $[0,1]$, bem como outros parâmetros de configuração da RNA, tais como a função de ativação, a tolerância do RMSE (RMSE_{tol}), o número máximo de iterações (itmax) e o número de neurônios na camada oculta ($\tilde{N}$). Caso o RMSE entre a MOS e a MOSp seja menor do que a tolerância (RMSE_{tol}), *i.e.*, $\text{RMSE} < \text{RMSE}_{tol}$, o algoritmo cessa a busca pelos melhores parâmetros de treinamento β, $\mathbf{w}$ e b. Entretanto, enquanto $\text{RMSE} \geq \text{RMSE}_{tol}$, o algoritmo continuará a busca até completar o número máximo de iterações ($k = \text{itmax}$). Na sequência, ocorre a seleção dos parâmetros β_p, $\mathbf{w}_p$ e b_p associados à iteração p (linha 13 do Algoritmo 1) que representa o menor valor de RMSE entre a MOS e a MOSp, com $p \in [1,\ldots,k]$. O vetor de saída (linhas 8 e 14 do Algoritmo 1) ou *output* ($\mathbf{o_j}$) corresponde aos escores objetivos (MOSp) mapeados na escala MOS, conforme Equação (30), normalizados no intervalo $[0,1]$ com o menor RMSE em k iterações.

Além disso, o método NRVQA-ELMtc faz com que os parâmetros $\mathbf{w}$, b e β correspondam ao ponto de mínimo no intervalo $[1,\ldots,k]$. Consequentemente, embora não exista garantia, este ponto de mínimo pode coincidir com um mínimo local ou, no melhor caso, este ponto pode coincidir com um mínimo global. A acurácia do método NRVQA-ELMtc está associada ao número máximo de iterações itmax. Logo, quanto maior for o seu valor, maior poderá ser a acurácia do método, bem como exigirá custo computacional adicional, em termos de memória e tempo de processamento. Nesta tese, foram utilizados os valores de RMSE_{tol} e itmax iguais a $5,0 \times 10^{-1}$ e $1,0 \times 10^{2}$, respectivamente.

A escolha de $\text{RMSE}_{tol} = 5,0 \times 10^{-1}$ é baseada nos resultados das Figuras 15, 16 e 17 que comparam as medidas de RMSE das métricas de referência completa PSNR, SSIM, MS-SSIM e sem referência JPEG-NR para as bases de dados LIVE e IVP, ambas divididas em três conjuntos de teste e do superconjunto S, dividido em blocos de teste, cuja formação desses grupos de teste está discutida em detalhes no Capítulo 5. A mediana (segundo quartil) do RMSE

Algoritmo 1: Algoritmo NRVQA-ELMtc.

Input:
características espaço-temporais: $\mathbf{x_j} = [\mathbf{A}, \mathbf{B}, \mathbf{Z}, \mathbf{TI}, \overline{\mathbf{MAD}}, \mathbf{MADw}]$
alvo: $\mathbf{t_j} = \mathbf{MOS}$, normalizada no intervalo $[0,1]$
função de ativação [radbas, sig, hardlim, sin]
tolerância do RMSE [$\text{RMSE}_{tol} = 5,0 \times 10^{-1}$]
número máximo de iterações [itmax = $1,0 \times 10^{2}$]
número de neurônios na camada oculta $\left[\tilde{N}\right]$

```
k ← 0
while k < itmax do
    k ← k + 1
    parâmetros ELM atuais: β_k, w_k e b_k
    MOSp_k ← ELM(β_k, w_k, b_k)
    RMSE_k ← RMSE entre MOS e MOSp_k
    if RMSE_k < RMSE_tol then
        MOSp ← MOSp_k
        break
    else
        continue
        if k = itmax then
            seleciona p com min(RMSE_{p=1,...,k})
            MOSp ← ELM(β_p, w_p, b_p)
        end
    end
end
```

Output: **MOSp** $(\mathbf{o_j})$ no intervalo $[0,1]$ com $j = 1, \ldots, N$ amostras

apresenta valores acima de 5 e 1 para base de dados LIVE e superconjunto *S*, respectivamente. A base de dados IVP expressa os menores valores de RMSE dentre as bases de dados utilizadas nesta tese, conforme a inspeção visual da Figura 16, com a mediana próxima a 0,5 nos conteúdos em Dirac, H.264 e MPEG-2, justificando o uso desse valor pela variável RMSE_{tol} no Algoritmo 1, baseado nos resultados experimentais obtidos com as métricas FR.

Quanto ao número de neurônios na camada oculta $(\tilde{N})$ há diversos trabalhos na literatura que se dedicam a este assunto, dentre os quais Panchal *et al.* (2011), González *et al.* (2011), Mao *et al.* (2012) que sugerem algumas abordagens: (i) o número de neurônios na camada oculta deve estar entre o tamanho da camada de entrada e o tamanho da camada de saída; (ii) o número de neurônios na camada oculta deve ser 2/3 do tamanho da camada de entrada mais o tamanho da camada de saída; (iii) o número de neurônios na camada oculta deve ser menor do que duas vezes o tamanho da camada de entrada. Nesta tese foram adotadas abordagens distintas quanto ao número de neurônios na camada oculta $(\tilde{N})$, conforme será detalhado no Capítulo 5.

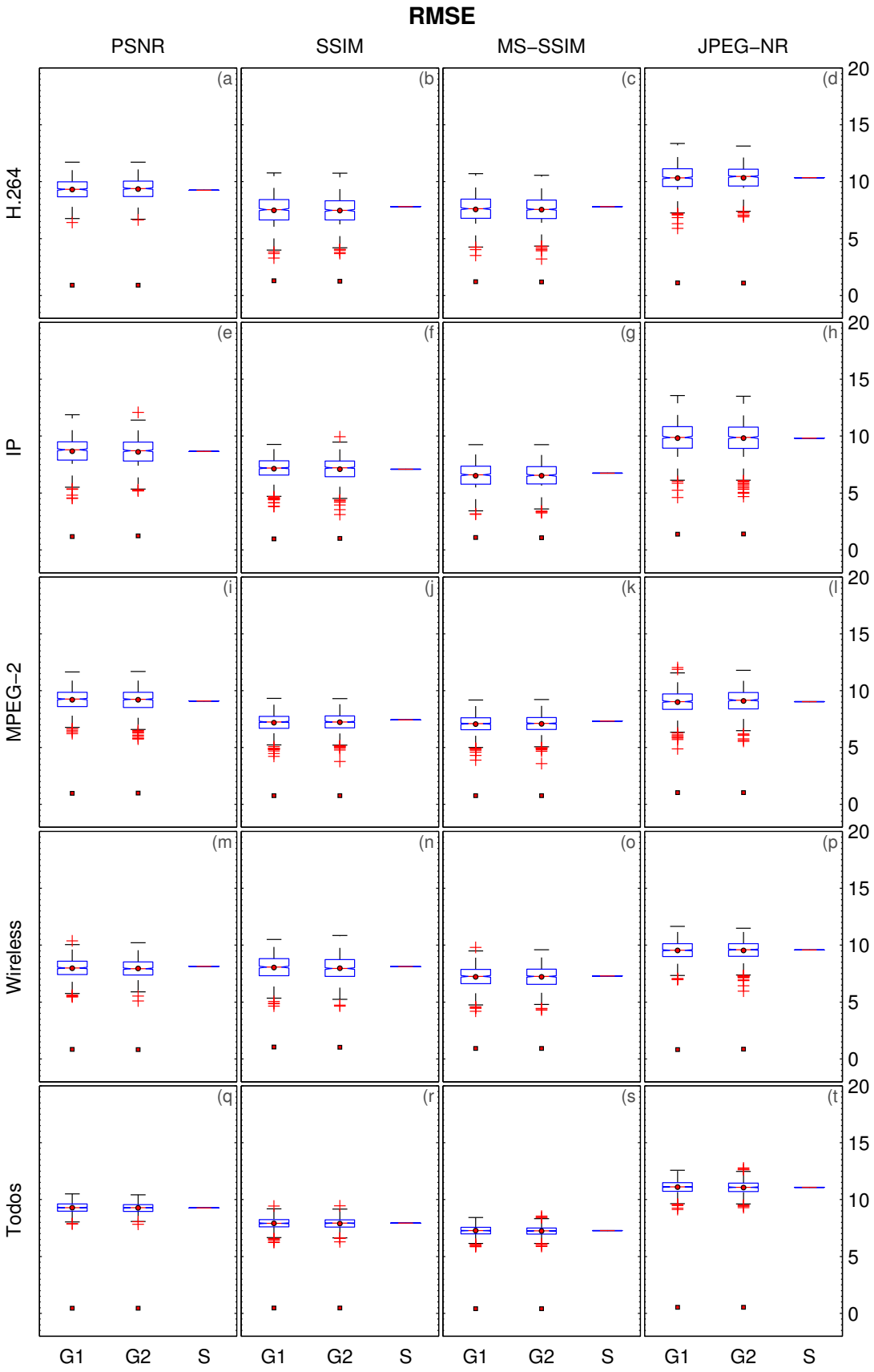

Figura 15: Comparação das medidas de RMSE entre as métricas PSNR, SSIM, MS-SSIM e JPEG-NR para a base de dados LIVE.

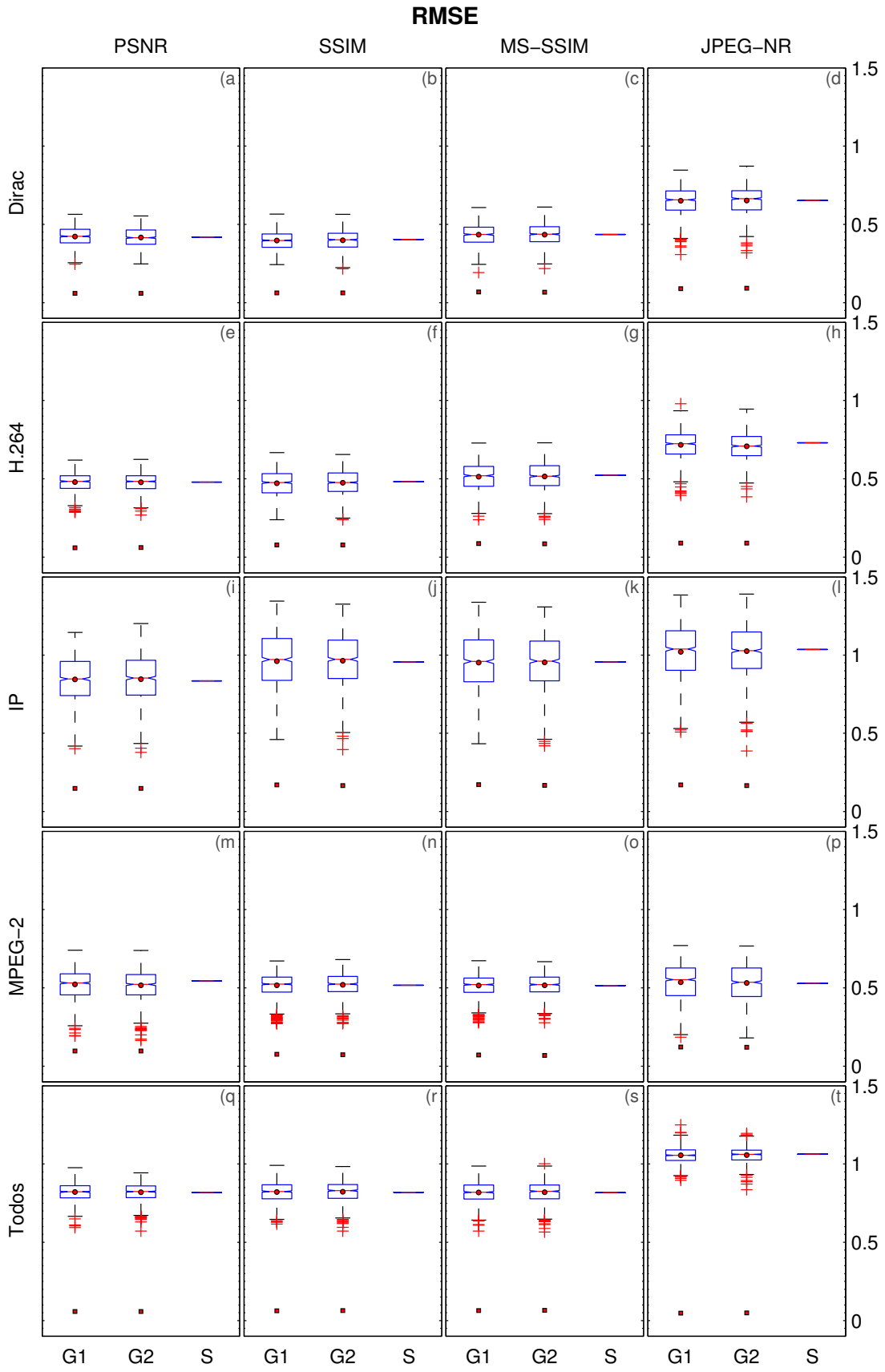

Figura 16: Comparação das medidas de RMSE entre as métricas PSNR, SSIM, MS-SSIM e JPEG-NR para a base de dados IVP.

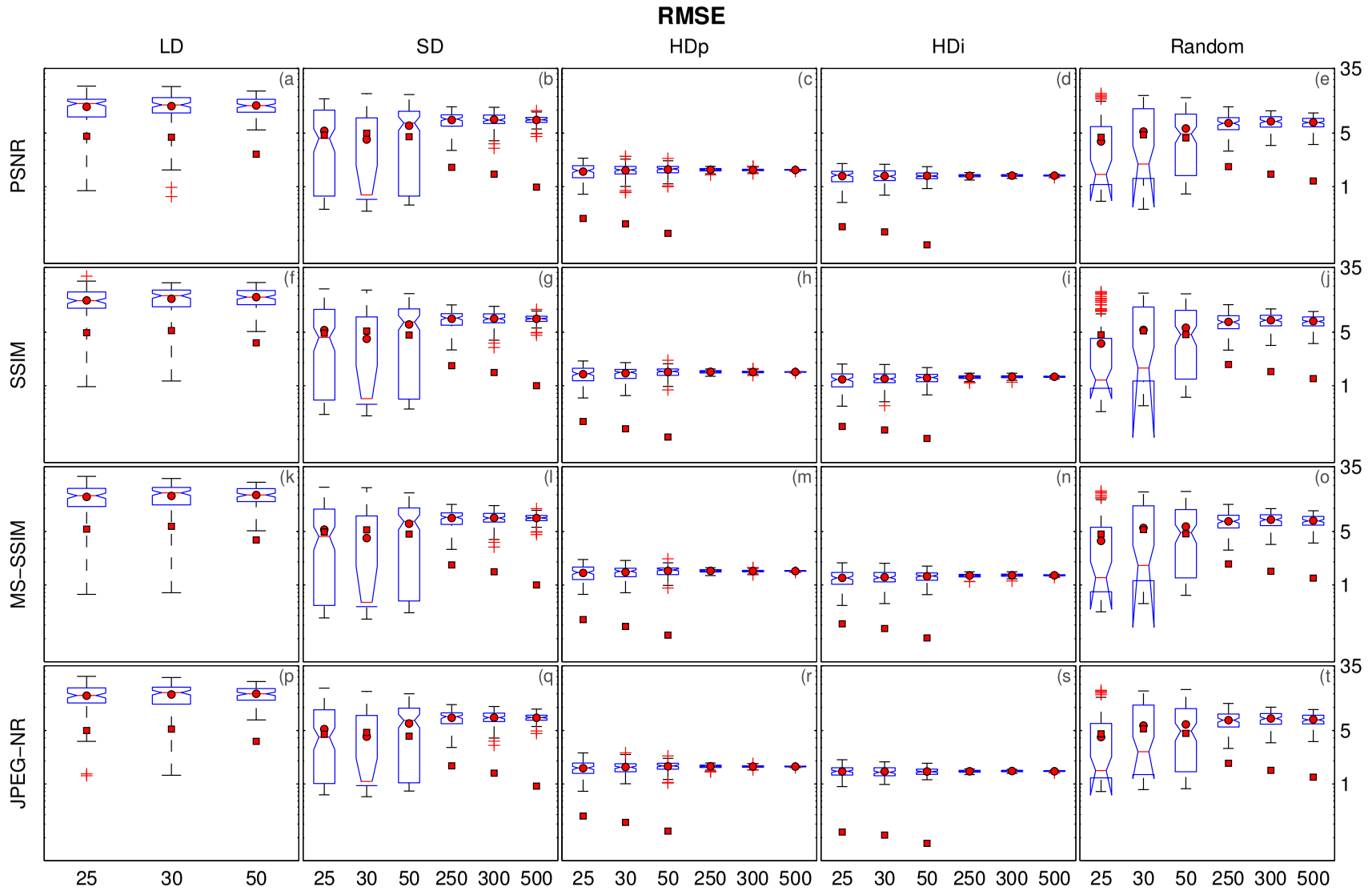

Figura 17: Comparação das medidas de RMSE entre as métricas PSNR, SSIM, MS-SSIM e JPEG-NR para o superconjunto S dividido em blocos de teste para as resoluções LD, SD, HDp, HDi e Todos.

4.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo apresentou e discutiu os detalhes de implementação dos métodos propostos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc. O método NRVQA-LM é fortemente dependente dos parâmetros β da Equação (77), os quais têm relação com o tipo de artefato de vídeo presente no treinamento, conforme será discutido no próximo capítulo. O algoritmo ELM clássico depende apenas dos parâmetros da RNA e da função de ativação escolhida. Dessa forma, o método NRVQA-ELMtc implementa um simples critério de parada, conforme o Algoritmo 1, que embora não seja direcionado, devido à geração aleatória dos parâmetros da RNA, apresenta maior desempenho do que a versão NRVQA-ELM, conforme mostram os resultados experimentais. O próximo capítulo descreve os experimentos realizados e a metodologia empregada no processo de validação cruzada, bem como discute os resultados obtidos.

5 RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados os resultados de desempenho dos métodos propostos. As medidas estatísticas, conforme descrição na Seção 3.2.1, determinam o grau de associação entre os escores objetivos e subjetivos. Alguns resultados são expressos por diagrama de caixa ou *box-plot*, que sintetiza estatisticamente a distribuição de dados. Detalhes acerca deste tipo de representação gráfica podem ser encontrados no Capítulo 3.

Nesta tese são conduzidos quatro experimentos com validação cruzada envolvendo os métodos propostos (NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc) e as métricas PSNR, SSIM, MS-SSIM e JPEG-NR, conforme a Figura 18 que ilustra esquematicamente os experimentos com validação cruzada.

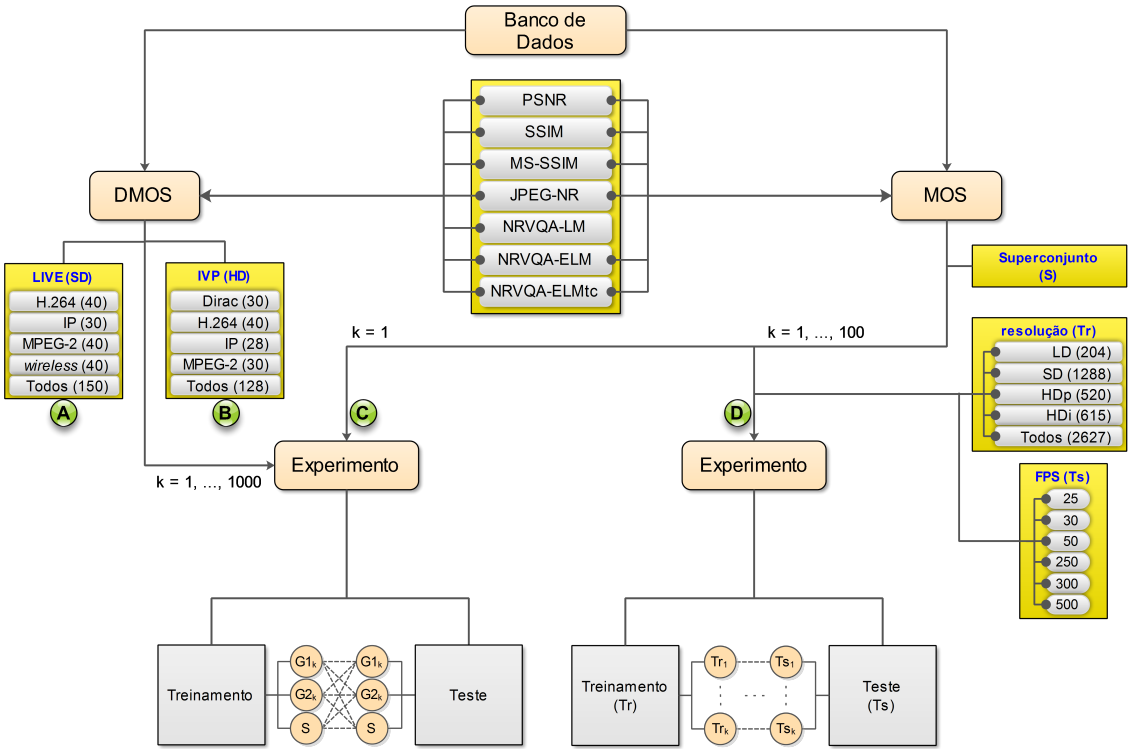

Figura 18: Diagrama esquemático dos experimentos A, B, C e D com validação cruzada.

Fonte: Autoria própria.

5.1 ARRANJO EXPERIMENTAL E O PROCESSO DE VALIDAÇÃO CRUZADA

Cada experimento da Figura 18 é composto de k repetições distintas. Os experimentos A e B usam a DMOS como alvo, enquanto que os experimentos C e D utilizam a MOS. Os experimentos A-B, C e D possuem uma distribuição com k igual a 1000, 1 e 100, respectivamente. Os pares treinamento-teste dos experimentos A, B e C são formados a partir de dois conjuntos disjuntos $G1_k$ e $G2_k$ e o conjunto S composto pela união desses, enquanto que no experimento D as amostras de teste são excluídas do treinamento. O processo de validação cruzada nos experimentos A-B e C são realizados com uma versão do método de resistência ou *holdout* (KOHAVI, 1995; HAYKIN, 1999), cuja abordagem empregada no experimento C com $k = 1$ é amplamente utilizada na literatura (WANG *et al.*, 2002; MOHAMED *et al.*, 2002; BARLAND; SAADANE, 2005; JANOWSKI; ROMANIAK, 2010; KHAN *et al.*, 2010; GU *et al.*, 2012). Quanto à validação no experimento D foi utilizado o método de validação cruzada múltipla ou *K-fold* (KOHAVI, 1995; HAYKIN, 1999; HASTIE *et al.*, 2009), conforme a descrição da Tabela 3. O método *K-fold* também é amplamente relatado na literatura que trata de métodos objetivos para avaliação de qualidade de imagem e vídeo (TONG *et al.*, 2005; GASTALDO; ZUNINO, 2005; van Marais *et al.*, 2008; LAHOUHOU *et al.*, 2010; STAELENS *et al.*, 2010; LIU *et al.*, 2011; HERZOG *et al.*, 2012; DECHERCHI *et al.*, 2013).

Tabela 3: Métodos de validação cruzada utilizados nos experimentos A, B, C e D.

Experimento	Método
A-B	Versão do método de resistência com a combinação dos pares treinamento-teste $G1_k$, $G2_k$ e S, em que $k = 1, \ldots, 1000$, representa a formação de grupos distintos
C	Idem, entretanto com $k = 1$, *i.e.*, apenas a combinação entre os grupos $G1$, $G2$ e S
D	*Método K-fold* com $K = 100$

A Tabela 4 apresenta a variação entre nove pares treinamento-teste, formados a partir dos conjuntos $G1$, $G2$ e S. Assim, três destes pares são idênticos e dificilmente, em condições práticas, os padrões de treinamento e teste serão exatamente iguais. Logo, restam seis pares que poderiam ser encontrados em condições práticas, sobretudo os pares disjuntos $G1$-$G2$ e $G2$-$G1$. Além disso, a Tabela 4 justifica a utilização de todos estes conjuntos na comparação do desempenho de métricas FR e NR, e.g., os pares $G1$-$G1$, $G2$-$G2$ e S-S podem ser utilizados para verificar a precisão do método proposto quanto à acurácia durante o treinamento, enquanto que os demais pares podem ser efetivamente usados na fase de teste ou no processo de validação. Entretanto, a discussão de resultados neste capítulo se concentrará nos pares de treinamento-teste disjuntos, devido à natureza de suas aplicações práticas.

Tabela 4: Características dos conjuntos de treinamento-teste no processo de validação cruzada.

Treinamento-Teste	Características
G1-G1 G2-G2 S-S	▷ Padrões de treinamento e de teste idênticos ▷ Verificação da acurácia no treinamento ▷ Calibração do método proposto
G1-G2 G2-G1	▷ Pares disjuntos ▷ Padrão de teste não contido no treinamento
G1-S G2-S	▷ Padrão de teste parcialmente contido no treinamento
S-G1 S-G2	▷ Padrão de teste completamente contido no treinamento

O processador utilizado nos experimentos foi um Intel Core 2 Quad Q9550, 64 *bits*, 2,83 GHz, cache de 6.144 k*byte*s e memória de 8 G*byte*s. A distribuição Debian Wheezy GNU/Linux 7.0 (DEBIAN, 2012) foi a distribuição do SO (Sistema Operacional) Linux utilizada nos experimentos.

As seções a seguir descrevem em pormenores os experimentos e seus respectivos resultados em termos da distribuição F percentual (F_p) entre os métodos propostos e a métrica FR MS-SSIM e da acurácia (PLCC), bem como os tempos de treinamento e de teste. Além disso, no final deste capítulo, a Seção 5.6 apresenta uma síntese dos resultados experimentais. Para exemplificar o uso de outras medidas estatísticas de desempenho de métodos NRVQA, o Apêndice B apresenta resultados adicionais acerca das medidas RMSE, SROCC, R-quadrado, OR e MAE do método NRVQA-LM com a base de dados IVP (experimento B).

5.2 EXPERIMENTO A

Experimento realizado com a base de dados LIVE (768×432p) que contém escores subjetivos (DMOS) de 150 vídeos divididos em distorções de codificação H.264 e MPEG-2, bem como em artefatos de transmissão em rede IP e sem fio (*Wireless*) com 40, 40, 30 e 40 amostras, respectivamente. Este experimento foi repetido mil vezes $(k = 1, \ldots, 1000)$ com sequências diferentes tanto para $G1_k$, quanto para $G2_k$, cuja formação do par treinamento-teste obedeceu à alternância entre todos os grupos, conforme a Figura 18. As métricas PSNR, SSIM, MS-SSIM, JPEG-NR e os métodos propostos NRVQA-LM, NRVQA-ELM e NRVQA-ELM são comparados no processo de validação com uma versão do método de resistência (KOHAVI, 1995; HAYKIN, 1999). A Equação (78) mostra a relação entre os conjuntos $G1_k$, $G2_k$ e S. Logo, a união entre $G1_k$ e $G2_k$ para qualquer k resulta em S, embora não exista interseção entre $G1_k$ e $G2_k$.

$$\begin{aligned} &\{G1_k \subset S\}, \\ &\{G2_k \subset S\}, \\ &S = \{G1_k \cup G2_k\}, \\ &\{G1_k \cap G2_k\} = \varnothing. \end{aligned} \tag{78}$$

Nos métodos NRVQA-ELM e NRVQA-ELMtc foi utilizado $\tilde{N} = \text{round}(\frac{2}{3}N)$, em que N é o número de amostras em cada categoria e $\tilde{N}$ é o número de neurônios na camada oculta. A Tabela 5 mostra a quantidade de amostras e o número de neurônios utilizados na camada oculta para cada conteúdo da base de dados LIVE (experimento A).

Tabela 5: Número de amostras e neurônios na camada oculta utilizados no experimento A.

Database	Conteúdo	Conjunto	Amostras (N)	$\tilde{N} = \text{round}\left(\frac{2}{3}N\right)$
LIVE	H.264	$G1/G2$	20	13
		S	40	27
	IP	$G1/G2$	15	10
		S	30	20
	MPEG-2	$G1/G2$	20	13
		S	40	27
	Wireless	$G1/G2$	20	13
		S	40	27
	Todos	$G1/G2$	75	50
		S	150	100

A Figura 19 compara a acurácia (PLCC) do método NRVQA-ELMtc em função do número de neurônios na camada oculta com as funções de ativação hardlim, radbas, sig e sin, conforme as Figuras 19-a a 19-d, respectivamente. O número de neurônios na camada oculta igual a $\tilde{N} = \text{round}(\frac{2}{3}N)$ é justificado pelos resultados da Figura 19. Quando o grupo de teste está contido no treinamento, *e.g.*, S como treinamento e $G1$ como teste aplicado ao conteúdo "Todos", observa-se uma convergência do PLCC próximo a 1 quando são usadas as funções de ativação sig e sin para $\tilde{N} = 100$, conforme inspeção visual das Figuras 19-c e 19-d, respectivamente.

A Figura 20 mostra a variação dos coeficientes β_1 a β_7 para os conjuntos de treinamento $G1$, $G2$ e S relacionados aos conteúdos da base de dados LIVE. A distribuição da acurácia entre as métricas PSNR, SSIM, MS-SSIM, JPEG-NR e o método proposto NRVQA-LM para cada categoria da base de dados LIVE é comparada na Figura 21. O método NRVQA-LM apresenta melhor desempenho para os conteúdos H.264 e MPEG-2, conforme inspeção visual dos intervalos interquartílicos das Figuras 21-e e 21-o, respectivamente. O intervalo interquartílico (amplitude interquartílica ente o primeiro e terceiro quartis) nos resultados com o método NRVQA-LM está comprimido no conteúdo em MPEG-2 com valores de acurácia acima de 0,8, conforme mostra a Figura 21-o, *i.e.*, o desempenho desse método é superior às métricas FR, mesmo nos pares treinamento-teste disjuntos $G1$-$G2$ e $G2$-$G1$. Os artefatos de compressão H.264 e MPEG-2 causam efeitos altamente estruturados nas amostras de vídeo (WU *et al.*, 2007) que são identificados pelo método NRVQA-LM. Esta pode ser a razão pela qual o método proposto tenha apresentado melhor desempenho. Entretanto, o mesmo apresenta menor acurácia quando os conteúdos estão embaralhados (categoria Todos), conforme a Figura 21-y, devido ao treinamento dos coeficientes β que estão associados com as características de um conteúdo específico de uma base de dados. Além disso, os artefatos para os conteúdos IP e *Wireless* foram gerados artificialmente e de uma maneira aleatória (SESHADRINATHAN *et al.*, 2010). Logo, estas condições reduzem a acurácia na predição de qualidade do método NRVQA-LM.

As Figuras 22 e 23 comparam o desempenho dos métodos propostos NRVQA-ELM e NRVQA-ELMtc em termos das funções de ativação, sendo que este último apresenta melhor desempenho, conforme compactação entre o primeiro e terceiro quartis, *i.e.*, menor amplitude interquartílica. Os resultados com as funções de ativação sin e sig são equivalentes, principalmente o método NRVQA-ELMtc, conforme a Figura 23. Exceto para os pares disjuntos $G1$-$G2$ e $G2$-$G1$, cujo desempenho dos métodos NRVQA-ELM e NRVQA-ELMtc, usando a função de ativação sin, são superiores às métricas PSNR e JPEG-NR nas categorias H.264, IP e MPEG-2, conforme a mediana da acurácia nas Figuras 22 e 23.

A Tabela 6 apresenta a mediana da distribuição dos tempos de treinamento dos métodos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc para cada grupo de treinamento da base de dados LIVE, *i.e.*, os conjuntos de treinamento $G1$, $G2$ e S. Embora o método NRVQA-ELMtc tenha mostrado maior desempenho, este tende a exigir maior custo computacional, conforme inspeção visual da Tabela 6, *e.g.*, para os conteúdos H.264, MPEG-2 e "Todos" relativos ao conjunto de treinamento S. Nos conteúdos IP e *Wireless* relacionados ao conjunto de treinamento S, o método NRVQA-ELMtc consumiu um tempo de treinamento próximo àquele observado na versão NRVQA-ELM, devido ao critério de parada do Algoritmo 1 descrito na Seção 4.2.1, o qual pode requerer um número reduzido de iterações. A mediana aproximada da distribuição dos tempos de teste do método NRVQA-LM apresenta um valor de 12 ms, enquanto os métodos NRVQA-ELM e NRVQA-ELMtc apresentam um tempo de teste de 10 ms.

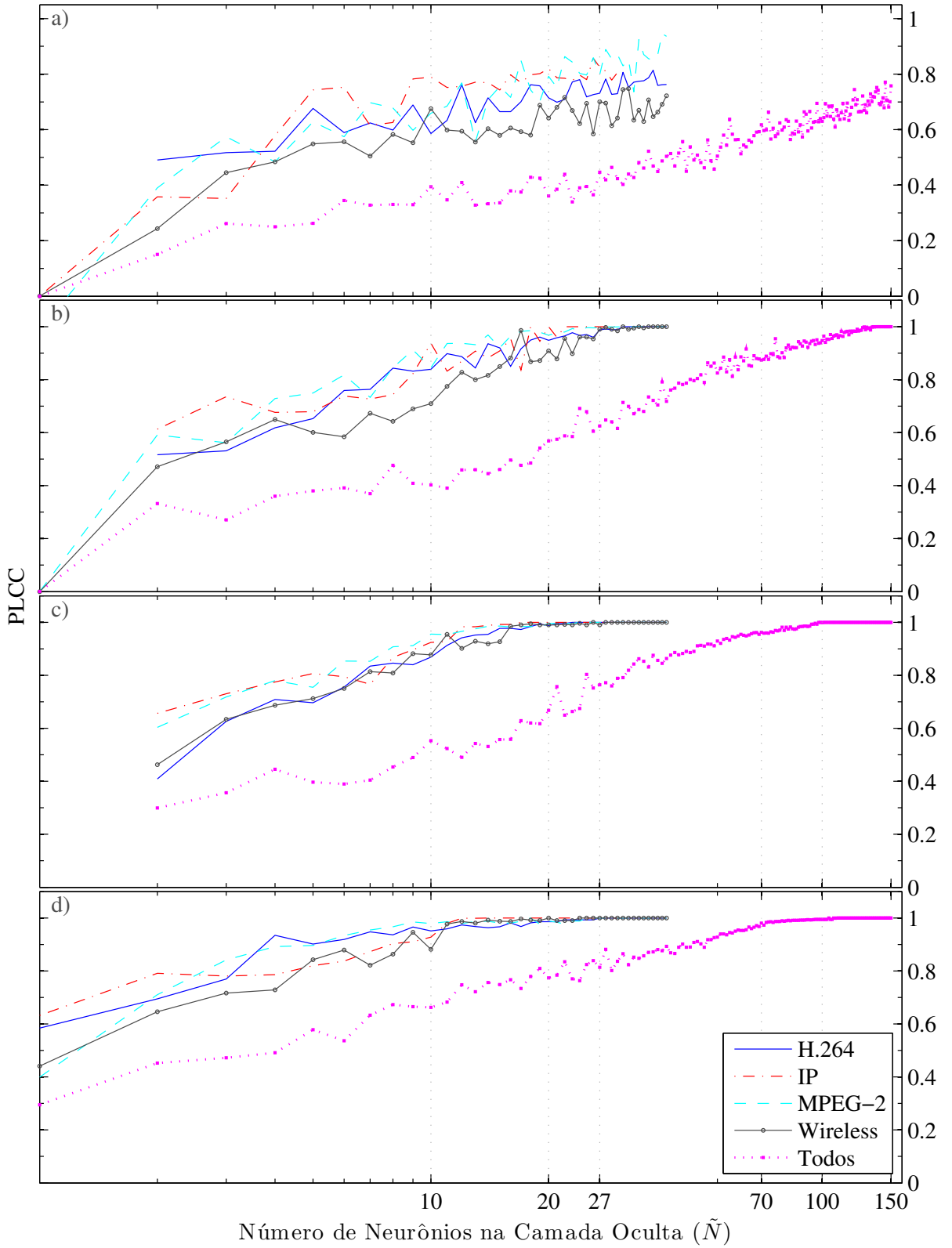

Figura 19: Comparação entre a acurácia e o número de neurônios ($\tilde{N}$) usando o método NRVQA-ELMtc, em que S é o treinamento e $G1$ é o teste com as funções de ativação: a) hardlim, b) radbas, c) sig e d) sin para os conteúdos do experimento A.

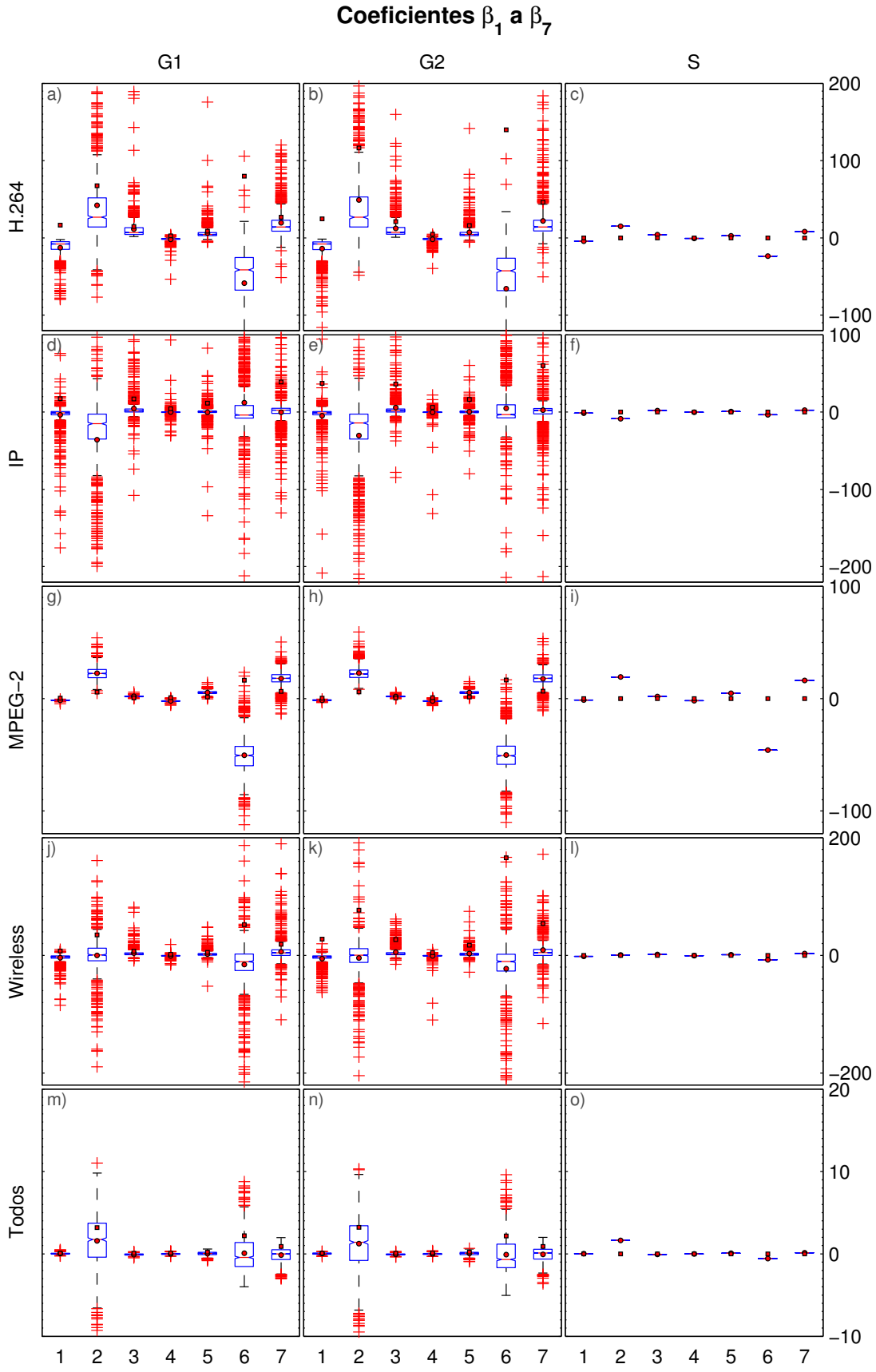

Figura 20: Coeficientes β_1 a β_7 do método NRVQA-LM nos grupos de treinamento $G1$, $G2$ e S para os conteúdos do experimento A.

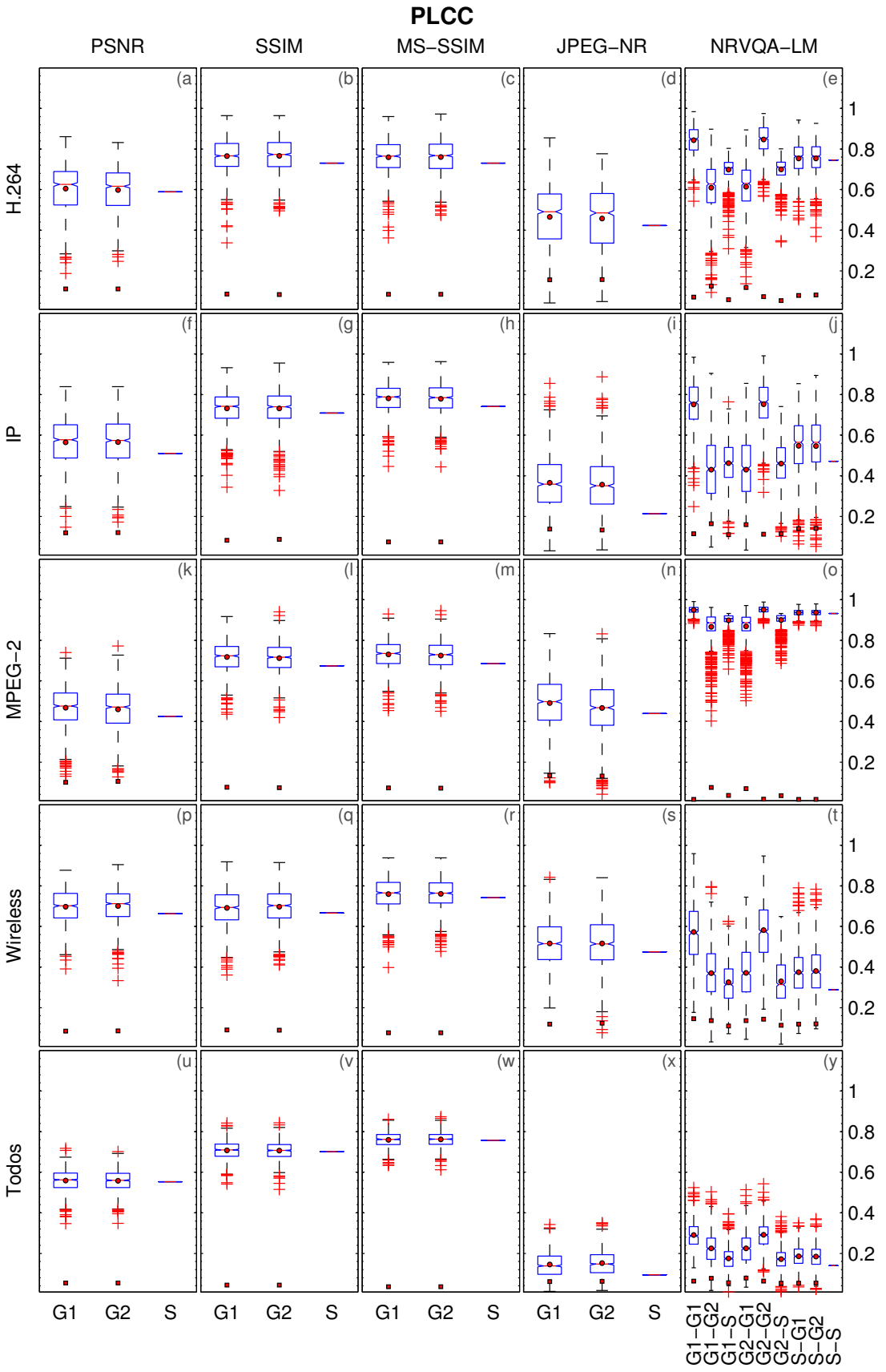

Figura 21: Comparação da acurácia (PLCC) entre as métricas PSNR, SSIM, MS-SSIM, JPEG-NR e NRVQA-LM para os conteúdos do experimento A.

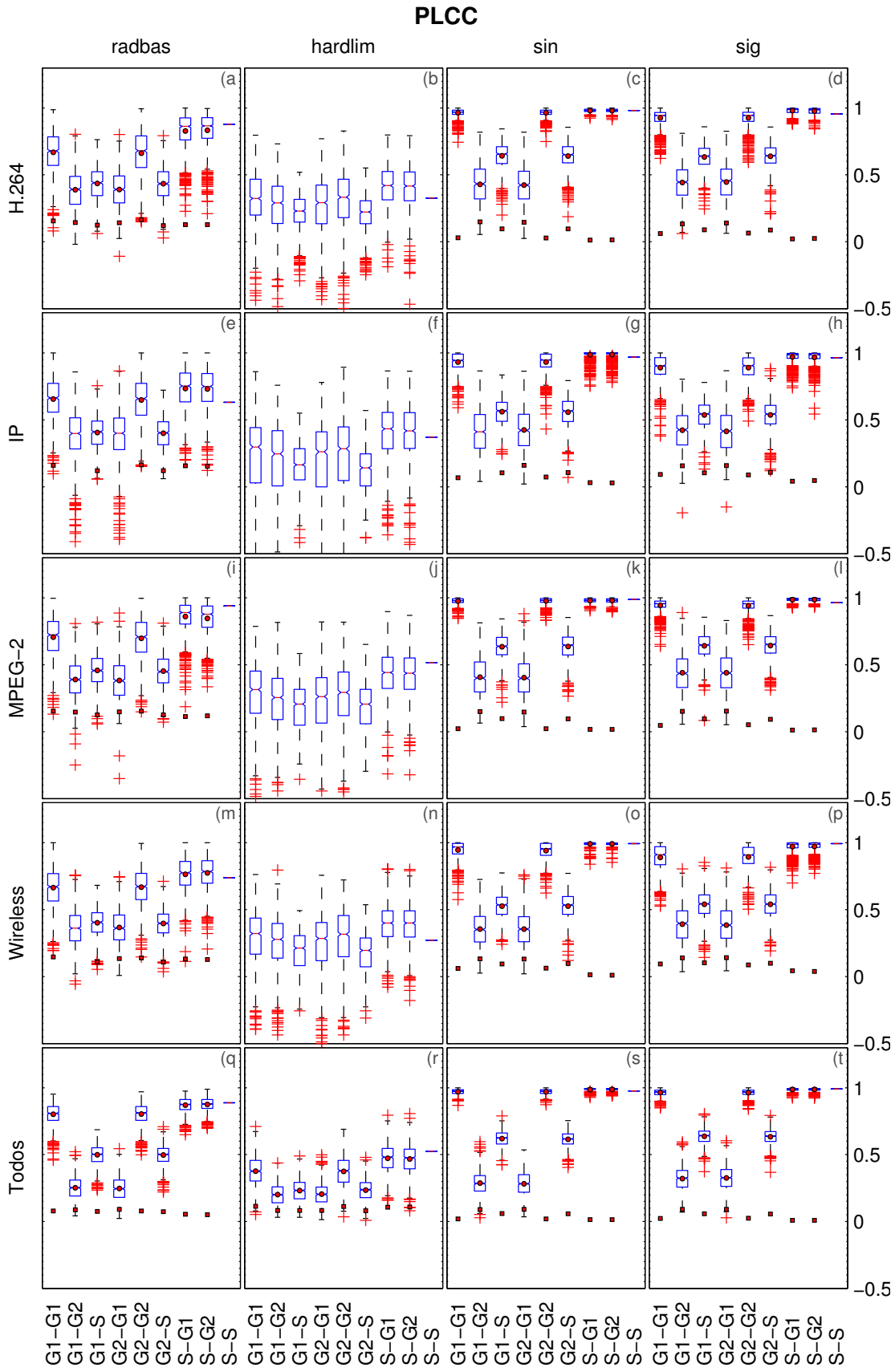

Figura 22: Comparação da acurácia (PLCC) do método NRVQA-ELM para os conteúdos do experimento A.

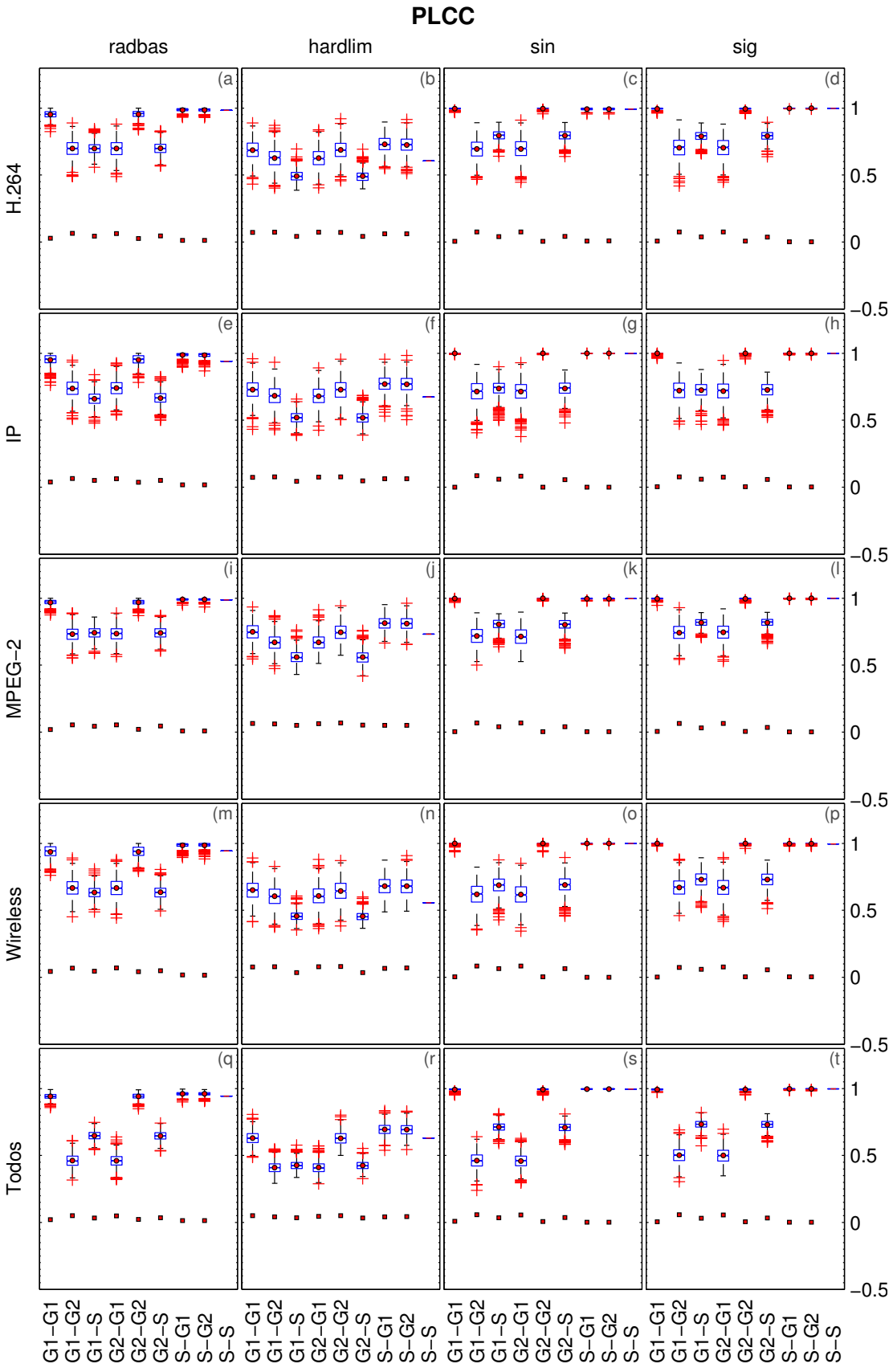

Figura 23: Comparação da acurácia (PLCC) do método NRVQA-ELMtc para os conteúdos do experimento A.

Tabela 6: Mediana do tempo de treinamento dos métodos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc para o experimento A.

Categoria	NRVQA	Conjunto de Treinamento (s)		
		*G*1	*G*2	*S*
H.264	LM	0,02	0,02	0,02
	ELM*	0,03	0,03	0,04
	ELMtc*	0,18	0,19	0,82
IP	LM	0,02	0,02	0,02
	ELM*	0,02	0,02	0,03
	ELMtc*	0,18	0,18	0,02
MPEG-2	LM	0,02	0,02	0,02
	ELM*	0,03	0,02	0,04
	ELMtc*	0,03	0,03	0,93
Wireless	LM	0,02	0,02	0,01
	ELM*	0,03	0,03	0,04
	ELMtc*	0,02	0,20	0,08
Todos	LM	0,01	0,01	0,01
	ELM*	0,07	0,06	0,02
	ELMtc*	1,65	1,97	4,04

* Função de ativação seno.

5.3 EXPERIMENTO B

Nesse experimento foi utilizada a base de dados IVP (1920×1080p) com 128 amostras de vídeo contendo a DMOS de distorções de codificação Dirac, H.264 e MPEG-2, bem como distorções de transmissão em rede IP com 30, 40, 30 e 28 amostras, respectivamente. O método de validação cruzada utilizado neste experimento foi uma versão do método de resistência (KOHAVI, 1995; HAYKIN, 1999) com a utilização das métricas PSNR, SSIM, MS-SSIM e JPEG-NR em comparação com os métodos propostos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc. Assim, a base de dados foi divida em três grupos, $G1$, $G2$ e S em cada uma das cinco categorias Dirac, H.264, IP, MPEG-2 e "Todos". Dessa forma, para a categoria Dirac, por exemplo, $G1$ e $G2$ possuem 15 amostras aleatórias e distintas em cada grupo e S possui todas as amostras, *i.e.*, 30 amostras de vídeo. Analogamente ao experimento A, a validação cruzada obedeceu à formação aleatória dos grupos $G1_k$ e $G2_k$, cujo k representa a repetição do experimento em mil vezes para cada categoria de distorção. Nos métodos NRVQA-ELM e NRVQA-ELMtc utilizou-se $\tilde{N} = \text{round}(\frac{2}{3}N)$ com N amostras em cada categoria e $\tilde{N}$ representando o número de neurônios na camada oculta. A Tabela 7 resume a quantidade de amostras e o número de neurônios utilizados na camada oculta para cada conteúdo da base de dados IVP (experimento B).

Tabela 7: Número de amostras e neurônios na camada oculta utilizados no experimento B.

Database	Conteúdo	Conjunto	Amostras (N)	$\tilde{N} = \text{round}\left(\frac{2}{3}N\right)$
IVP	Dirac	$G1/G2$	15	10
		S	30	20
	H.264	$G1/G2$	20	13
		S	40	27
	IP	$G1/G2$	14	9
		S	28	19
	MPEG-2	$G1/G2$	15	10
		S	30	20
	Todos	$G1/G2$	64	43
		S	128	85

A Figura 24 mostra a variação dos coeficientes obtidos para os conjuntos $G1$, $G2$ e S relacionados aos conteúdos da base de dados IVP. Os valores das medianas dos coeficientes β, tanto no experimento anterior quanto neste, poderiam ser incorporados ao modelo da Equação (77), a fim de realizar uma avaliação objetiva de qualidade de vídeo sem referência com

valores fixos de β para tipos específicos de vídeo ou distorções, e.g., compressão Dirac, H.264 e MPEG-2.

A Figura 25 compara a acurácia do método proposto NRVQA-LM com as métricas PSNR, SSIM, MS-SSIM e JPEG-NR para todas as categorias da base de dados IVP. Nesse experimento, o método NRVQA-LM nos pares treinamento-teste disjuntos apresenta uma mediana da acurácia equivalente às métricas FR nos conteúdos em Dirac e H.264, conforme as Figuras 25-e e 25-j. No conteúdo contendo perda de pacotes (distorções em IP) da Figura 25-o, esse método nos pares disjuntos apresenta desempenho equivalente à métrica PSNR, bem como expressa maior acurácia do que as métricas SSIM, MS-SSIM e JPEG-NR. Analogamente ao experimento A, mesmo nos pares disjuntos, o método NRVQA-LM expressa maior desempenho do que as métricas FR no conteúdo em MPEG-2, conforme mostra a Figura 25-t. Entretanto, tal qual ocorreu no experimento A, quando esse método é aplicado em amostras de vídeos embaralhadas, conforme a Figura 25-y, reduz o seu desempenho, devido à exigência de um treinamento especializado dos parâmetros β, *e.g.*, tipo de compressão, distorção e sua intensidade.

As Figuras 26 e 27 apresentam a acurácia dos métodos propostos NRVQA-ELM e NRVQA-ELMtc, respectivamente. Analogamente ao experimento A, as funções de ativação sin e sig apresentam desempenho equivalente no método NRVQA-ELMtc. No entanto, ele apresenta menor amplitude interquartílica do que o método NRVQA-ELM, conforme inspeção visual da Figura 27. Nesse experimento o método NRVQA-ELMtc com a função de ativação sin nos pares disjuntos $G1$-$G2$ e $G2$-$G1$ do conteúdo em IP expressa uma mediana do PLCC igual a 0,75 que é superior às medianas de PLCC das métricas FR, cuja métrica MS-SSIM apresenta uma mediana da acurácia igual a 0,54. Nos conteúdos em H.264 e MPEG-2 esse método mostra um desempenho equivalente às métricas FR com a mediana do PLCC igual a 0,82 e 0,73 contra as medianas do PLCC de 0,85 e 0,78 da métrica MS-SSIM, respectivamente. Porém, nos conteúdos em Dirac e Todos com os pares disjuntos, o método NRVQA-ELMtc apresenta uma acurácia inferior à métrica FR SSIM, *i.e.*, os valores da mediana do PLCC desse método são de 0,77 e 0,49 contra 0,87 e 0,65 da métrica MS-SSIM, respectivamente.

A Tabela 8 mostra a mediana da distribuição dos tempos de treinamento dos métodos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc para cada grupo de treinamento da base de dados IVP. Os métodos NRVQA-LM e NRVQA-ELM consomem um tempo de treinamento entre 20 ms e 30 ms, exceto para a categoria “Todos”, cujo tempo de treinamento do método NRVQA-ELM foi de 10 ms. O método NRVQA-ELMtc devido ao critério de parada descrito no Algoritmo 1 (Seção 4.2.1) apresenta maior variação no tempo de treinamento, principalmente para o grupo $G2$, com uma variação entre 0,18 s e 1,10 s. Analogamente ao experimento A, a

mediana aproximada da distribuição dos tempos de teste do método NRVQA-LM foi de 12 ms e para os métodos NRVQA-ELM e NRVQA-ELMtc foi de 10 ms.

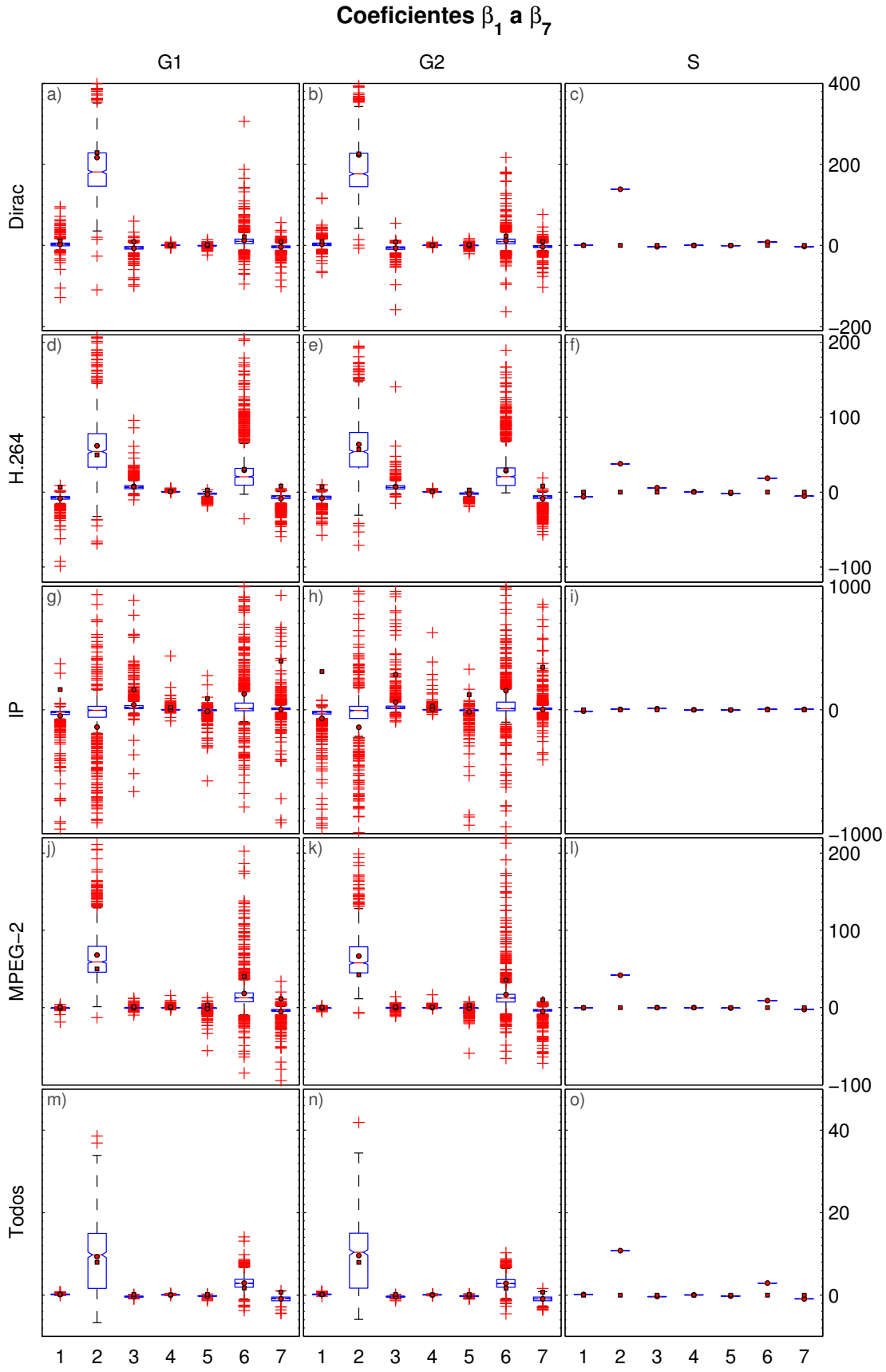

Figura 24: Coeficientes β_1 a β_7 do método NRVQA-LM nos grupos de treinamento $G1$, $G2$ e S para os conteúdos do experimento B.

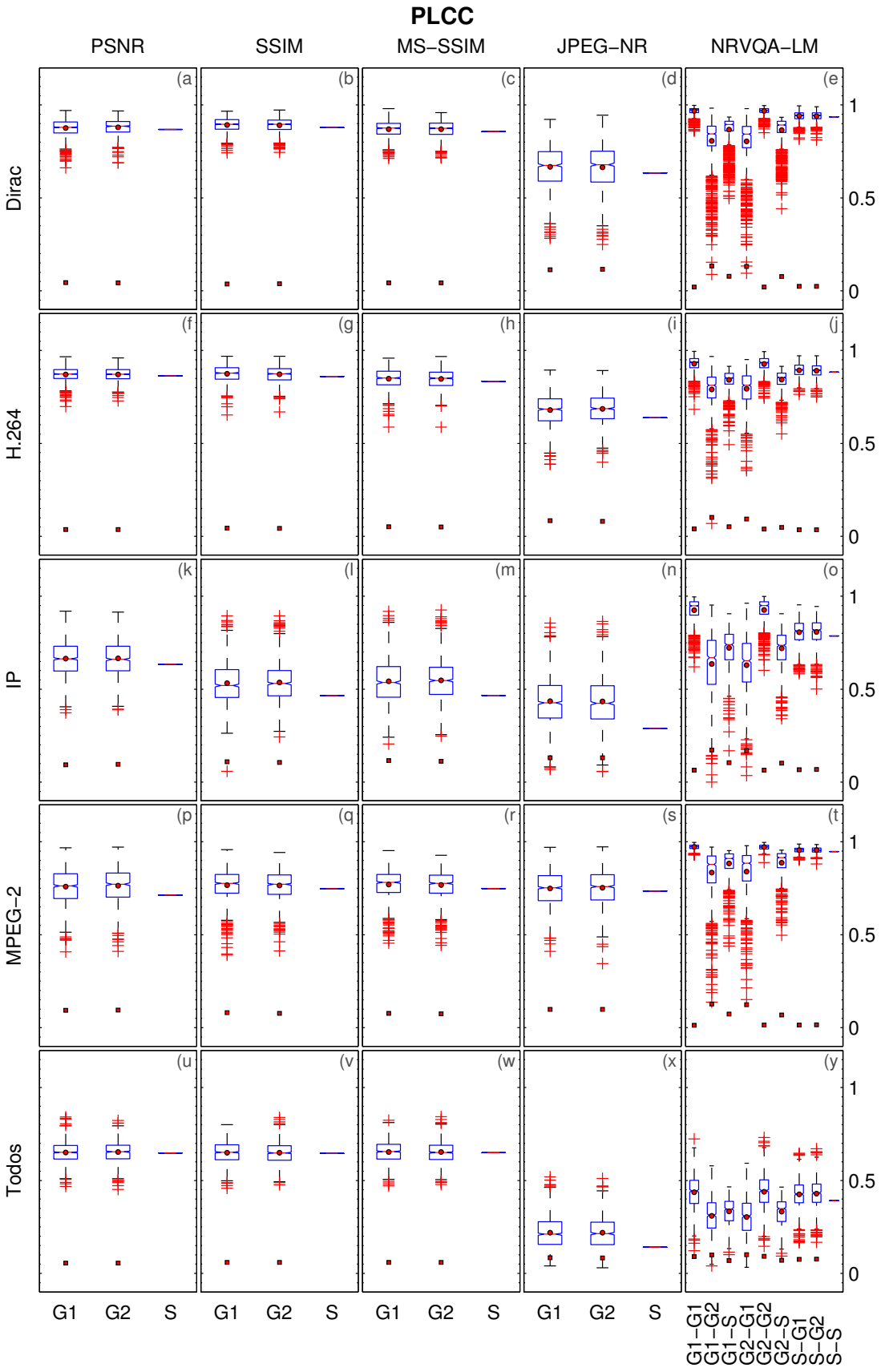

Figura 25: Comparação da acurácia (PLCC) entre as métricas PSNR, SSIM, MS-SSIM, JPEG-NR e NRVQA-LM para os conteúdos do experimento B.

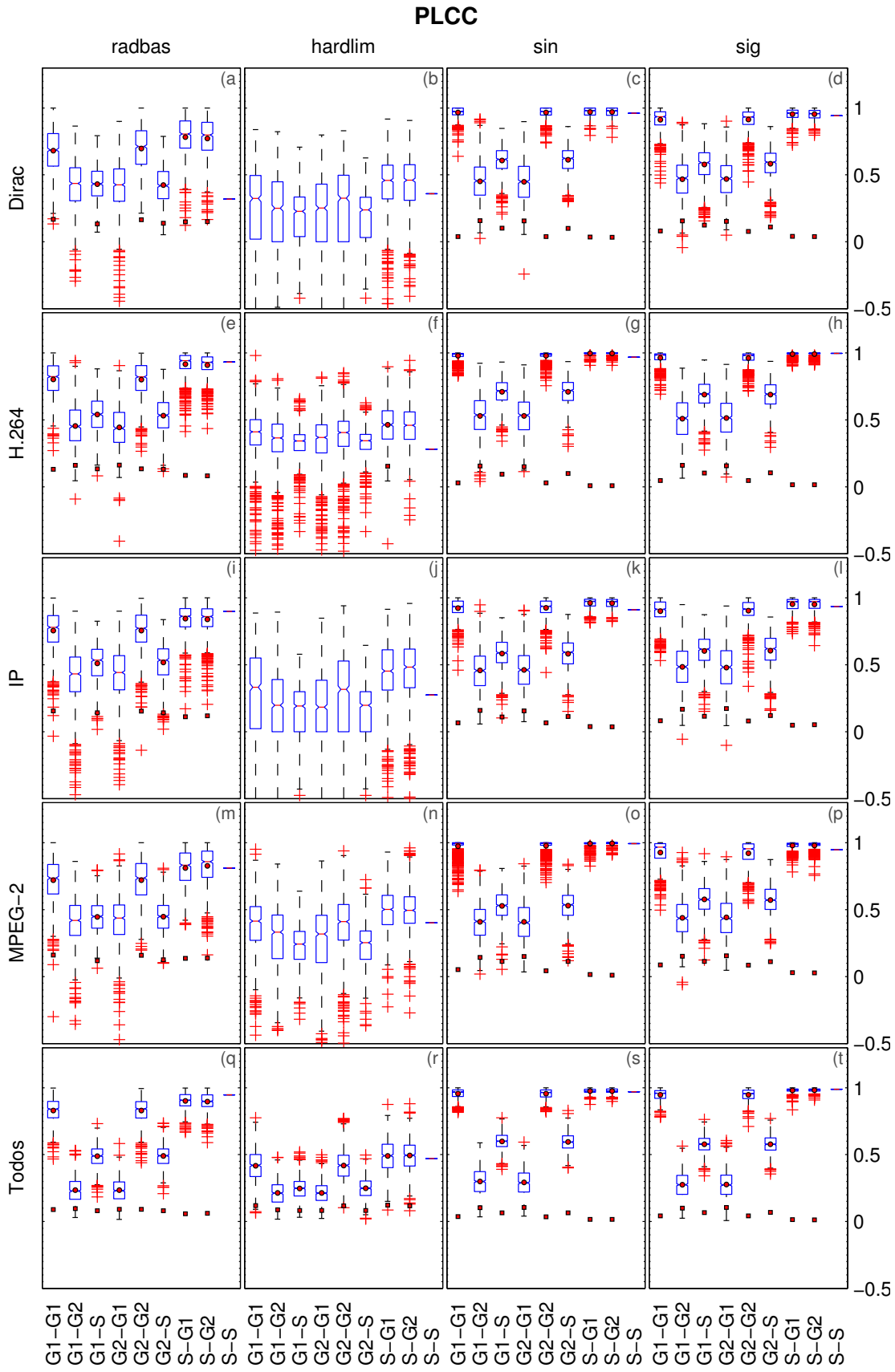

Figura 26: Comparação da acurácia (PLCC) do método NRVQA-ELM para os conteúdos do experimento B.

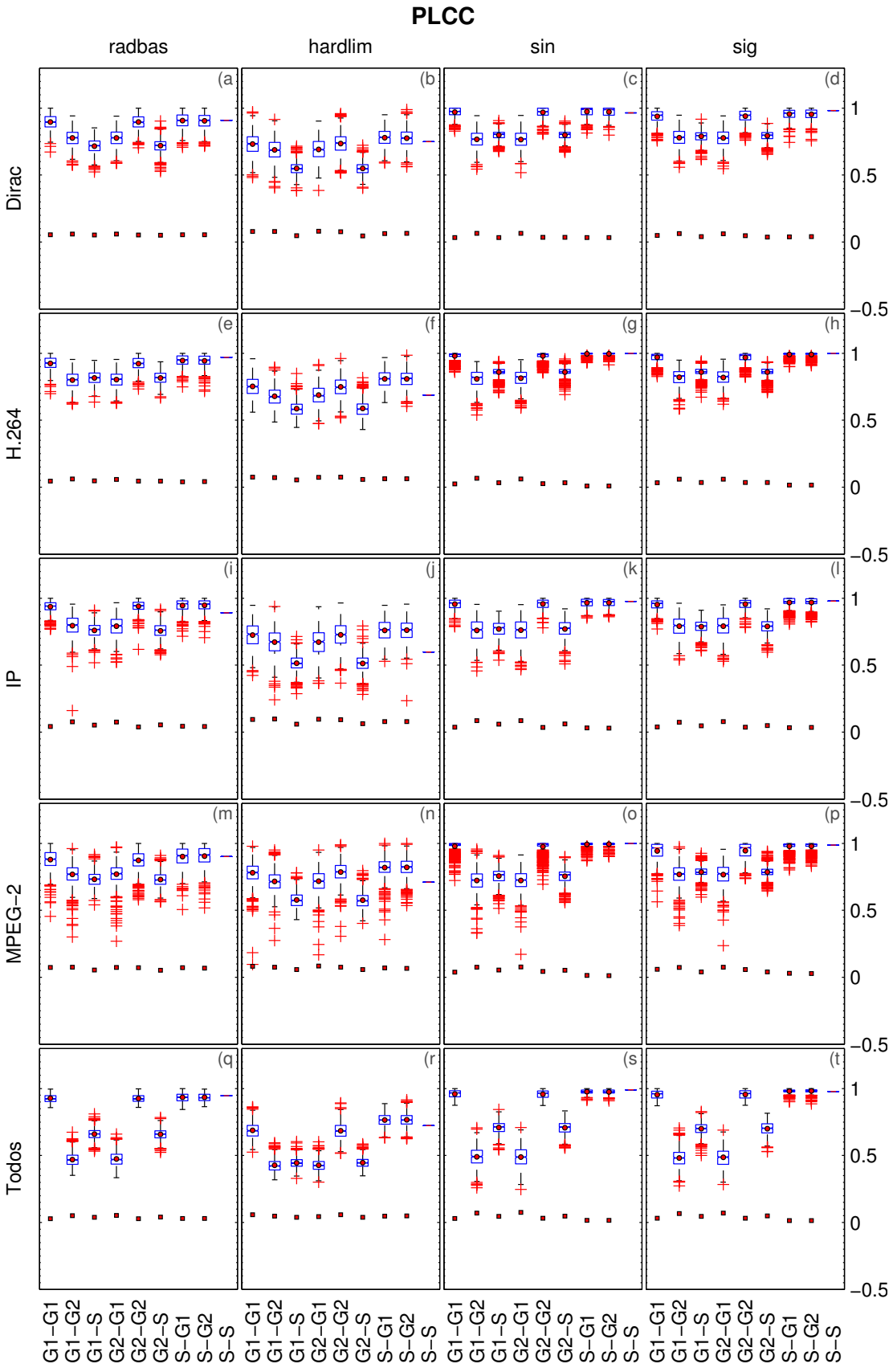

Figura 27: Comparação da acurácia (PLCC) do método NRVQA-ELMtc para os conteúdos do experimento B.

Tabela 8: Mediana do tempo de treinamento dos métodos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc para o experimento B.

Categoria	NRVQA	Conjunto de Treinamento (s)		
		*G*1	*G*2	*S*
Dirac	LM	0,02	0,02	0,02
	ELM*	0,03	0,03	0,03
	ELMtc*	0,03	0,18	0,01
H.264	LM	0,03	0,03	0,03
	ELM*	0,02	0,02	0,04
	ELMtc*	0,03	0,40	0,02
IP	LM	0,03	0,02	0,02
	ELM*	0,03	0,02	0,03
	ELMtc*	0,03	0,03	0,93
MPEG-2	LM	0,02	0,02	0,02
	ELM*	0,02	0,03	0,03
	ELMtc*	0,03	0,19	0,01
Todos	LM	0,02	0,02	0,02
	ELM*	0,01	0,01	0,01
	ELMtc*	0,02	1,10	0,02

* Função de ativação seno.

5.4 EXPERIMENTO C

Experimento que envolveu 2.627 amostras de vídeo com 17 bases de dados denominadas superconjunto S, contendo a MOS e diversas distorções, conforme descrição da Subseção 3.1.3. A mesma metodologia aplicada nos experimentos A e B quanto ao método de validação cruzada foi utilizada neste, com uma versão do método de resistência (KOHAVI, 1995; HAYKIN, 1999) com $k = 1$, conforme a Figura 18. Além disso, os grupos $G1_k$ e $G2_k$ são formados de forma distinta e aleatória, ambos com 1.313 e 1.314 amostras, respectivamente. As mesmas métricas utilizadas nos experimentos anteriores também são empregadas neste ensaio, exceto a proposta NRVQA-LM que, devido ao seu treinamento especializado, exige o treinamento de categorias de vídeos com características similares, em termos de distorção ou codificação. Pois, no superconjunto S, a maioria das bases de dados não explicitam as suas características, tais como padrão de codificação e taxas de compressão. Dessa forma, neste experimento são comparadas as métricas PSNR, SSIM, MS-SSIM e JPEG-NR e as propostas NRVQA-ELM e NRVQA-ELMtc com o número de neurônios na camada oculta no intervalo $\tilde{N} = 1, \ldots, N$. Assim, enquanto PLCC < 1, $\tilde{N}$ é incrementado até chegar aos valores da quantidade de amostras contidas nos grupos $G1_k$, $G2_k$ e S, *i.e.*, 1.313, 1.314 e 2.627, respectivamente.

A Figura 28 apresenta a medida de acurácia (PLCC) das métricas PSNR, SSIM, MS-SSIM e JPEG-NR para os conjuntos $G1$, $G2$ e S. Os valores dessa medida são inferiores a 0,17 para todos os conjuntos tratados, devido ao tamanho das amostras contidas nos grupos $G1$, $G2$ e S, *i.e.*, essas métricas não conseguem estabelecer uma maior predição de acurácia quando são formados grupos contendo muitas amostras distintas e aleatórias.

As Figuras 29 e 30 apresentam a acurácia em relação ao número de neurônios utilizados na camada oculta e da função de ativação usada. O par treinamento-teste é representado pela combinação linha *vs.* coluna, *e.g.*, as Figuras 29-d e 30-d mostram os resultados para o par treinamento-teste $G2$-$G1$, *i.e.*, $Tr(G2)$ na linha e $Ts(G1)$ na coluna. Logo, a validação cruzada ocorre com nove variações possíveis entre $G1$, $G2$ e S. A comparação entre as Figuras 29 e 30 mostra que o método proposto NRVQA-ELMtc conduz a resultados mais estáveis (menos oscilação nas curvas de PLCC) do que o método NRVQA-ELM. Assim, as Figuras 30-a, 30-e e 30-i em diagonal mostram que desempenho da acurácia é maior quando o conjunto de teste é idêntico ao do treinamento, *i.e.*, o treinamento reconhece integralmente o padrão de teste mesmo com um número menor de neurônios na camada oculta. Quando são usados os pares de treinamento-teste $G1$-S e $G2$-S (Figuras 30-c e 30-f) observa-se que a partir de 400 neurônios na camada oculta, ocorre redução na acurácia até convergir entre 0,7 e 0,8 com $\tilde{N} = 2.627$, devido à saturação de neurônios na camada oculta. Os pares de treinamento-teste S-$G1$ e S-$G2$ (Figuras 30-g e

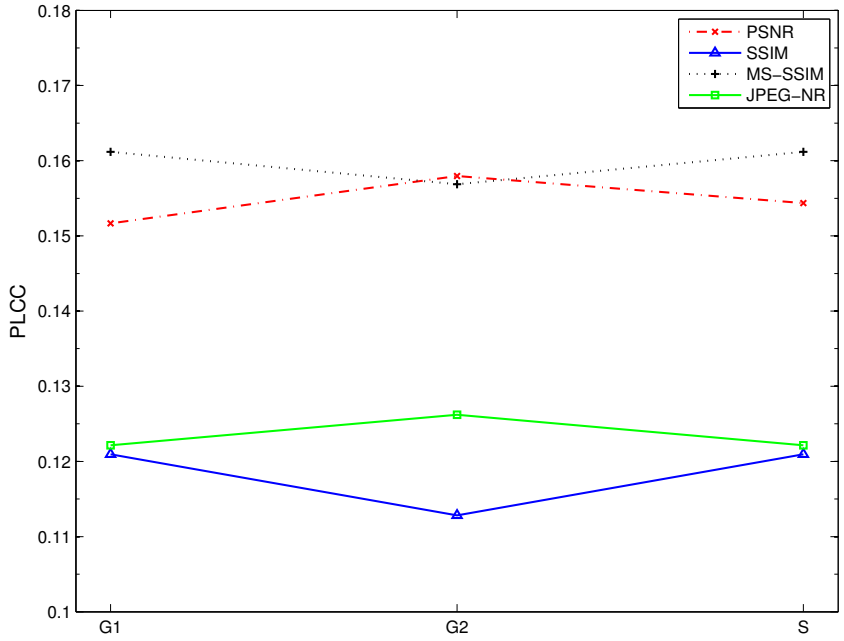

Figura 28: Comparação da acurácia (PLCC) entre as métricas PSNR, SSIM, MS-SSIM e JPEG-NR para os grupos de teste $G1$, $G2$ e S do experimento C.

30-h) convergem com PLCC próximo a 1 para $\tilde{N} > 800$. Entretanto, no cenário prático, quando os pares treinamento-teste são disjuntos, *i.e.*, $Tr(G1)$-$Ts(G2)$ e $Tr(G2)$-$Ts(G1)$, conforme as Figuras 30-b e 30-d, respectivamente, o método NRVQA-ELMtc com a função de ativação sin e $\tilde{N} = 400$, expressa um valor de acurácia maior do que 0,8. No entanto, a sua acurácia decresce quando $\tilde{N} > 400$, *i.e.*, o superdimensionamento de neurônios na camada oculta reduz a capacidade de generalização do método NRVQA-ELMtc. Logo, comparando os resultados das métricas FR nos grupos de teste $G1$ e $G2$ da Figura 28, com os resultados obtidos pelo método NRVQA-ELMtc nos pares disjuntos, conforme as 30-b e 30-d, observa-se que este método apresenta desempenho superior às métricas FR, quando é utilizada uma grande quantidade de amostras de treinamento com características espaço-temporais distintas.

As Figuras 31 e 32 representam a significância estatística, segundo uma distribuição F percentual, conforme a Fórmula (40), entre a métrica FR MS-SSIM e os métodos propostos NRVQA-ELM e NRVQA-ELMtc, respectivamente. A linha sombreada indica o valor definido pela iCDF que é obtido com a substituição de ζ da Fórmula (41) na Fórmula (40), a qual expressa o limiar de significância estatística positiva, *i.e.*, quando o valor de F_p é maior do que a linha sombreada, há uma diferença significativa positiva do método proposto em relação à métrica de referência completa MS-SSIM. Logo, nesse caso, o método proposto apresenta desempenho superior à métrica FR MS-SSIM. Nesse experimento não foi representado o limiar da significância estatística negativa, determinado pela substituição de ζ da Fórmula (44) na Fórmula (43), pois as Figuras 31 e 32 possuem apenas valores positivos no eixo das ordena-

das. Analogamente aos resultados de acurácia, o método NRVQA-ELMtc apresenta melhor desempenho do que a versão NRVQA-ELM. Nos pares de treinamento-teste disjuntos ($G1$-$G2$ e $G2$-$G1$), os quais representam cenários de aplicação prática, o método NRVQA-ELMtc com a função de ativação sin, em termos de F_p, expressa um desempenho superior à métrica de referência completa MS-SSIM para $\tilde{N} < 400$, *i.e.*, quando o número de neurônios na camada oculta é menor do que 400, conforme mostram as Figuras 32-b e 32-d, respectivamente. Além disso, para $\tilde{N} < 1.000$, há uma diferença significativa positiva entre a métrica FR MS-SSIM e o método NRVQA-ELM com qualquer função de ativação usada, conforme o limiar representado pela linha sombreada nessas figuras.

Os tempos de treinamento dos métodos NRVQA-ELM e NRVQA-ELMtc estão representados nas Figuras 33 e 34, respectivamente. Como já fora mencionado, o método NRVQA-ELMtc pode exigir maior custo computacional do que a versão NRVQA-ELM, cuja diferença no tempo de treinamento é aproximadamente igual ao número máximo de iterações (variável itmax do Algoritmo 1, descrito na Seção 4.2) que método NRVQA-ELMtc pode alcançar, *i.e.*, na ordem de 10^2 vezes o tempo da versão NRVQA-ELM. Na fase de teste, analisando os pares de treinamento-teste disjuntos $G1$-$G2$ e $G2$-$G1$, o método NRVQA-ELMtc recorre aos parâmetros de treinamento da RNA e apresenta cerca de um segundo a mais, do que a versão NRVQA-ELM, conforme descreve as Figuras 35 e 36, respectivamente.

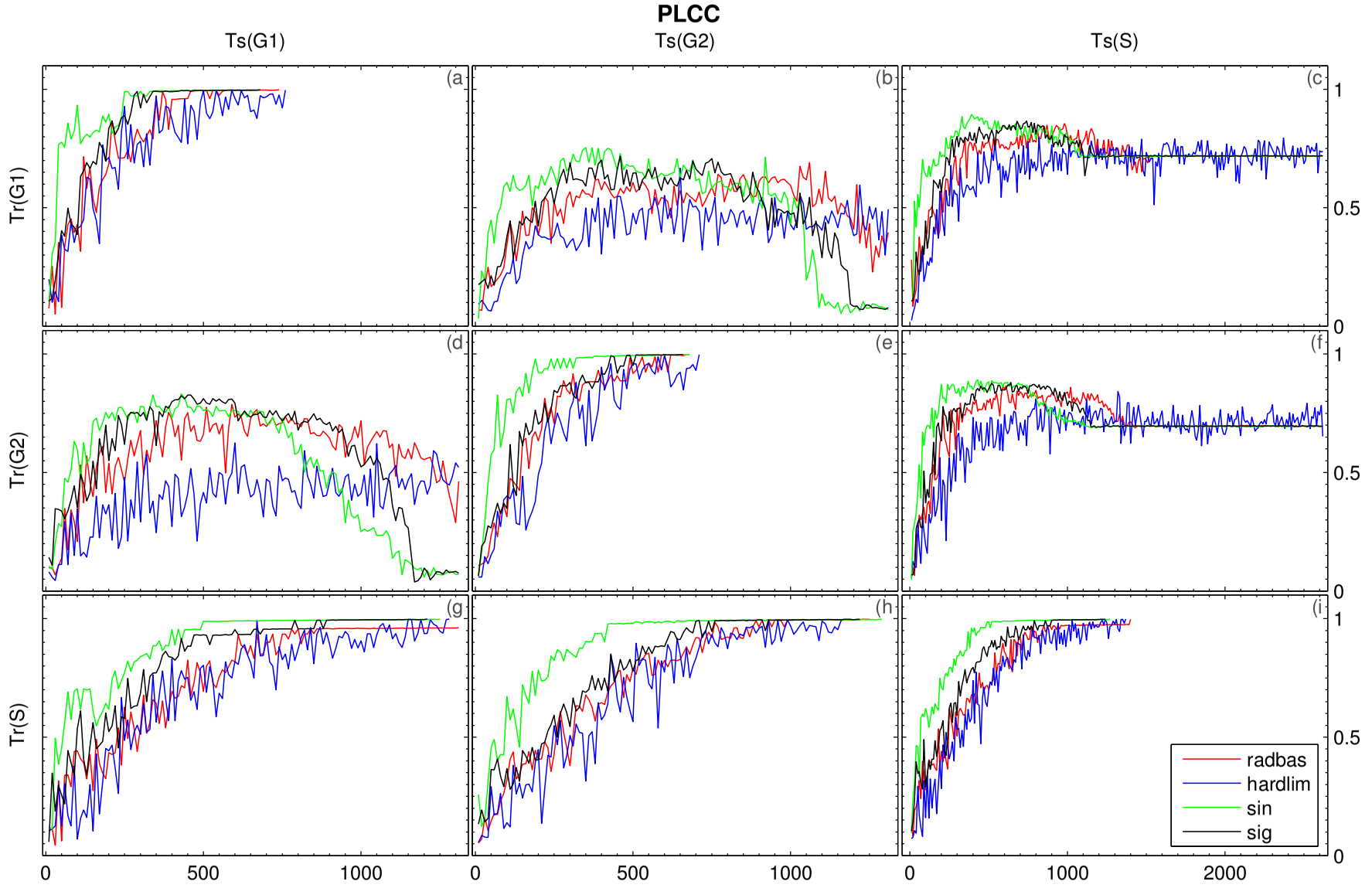

Figura 29: Comparação da acurácia (PLCC) do método NRVQA-ELM com a validação cruzada entre os grupos $G1$, $G2$ e S do experimento C.

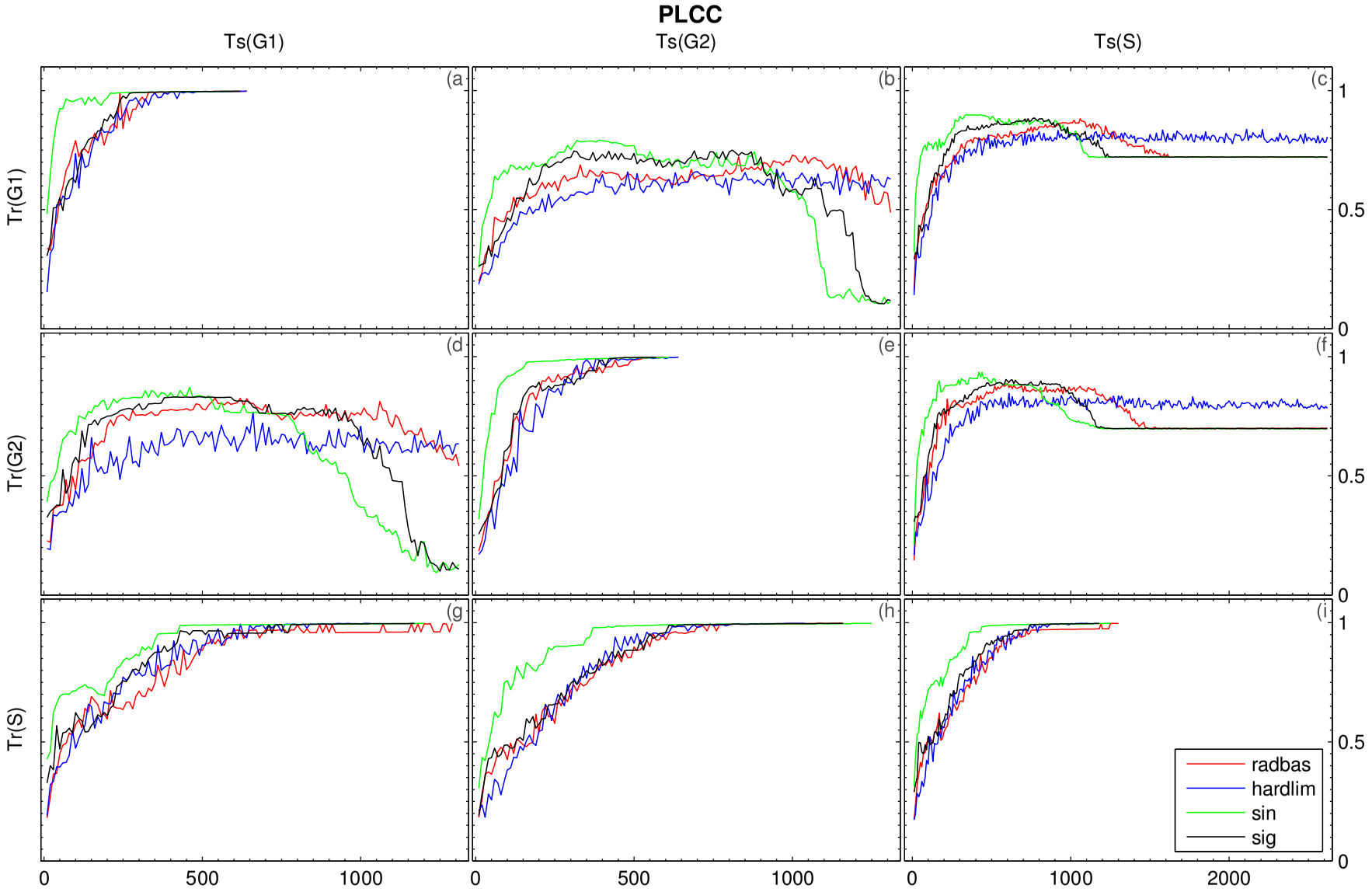

Figura 30: Comparação da acurácia (PLCC) do método NRVQA-ELMtc com a validação cruzada entre os grupos $G1$, $G2$ e S do experimento C.

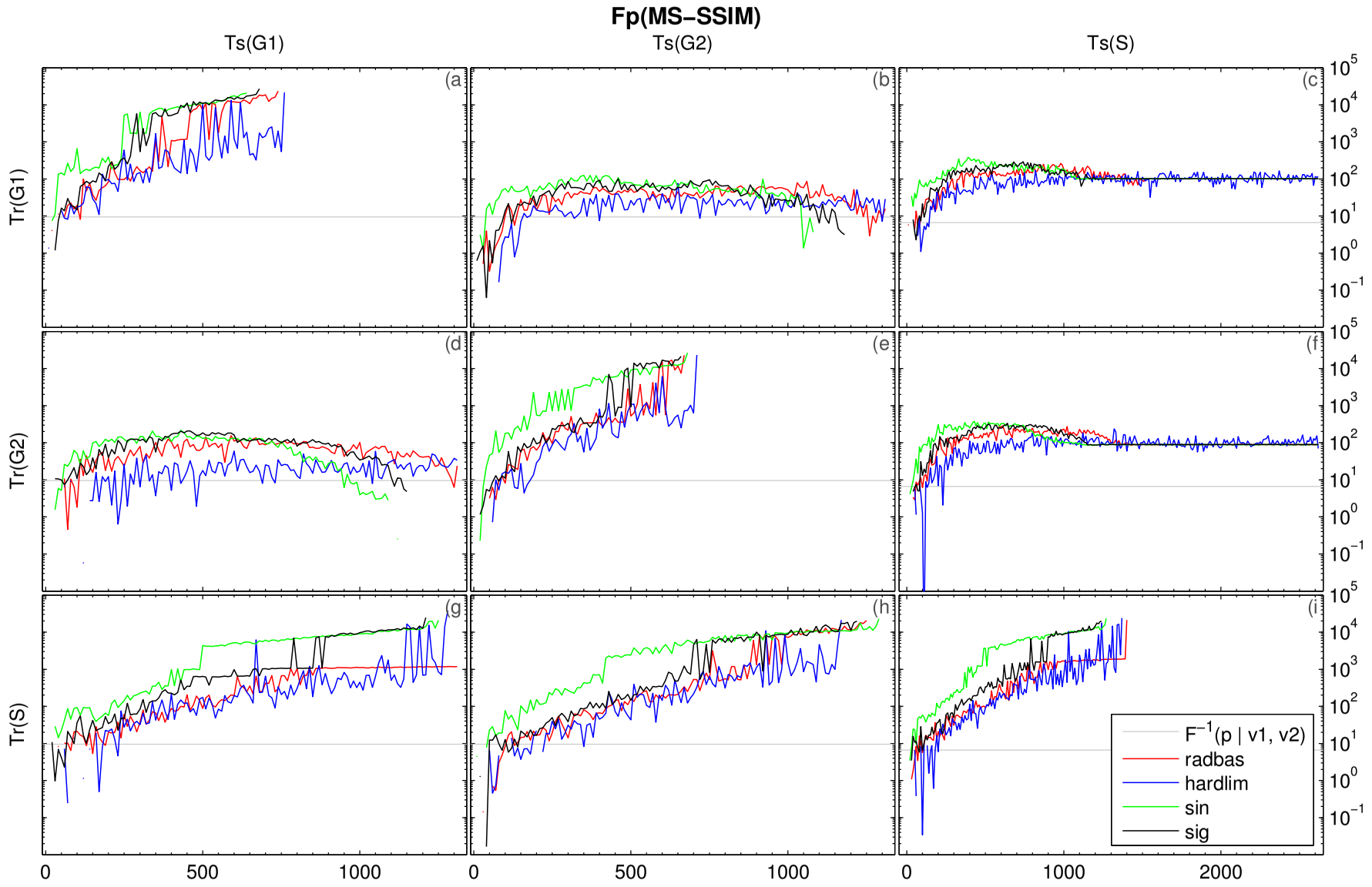

Figura 31: Distribuição F percentual (F_p) entre a métrica FR MS-SSIM e o método NRVQA-ELM com validação cruzada entre os grupos $G1$, $G2$ e S do experimento C.

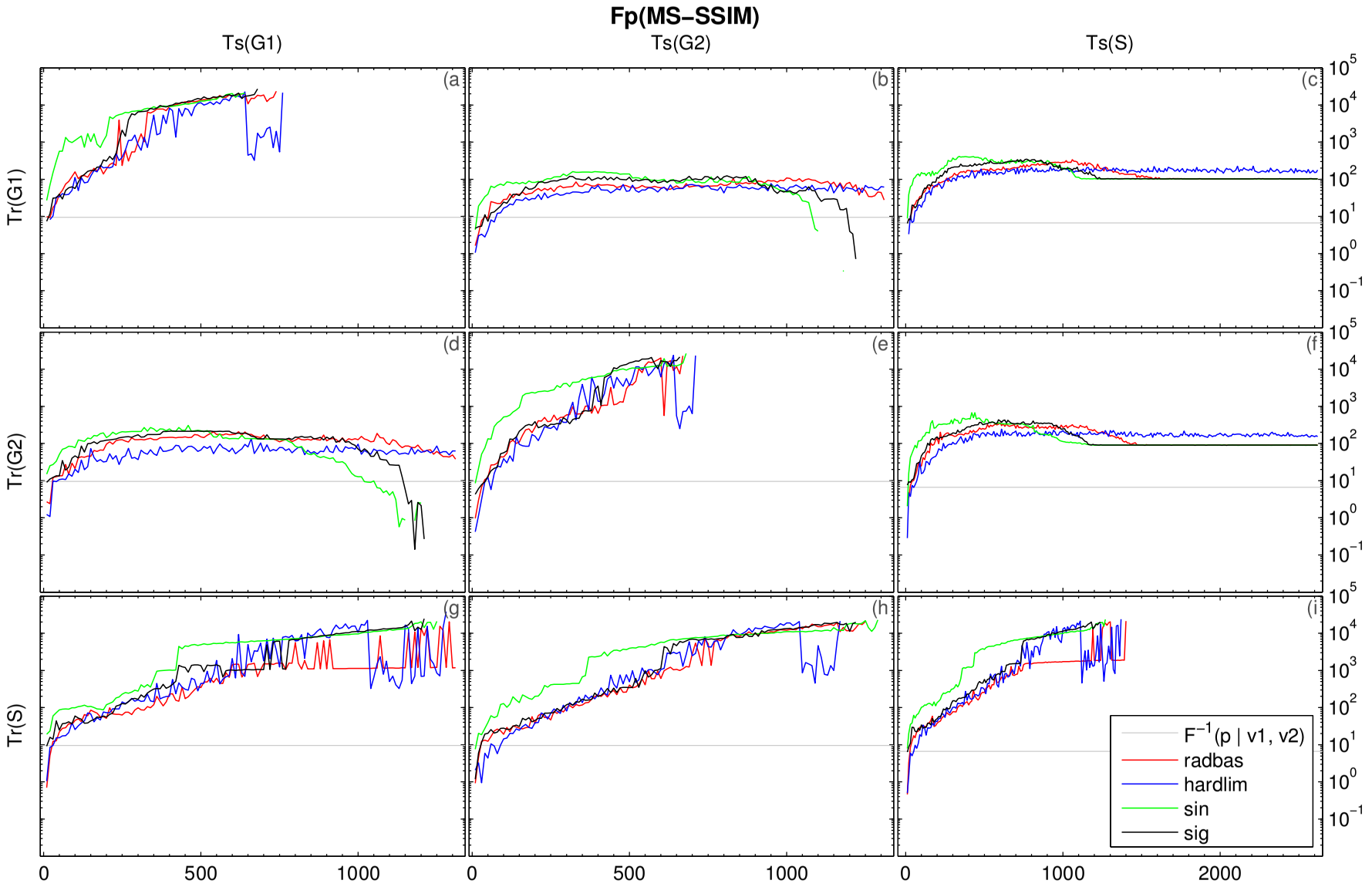

Figura 32: Distribuição F percentual (F_p) entre a métrica FR MS-SSIM e o método NRVQA-ELMtc com validação cruzada entre os grupos $G1$, $G2$ e S do experimento C.

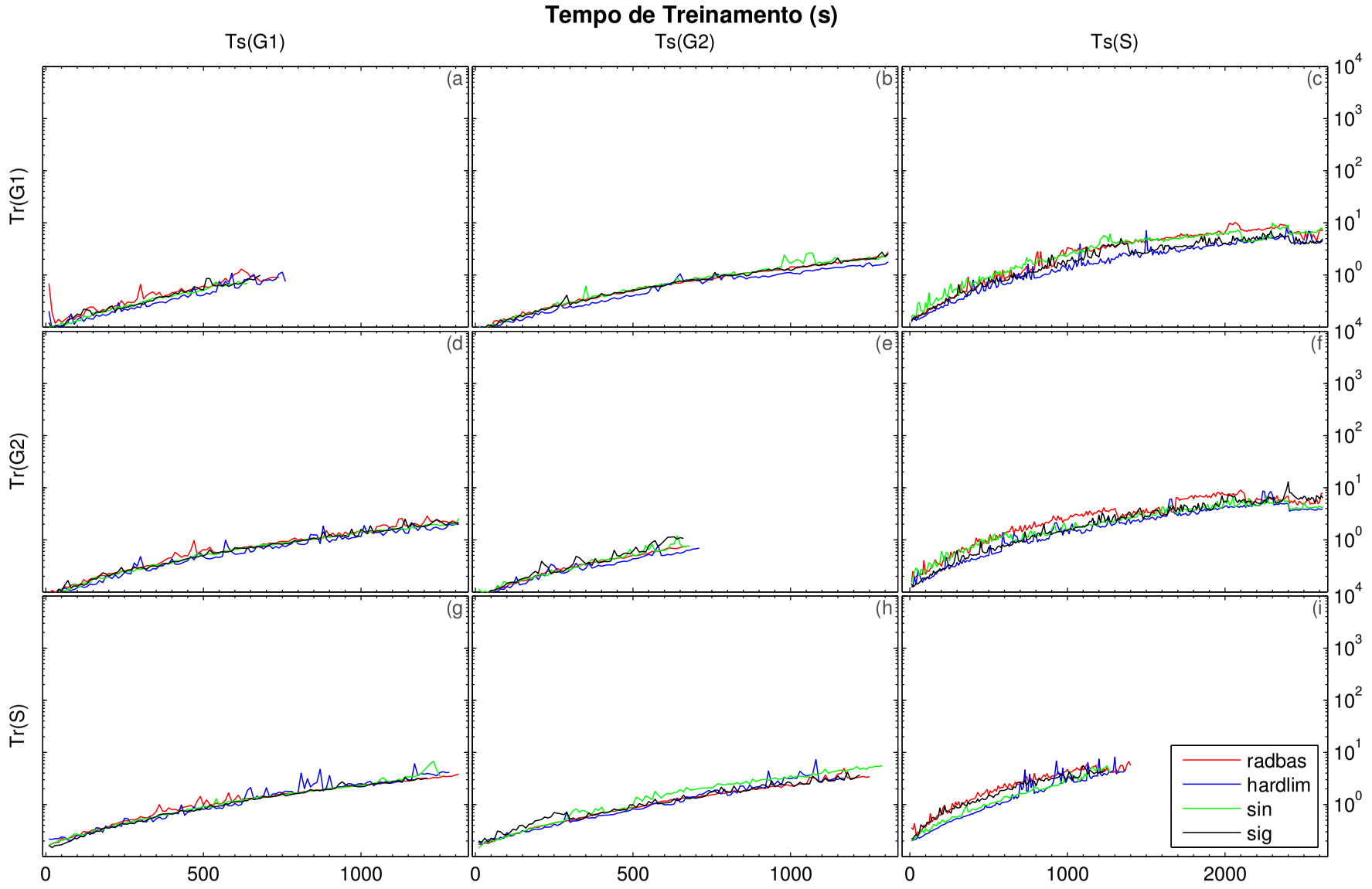

Figura 33: Tempo de treinamento do método NRVQA-ELM com a validação cruzada entre os grupos $G1$, $G2$ e S do experimento C.

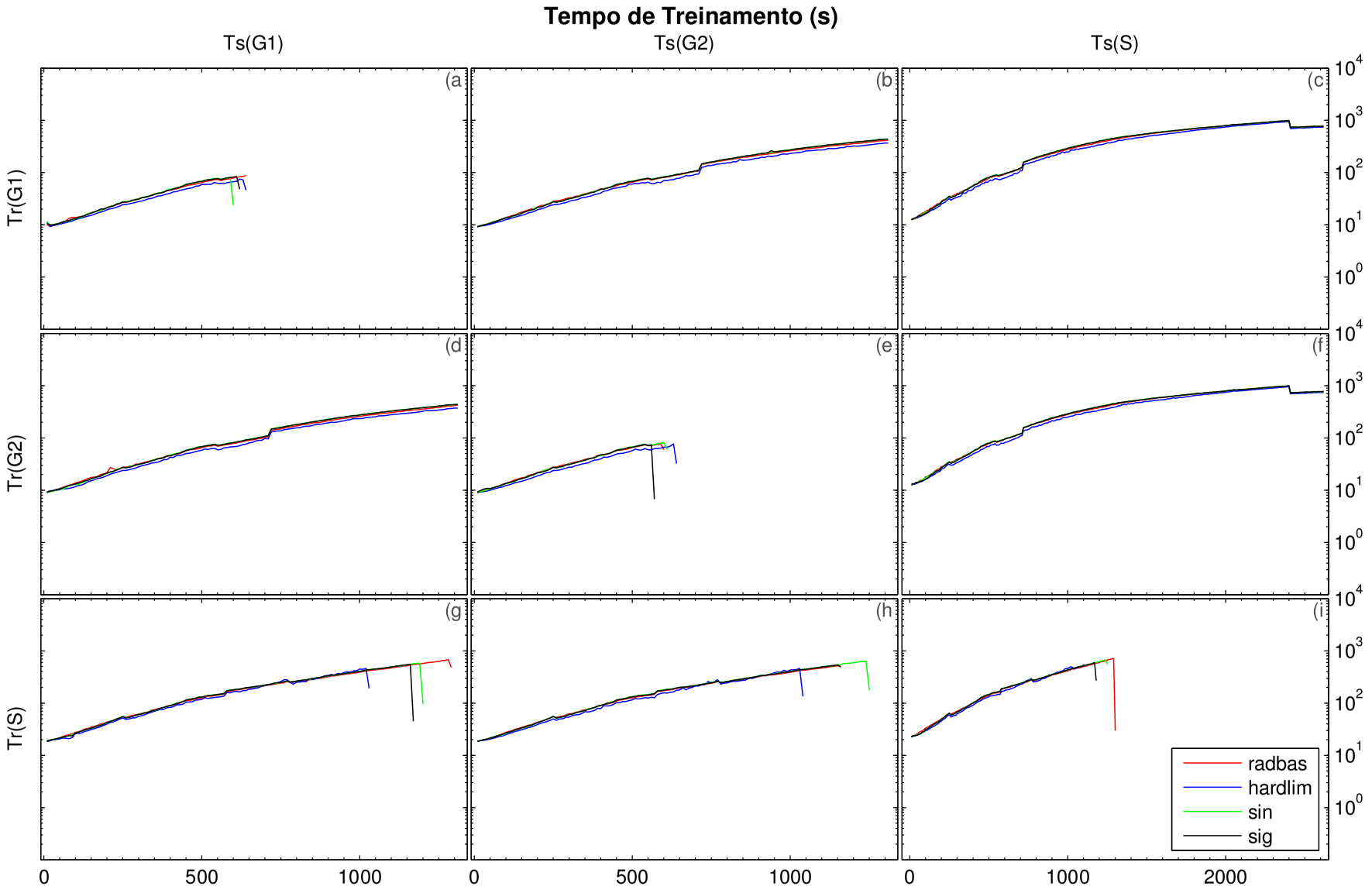

Figura 34: Tempo de treinamento do método NRVQA-ELMtc com a validação cruzada entre os grupos $G1$, $G2$ e S do experimento C.

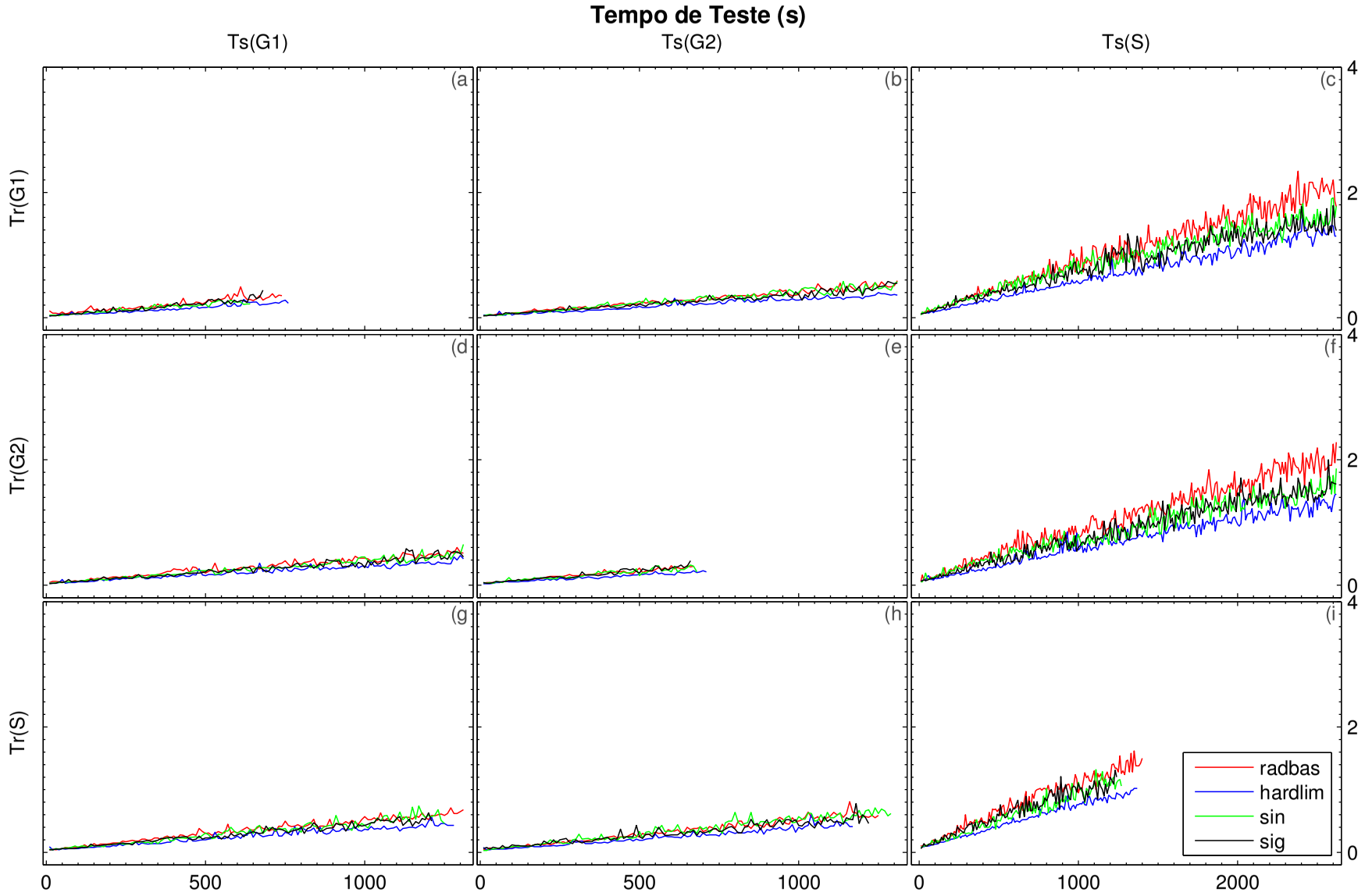

Figura 35: Tempo de teste do método NRVQA-ELM com a validação cruzada entre os grupos $G1$, $G2$ e S do experimento C.

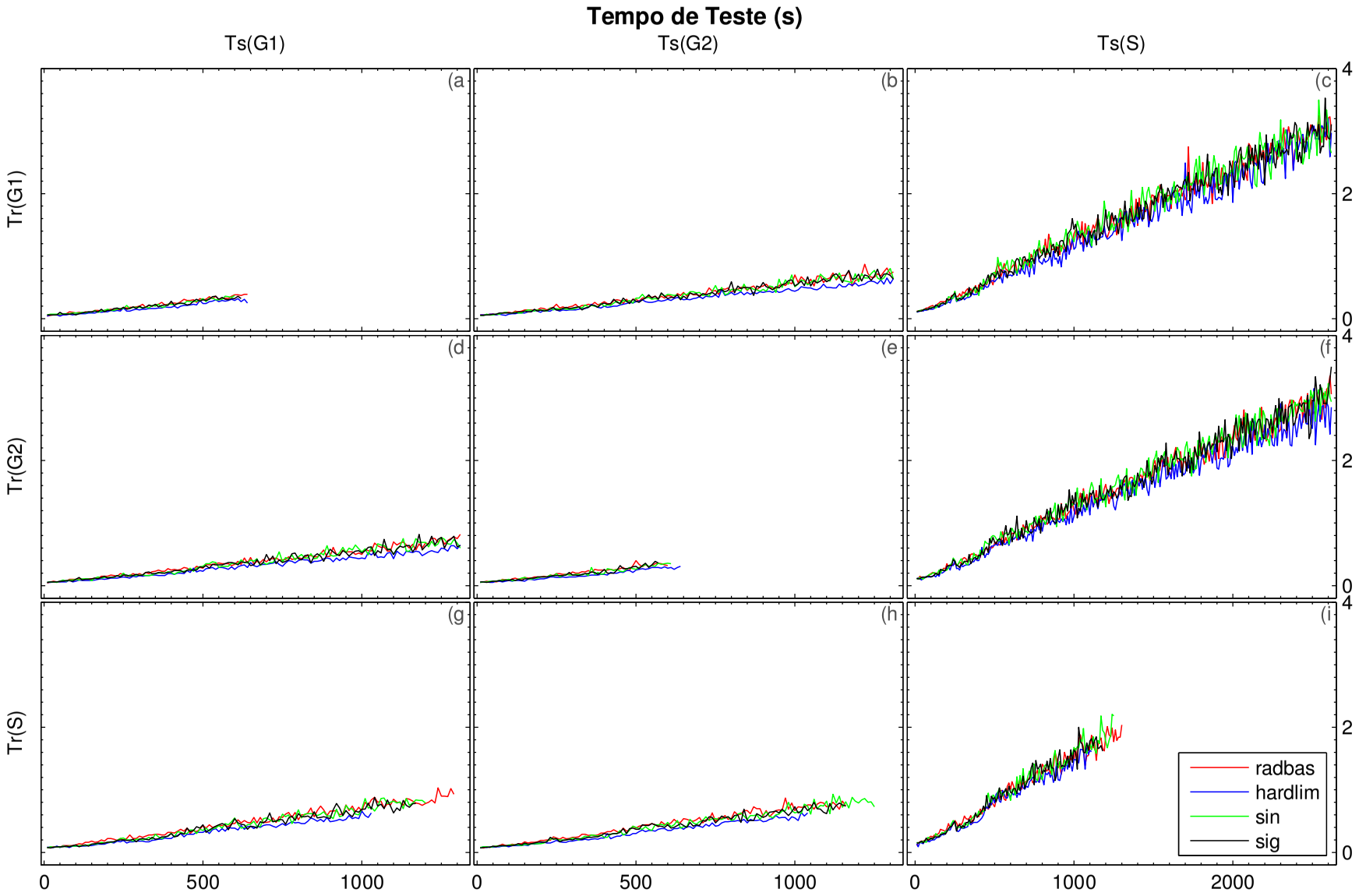

Figura 36: Tempo de teste do método NRVQA-ELMtc com a validação cruzada entre os grupos $G1$, $G2$ e S do experimento C.

5.5 EXPERIMENTO D

Experimento com a distribuição de 100 pontos ($k = 1, ..., 100$) e escores subjetivos (MOS) do superconjunto S contendo 2.627 amostras de vídeo, em que cada k representa um par treinamento-teste Tr_k-Ts_k distinto em cada categoria (resoluções LD, SD, HDp, HDi e "Todos"), cujo k-ésimo teste não é subconjunto do k-ésimo treinamento, ou seja, $Ts_k \not\subset Tr_k$, *e.g.*, na categoria LD com 204 vídeos, um grupo de treinamento (Tr) é formado por 179 amostras, caso o conjunto de teste (Ts) contenha 25 amostras (fps e $\tilde{N}$ iguais a 25), conforme mostra a Figura 18. Assim, o tamanho do grupo de treinamento é determinado pela diferença entre a categoria do superconjunto S e o seu respectivo grupo de teste, *i.e.*, $Tr_k = S_{\text{resolução}} - Ts_k$, em que Tr_k é o conjunto de treinamento, $S_{\text{resolução}}$ é a quantidade total de amostras na categoria de vídeo e Ts_k é constituído a partir da quantidade de amostras de vídeos, *i.e.*, é uma aproximação de quadros por segundo (fps). Os atuais sistemas de codificação apresentam valores típicos de 25, 30 e 50 quadros por segundo e as quantidades de amostras 250, 300 e 500 representam dez segundos de um vídeo que também são quantidades típicas encontradas em diversas bases de dados. Assim, cada resolução (LD, SD, HDp, HDi ou "Todos") contém seis variações, exceto a resolução LD que apresenta fps com valores de 25, 30 e 50, pois no superconjunto S há apenas 204 amostras de vídeos em LD. No processo de validação são empregadas as métricas PSNR, SSIM, MS-SSIM, JPEG-NR em comparação com os métodos NRVQA-ELM e NRVQA-ELMtc, ambos com o número de neurônios na camada oculta igual ao número de amostras de vídeo, *i.e.*, $\tilde{N} = \text{fps}$.

A Figura 37 compara a distribuição da acurácia entre as métricas de referência completa PSNR, SSIM, MS-SSIM e sem referência JPEG-NR para cada categoria do superconjunto S em função da quantidade de amostras de teste (fps). Pela inspeção visual da Figura 37, observa-se que a acurácia dessas métricas diminui quando fps > 50 nas categorias SD, HDp, HDi e "Todos". Essas métricas diminuem o desempenho na predição de qualidade quando são usadas muitas amostras de teste aleatórias, tal qual ocorre no experimento C, conforme a Figura 28 nos grupos de teste $G1$ e $G2$.

As Figuras 38 e 39 apresentam a distribuição da acurácia dos métodos NRVQA-ELM e NRVQA-ELMtc, respectivamente. A partir da comparação entre os resultados destes dois métodos propostos, observa-se que o último apresenta desempenho na predição de qualidade superior à versão NRVQA-ELM e às métricas FR nos conteúdos em LD, SD e "Todos". Para os conteúdos em HD (HDp e HDi) e valores de fps iguais a 250, 300 e 500, os métodos NRVQA-ELM e NRVQA-ELMtc apresentam uma redução de desempenho na distribuição da acurácia, devido ao decréscimo do número de amostras de treinamento, *e.g.*, o conteúdo em HDi possui

615 amostras de vídeo, quando fps $= 500$, são usadas 115 amostras apenas no treinamento com $\tilde{N} = 500$. Pois, neste experimento, o conjunto de treinamento Tr_k não inclui as amostras de teste Ts_k, conforme ilustra a Figura 18. Além disso, há uma saturação do número de neurônios na camada oculta, ou seja, $\tilde{N}$ é maior do que o conjunto de treinamento ($\tilde{N} > 115$), tal qual ocorre no experimento C com os pares treinamento-teste disjuntos $Tr(G1)$-$Ts(G2)$ e $Tr(G2)$-$Ts(G1)$, conforme as Figuras 30-b e 30-d, respectivamente. Entretanto, mesmo com fps e $\tilde{N}$ maiores do que 50, a métrica proposta NRVQA-ELMtc apresenta um desempenho na distribuição de acurácia superior àqueles observados nas métricas de referência completa na categoria "Todos", pois neste caso o número de amostras de treinamento é maior do que o a quantidade de amostras de teste e de neurônios na camada oculta. Em relação às funções de ativação, os resultados apresentam algumas variações, *e.g.*, entre as funções de ativação hardlim e sin, conforme as Figuras 39-f e 39-k, respectivamente. Esta última mostra uma maior compactação do desvio-quartil com a distribuição do PLCC próximo a 1. Ressalta-se que este experimento representa um cenário de aplicação prática, *i.e.*, quando o conjunto de treinamento e de teste são disjuntos. Assim, o método NRVQA-ELMtc nos conteúdos LD, SD e "Todos" apresenta maior desempenho do que as métricas FR da Figura 37, cujo intervalo interquartílico do PLCC, está confinado entre 0,75 e 0,95, e.g., na Figura 39-o para fps > 50 com a função de ativação sin.

A Tabela 9 exibe a mediana da distribuição dos tempos de treinamento dos métodos NRVQA-ELM e NRVQA-ELMtc para cada conjunto de treinamento do superconjunto S. O método NRVQA-ELMtc requer maior custo computacional, logo, exige maior tempo de treinamento. Esta variável cresce com o incremento do número de amostras (fps) e de neurônios na camada oculta $(\tilde{N})$. Embora a Tabela 9 mostre um crescimento do número de neurônios na camada oculta, quando fps é maior do que 250 e 300, há menos amostras de treinamento do que de teste para os conteúdos HDp e HDi, respectivamente. Logo, o tempo de processamento diminui, conforme inspeção visual dessa tabela. A mediana aproximada da distribuição dos tempos de teste do método NRVQA-ELM varia entre 10 ms e 160 ms, enquanto que o método NRVQA-ELMtc apresenta uma variação entre 30 ms e 180 ms.

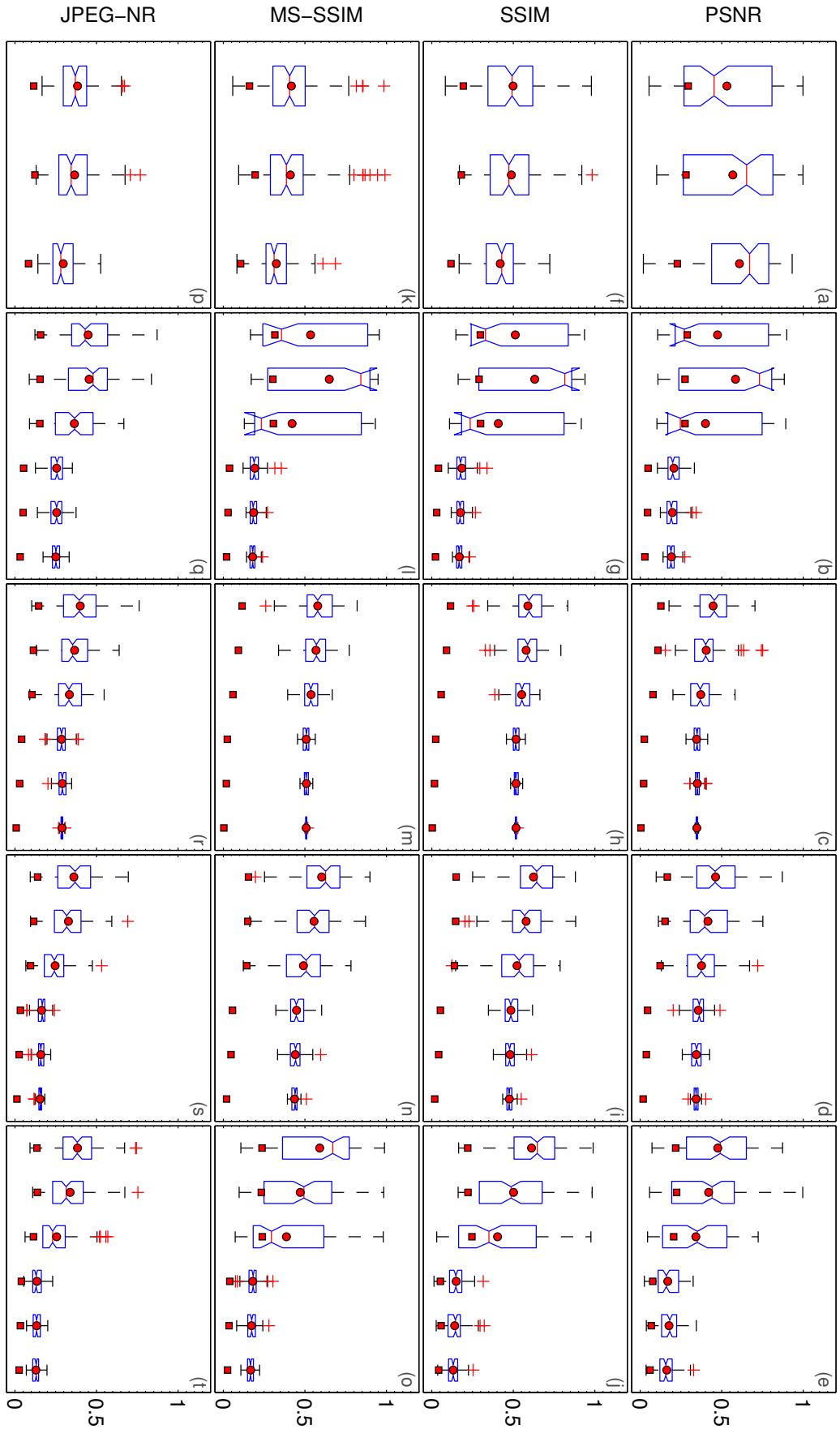

Figura 37: Comparação da acurácia (PLCC) entre as métricas PSNR, SSIM, MS-SSIM e JPEG-NR para os conteúdos do experimento D.

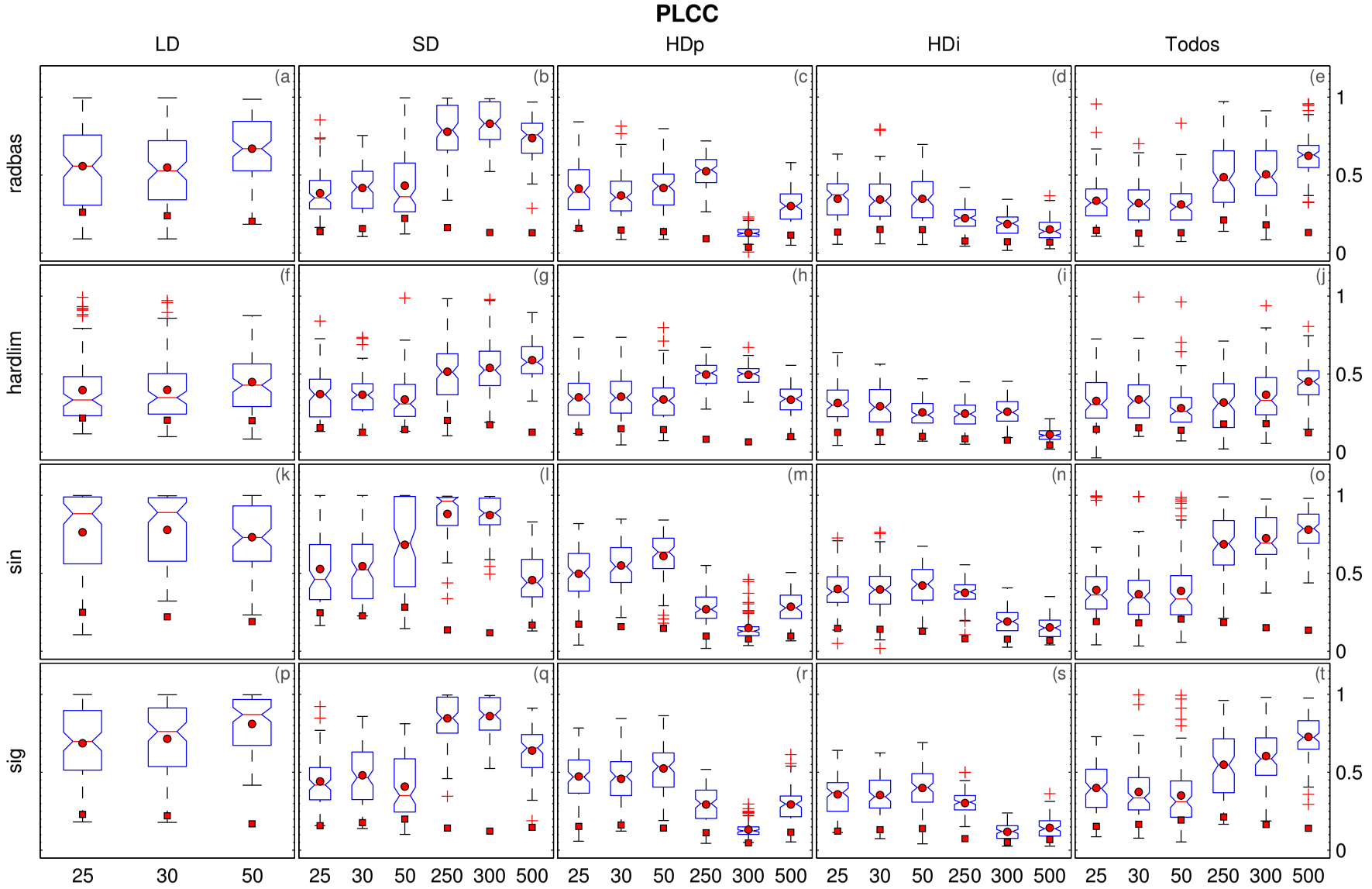

Figura 38: Comparação da acurácia (PLCC) do método NRVQA-ELM para os conteúdos do experimento D.

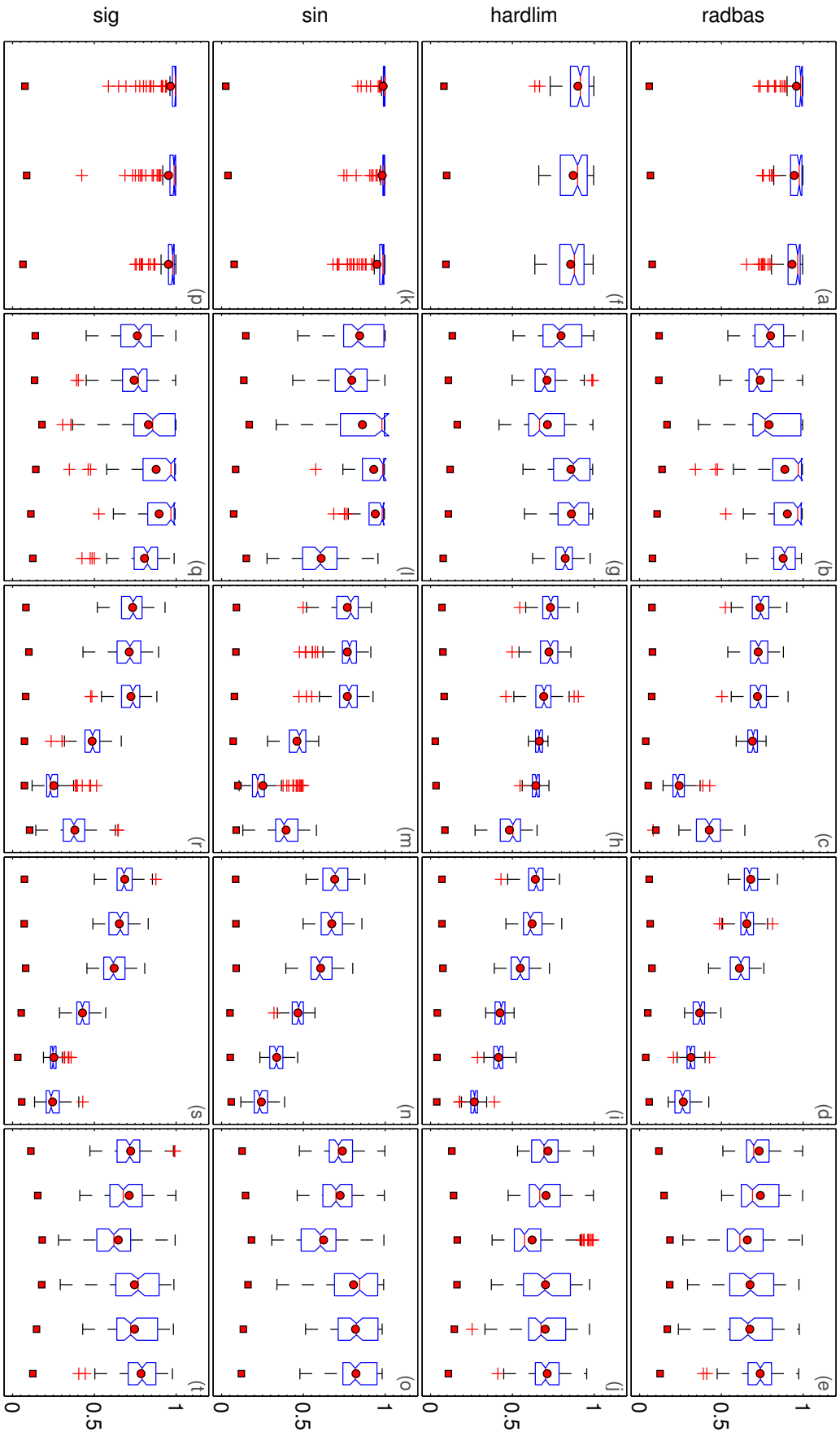

Figura 39: Comparação da acurácia (PLCC) do método NRVQA-ELMtc para os conteúdos do experimento D.

Tabela 9: Mediana do tempo de treinamento dos métodos NRVQA-ELM (1) e NRVQA-ELMtc (2), ambos usando a função de ativação seno para o experimento D.

Categoria	Métrica	fps (s)					
		25	30	50	250	300	500
LD	1	0,0183	0,0203	0,0162			
	2	0,7922	0,8011	0,9141			
SD	1	0,0749	0,0904	0,0972	0,2768	0,3909	0,8741
	2	5,2440	5,3269	5,7579	15,7164	18,6784	38,1168
HDp	1	0,0309	0,0299	0,0310	0,1113	0,0993	0,0133
	2	1,4361	1,4889	1,7053	5,9346	5,4848	2,0022
HDi	1	0,0441	0,0365	0,0481	0,1624	0,1511	0,0407
	2	2,0202	2,0672	2,3427	7,4142	8,6250	4,4043
Todos	1	0,2238	0,2638	0,2903	0,6278	0,6915	1,7171
	2	15,0703	15,6633	16,6005	41,0398	43,9675	98,5309

5.6 SÍNTESE DOS RESULTADOS EXPERIMENTAIS

As Figuras 40, 41 e 42 expressam os valores de significância estatística, segundo uma distribuição F percentual, determinada pela Fórmula (40), entre a métrica de referência completa MS-SSIM e os métodos propostos. As linhas sombreadas inferior e superior indicam os valores determinados pela iCDF, obtidos com a substituição de ζ das Fórmulas (41) e (44) nas Fórmulas (40) e (43), respectivamente. Além disso, as linhas sombreadas inferior e superior expressam os limiares de significância estatística negativo e positivo, respectivamente. Assim, quando o valor de F_p da Fórmula (43) na região de interesse do *box-plot* (*e.g.*, desvio-quartil, primeiro ou terceiro quartis, mediana ou média) é menor do que o valor da linha sombreada inferior, há uma diferença significativa negativa do método proposto em relação à métrica FR MS-SSIM, ou seja, o método proposto apresenta desempenho inferior à métrica MS-SSIM nessa região. Caso o valor de F_p da Fórmula (40), na região de interesse do *box-plot*, seja maior do que o valor representado pela linha sombreada superior, há uma diferença significativa positiva do método proposto em relação à métrica FR MS-SSIM, ou seja, o método proposto apresenta desempenho superior à métrica MS-SSIM nessa região. As Figuras 43-45 comparam a distribuição da acurácia entre os métodos propostos e a métrica MS-SSIM. Essas figuras sintetizam os resultados envolvendo os experimentos A, B e D. As Tabelas 10, 11 e 12 exibem a mediana (segundo quartil que compreende 50% dos dados) das distribuições da acurácia da predição de qualidade (PLCC) das Figuras 43, 44 e 45, respectivamente. Os valores em negrito nessas tabelas indicam as maiores medidas de acurácia para os pares treinamento-teste relacionados às categorias de vídeo desses experimentos. Os experimentos com os pares disjuntos $G1$-$G2$ e $G2$-$G2$, que representam cenários reais de aplicação, estão destacadas nas colunas dessas tabelas. Nesta seção, os métodos propostos NRVQA-ELM e NRVQA-ELMtc são comparados com a função de ativação sin.

As Figuras 40 e 41 expressam a medida F_p, entre a métrica FR MS-SSIM e os métodos propostos NRVQA-LM, NRVQA-ELM e NRVQA-ELMtc para os pares disjuntos $G1$-$G2$ e $G2$-$G1$ dos experimentos A e B, respectivamente. O método NRVQA-LM, em relação à métrica de referência completa MS-SSIM, expressa uma diferença significativa entre a mediana e o terceiro quartil nos conteúdos MPEG-2 desses experimentos, conforme mostram as Figuras 40-g e 41-j. O método NRVQA-ELMtc apresenta melhor desempenho do que a versão NRVQA-ELM nos experimentos A e B, conforme a comparação entre a segunda e terceira colunas das Figuras 40 e 41, respectivamente. Os limiares inferior e superior dessas figuras, representados pelas linhas sombreadas inferior e superior, mostram que o método NRVQA-ELMtc com a função de ativação sin apresenta uma diferença significativa em relação à métrica FR MS-SSIM no ter-

ceiro quartil para o conteúdo IP do experimento B, conforme mostra a Figura 41-i. Entretanto, o método NRVQA-LM expressa menor desempenho do que a métrica FR MS-SSIM nos conteúdos IP, *Wireless* e "Todos" do experimento A. Nos demais conteúdos dos experimentos A e B, considerando a média e a mediana, não há diferenças significativas negativas ou positivas da métrica FR MS-SSIM em relação aos métodos NRVQA-LM e NRVQA-ELMtc, *i.e.*, segundo uma distribuição F de Snedecor (VEERARAJAN, 2008), esses métodos propostos apresentam desempenho equivalente à métrica FR MS-SSIM na predição de qualidade de vídeo.

A Figura 42 mostra a medida F_p da métrica FR MS-SSIM em relação aos métodos propostos NRVQA-ELM e NRVQA-ELMtc, ambos usando a função de ativação sin, para os conteúdos do experimento D. Analogamente aos experimentos A, B e C, O método NRVQA-ELMtc expressa maior desempenho do que a versão NRVQA-ELM. O método NRVQA-ELMtc com a função de ativação sin apresenta maior desempenho do que a métrica FR MS-SSIM em qualquer região de interesse do *box-plot* no conteúdo LD da Figura 42-b e nos conteúdos SD e "Todos", ambos com $\tilde{N} > 50$, conforme as Figuras 42-d e 42-j. Considerando a média e mediana, o método NRVQA-ELMtc usando a função de ativação sin apresenta desempenho equivalente à métrica FR MS-SSIM no conteúdo em HDi, conforme a Figura 42-h. No conteúdo em HDp, o método NRVQA-ELMtc com a função de ativação sin expressa um desempenho equivalente à métrica FR MS-SSIM para $\tilde{N} < 250$ e $\tilde{N} = 300$, conforme ilustra a Figura 42-f.

O método NRVQA-LM depende do treinamento dos parâmetros β que estão intrinsecamente relacionados com as características do conteúdo da base de dados usada no treinamento. Logo, o método NRVQA-LM nos pares treinamento-teste disjuntos do experimento A, conforme Figura 43, apresenta maior desempenho (amplitude interquartílica) do que a métrica de referência completa MS-SSIM no conteúdo MPEG-2 da base de dados LIVE, conforme a comparação entre as Figuras 43-i e 43-j. No experimento B, com os pares treinamento-teste disjuntos, esse método apresenta um desempenho maior do que a métrica FR MS-SSIM nos conteúdos IP e MPEG-2, conforme a comparação entre os resultados das Figuras 44-(i/m) e 44-(j/n). Além disso, o método NRVQA-LM nos conteúdos Dirac e H.264 expressa um intervalo interquartílico próximo ao que é observado na métrica FR MS-SSIM nesses conteúdos, conforme mostra as Figuras 44-(a/e) e 44-(b/f). Entretanto, quando o treinamento desse método envolve todos os vídeos com quatro diferentes degradações há uma queda no seu desempenho, conforme as distribuições da acurácia ilustradas nas Figuras (43/44)-r.

Os resultados obtidos pelo método NRVQA-ELMtc apresentam maior desempenho do que o método NRVQA-ELM nos experimentos realizados nesta tese, sobretudo quando são usados os pares treinamento-teste disjuntos, conforme Figuras 43, 44 e 45. O método NRVQA-

ELMtc com a função de ativação sin no experimento A, usando pares de treino e de teste disjuntos, apresenta uma amplitude interquartílica próxima à métrica FR MS-SSIM nos conteúdos H.264, IP e MPEG-2, conforme inspeção visual das Figuras 43-(a/e/i) e Figuras 43-(d/h/l). No experimento B, também considerando os pares treinamento-teste disjuntos e a função de ativação sin, esse método apresenta maior desempenho do que a métrica de referência completa MS-SSIM no conteúdo IP, conforme as Figuras 44-i e Figuras 44-l. Além disso, o método NRVQA-ELMtc (sin) expressa uma amplitude interquartílica próxima àquela observada na métrica FR MS-SSIM nos conteúdos H.264 e MPEG-2, conforme ilustram as Figuras 44-(e/m) e Figuras 44-(h/p).

A Figura 45 sintetiza o experimento D, o qual representa um cenário real de aplicação, em que o conjunto de teste não está contido no treinamento. Neste experimento, o método NRVQA-ELMtc apresenta maior desempenho do que a versão NRVQA-ELM, independentemente da quantidade de neurônios usada na camada oculta e do número de amostras de teste. Assim, o método proposto NRVQA-ELMtc com a função de ativação sin, apresenta maior desempenho do que a métrica FR MS-SSIM nos conteúdos LD e SD, conforme mostram as Figuras 45-(a/d) e 45-(c/f). Destaca-se a compactação da amplitude interquartílica da acurácia (PLCC $\cong 1$) desse método no conteúdo LD da Figura 45-c. Este resultado mostra a eficiência do critério de parada proposto no Algoritmo 1 da Seção 4.2. As métricas FR apresentam uma tendência de queda de desempenho na predição da qualidade quando são formados grupos de teste com muitas amostras de vídeos aleatórias, analogamente ao que ocorre no experimento C, em que os resultados de PLCC dessas métricas são inferiores a 0,17, conforme descreve a Figura 28 da Seção 5.4. Essa tendência também ocorre nesse experimento, nos conteúdos com fps > 50 para a métrica MS-SSIM, conforme mostra a primeira coluna da Figura 45. Entretanto, o método NRVQA-ELMtc com a função de ativação sin, apresenta uma amplitude interquartílica da acurácia confinada entre 0,75 e 0,85 para o conteúdo HDp e no conteúdo em HDi essa amplitude está confinada entre 0,55 e 0,80, ambos os conteúdos com fps e $\tilde{N}$ entre 25 e 50, conforme mostram as Figuras 45-i e 45-l, respectivamente.

Tal como discutido na Seção 5.5, quando o número de amostras de teste e de neurônios na camada oculta são superiores à quantidade de amostras de treinamento, *i.e.*, fps $> N$ e $\tilde{N} > N$, ocorre um decréscimo na acurácia do método NRVQA-ELMtc, conforme mostram as Figuras 45-i e 45-l nos conteúdos em HDp e HDi, respectivamente. A análise de resultados da distribuição da acurácia do método NRVQA-ELMtc com a função de ativação sin para o conteúdo "Todos", quando são usadas amostras de vídeo embaralhadas, mostra o potencial do método NRVQA-ELMtc na avaliação de qualidade de vídeo digital, em que seu desempenho é superior à métrica de referência completa MS-SSIM, cuja amplitude interquartílica do método

NRVQA-ELMtc está confinada no intervalo [0,75; 0,95] quando $\tilde{N} > 50$, conforme descreve a Figura 45-o.

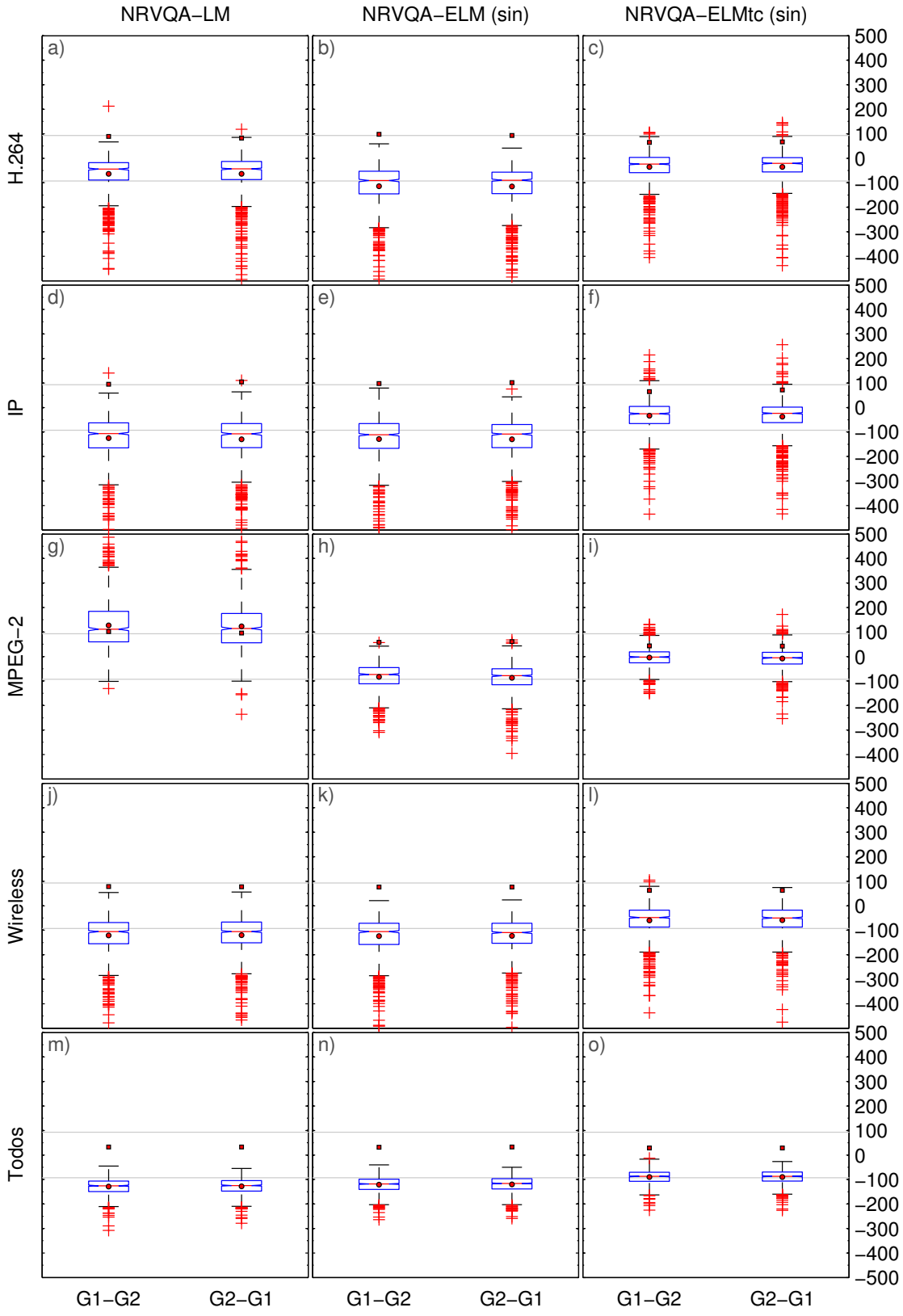

Figura 40: Comparação da distribuição F percentual (F_p) entre a métrica MS-SSIM e os métodos NRVQA-LM, NRVQA-ELM (sin) e NRVQA-ELMtc (sin) para os conteúdos do experimento A.

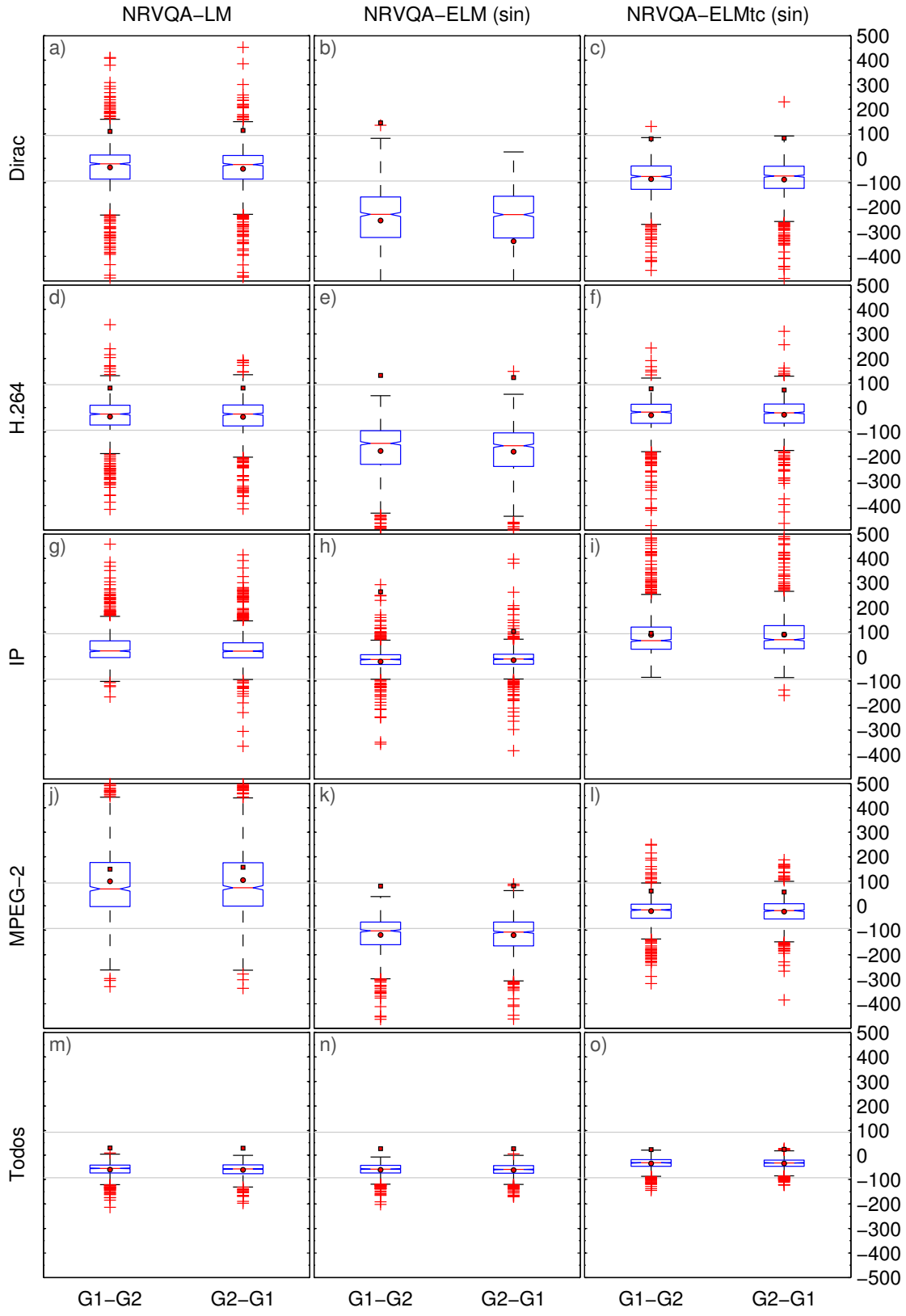

Figura 41: Comparação da distribuição F percentual (F_p) entre a métrica MS-SSIM e os métodos NRVQA-LM, NRVQA-ELM (sin) e NRVQA-ELMtc (sin) para os conteúdos do experimento B.

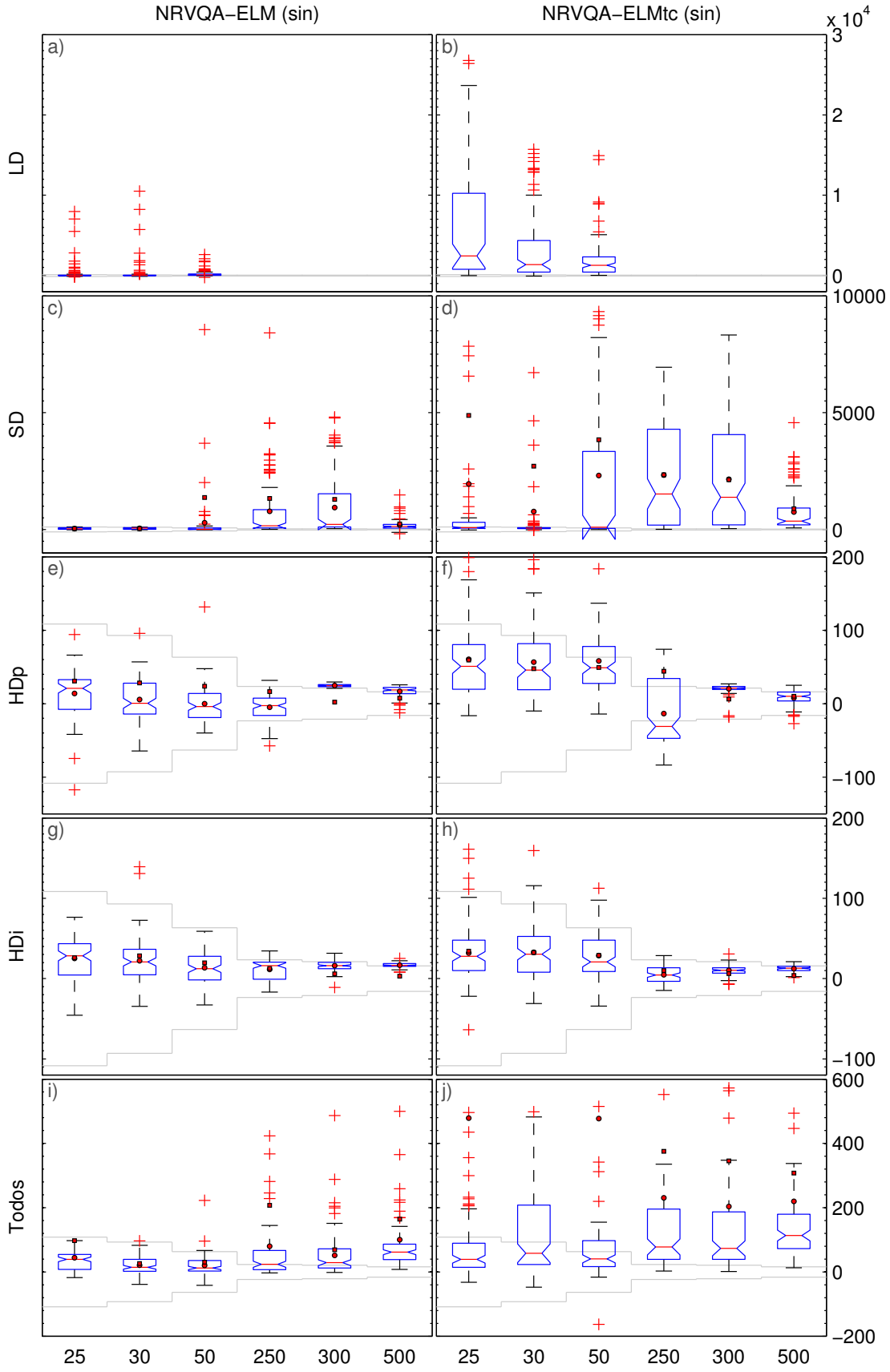

Figura 42: Comparação da distribuição F percentual (F_p) entre a métrica MS-SSIM e os métodos NRVQA-ELM (sin) e NRVQA-ELMtc (sin) para os conteúdos do experimento D.

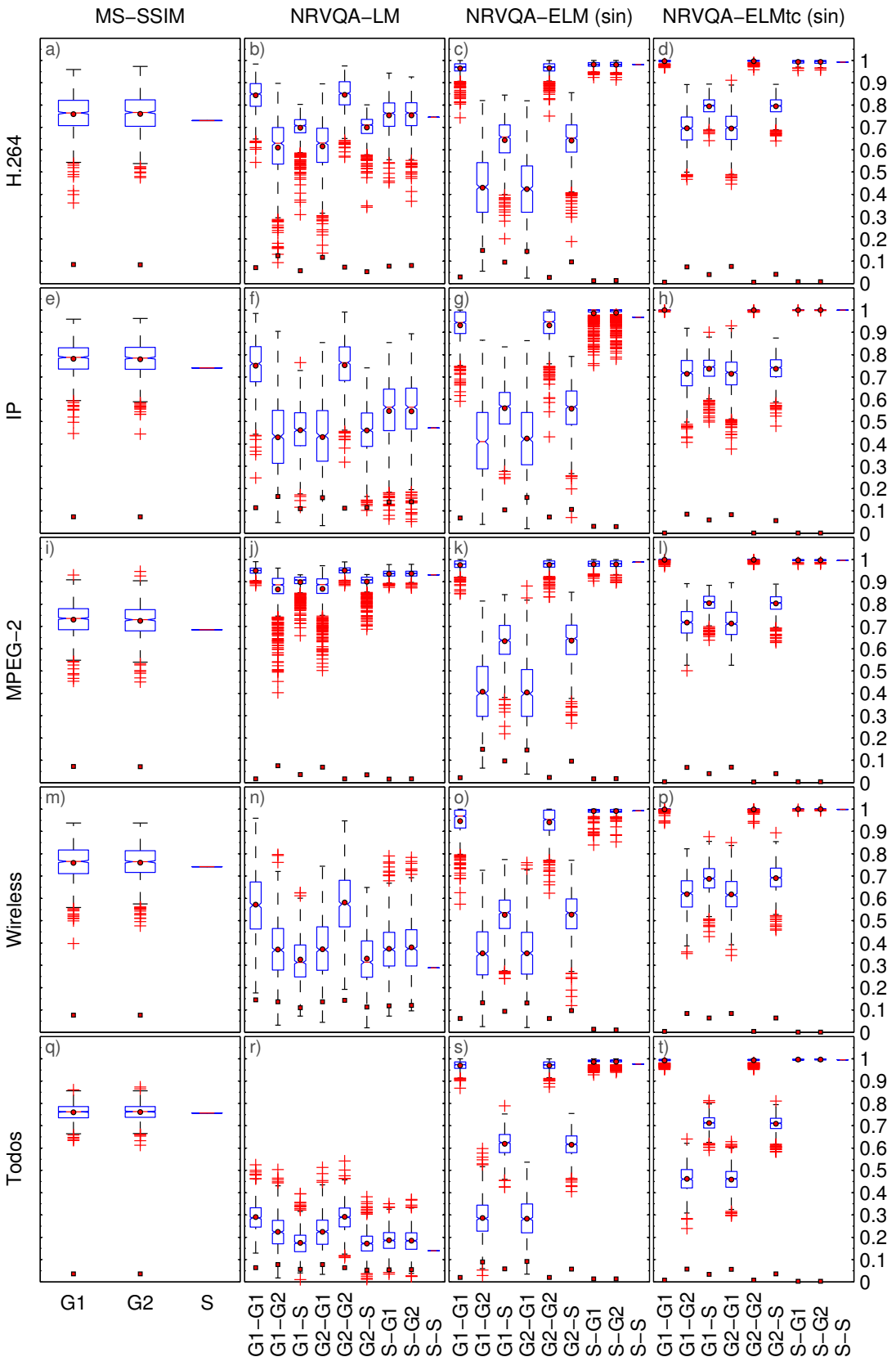

Figura 43: Comparação da acurácia (PLCC) entre a métrica MS-SSIM e os métodos NRVQA-LM, NRVQA-ELM (sin) e NRVQA-ELMtc (sin) para os conteúdos do experimento A.

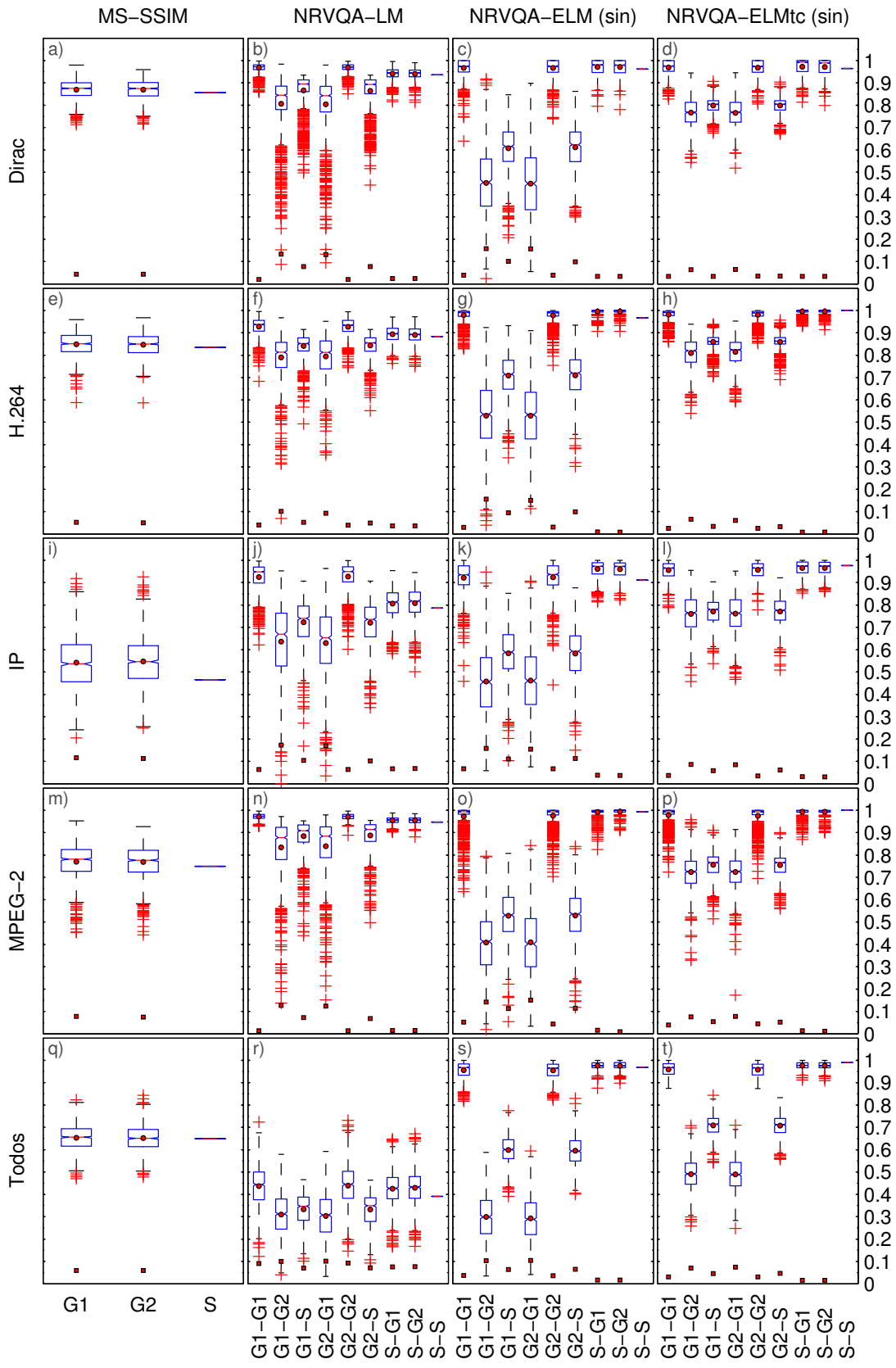

Figura 44: Comparação da acurácia (PLCC) entre a métrica MS-SSIM e os métodos NRVQA-LM, NRVQA-ELM (sin) e NRVQA-ELMtc (sin) para os conteúdos do experimento B.

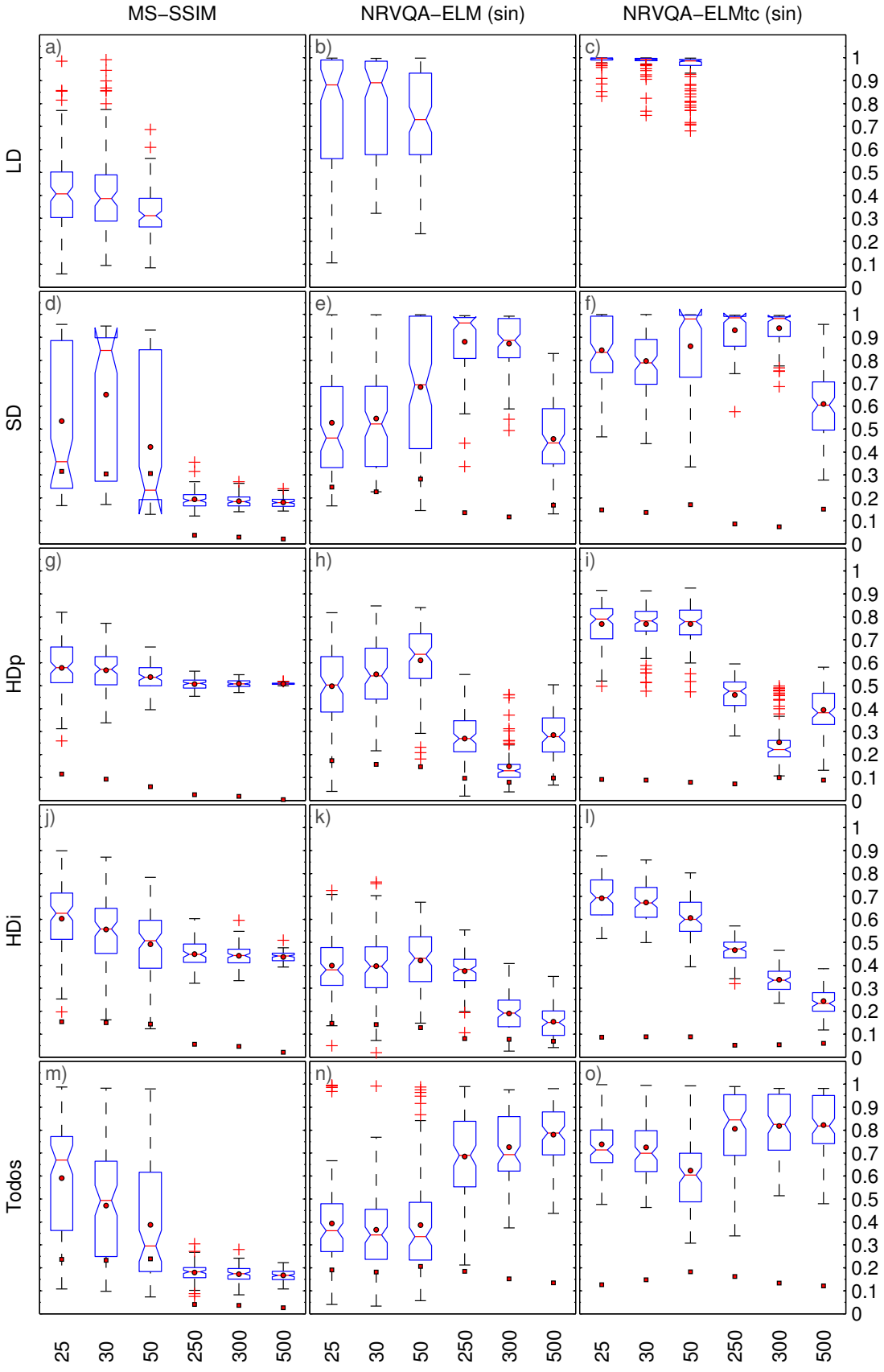

Figura 45: Comparação da acurácia (PLCC) entre a métrica MS-SSIM e os métodos NRVQA-ELM (sin) e NRVQA-ELMtc (sin) para os conteúdos do experimento D.

O método NRVQA-LM, a partir da mediana da distribuição das Figuras 43 e 44 nos pares treinamento-teste disjuntos, expressa uma acurácia superior à métrica FR MS-SSIM no conteúdo MPEG-2 do experimento A (Tabela 10) e nos conteúdos MPEG-2 e IP do experimento B (Tabela 11). Embora nos experimentos (A e B) há amostras de vídeo com conteúdos em IP, seus artefatos foram gerados artificialmente e com intensidades diferentes. Na base de dados LIVE (SESHADRINATHAN *et al.*, 2010), a simulação da perda de pacotes foi realizada com taxas mais altas (3%, 5%, 10% e 20%), enquanto que na base de dados IVP (LI; MA, 2012) as taxas são mais moderadas (0,1%, 0,5%, 1%, 3% e 5%). Assim, como fora já mencionado anteriormente, os artefatos provenientes de compressão causam efeitos altamente estruturados, enquanto que os artefatos oriundos de erros de transmissão, e.g., *wireless* e IP, causam efeitos não estruturados, devido à intensidade e variação aleatórias da taxa de perdas de dados, cujo modelo sigmoidal proposto na Equação (77) não consegue realizar uma predição de qualidade com o mesmo desempenho observado nos conteúdos em H.264, MPEG-2 e Dirac.

O método NRVQA-ELMtc no experimento A, com a função de ativação sin, apresenta uma mediana da acurácia superior à versão NRVQA-ELM. A Tabela 10 destaca os pares de treinamento-teste disjuntos $G1$-$G2$ e $G2$-$G1$, em que o método proposto NRVQA-ELMtc expressa medidas de PLCC próximas àquelas observadas na métrica FR MS-SSIM nos conteúdos H.264, IP e MPEG-2. No entanto, no experimento B, esse método usando a função de ativação sin nos pares de treino e de teste disjuntos, apresenta uma mediana da acurácia superior a da métrica FR MS-SSIM no conteúdo IP e valores dessa medida próximos aos da métrica FR MS-SSIM nos conteúdos H.264, e MPEG-2, conforme inspeção visual da Tabela 11.

A mediana da distribuição da acurácia do experimento D estão descritos na Tabela 12, em que o método NRVQA-ELMtc apresenta melhor desempenho do que a métrica FR MS-SSIM nos conteúdos em LD, SD (exceto para fps = 30) e "Todos", bem como nos conteúdos em HD (HDp e HDi) para fps < 250. Pois, nestes conteúdos com fps > 250 há menos amostras de treinamento do que de teste, e.g., na categoria HDp há 520 vídeos; caso fps seja igual a 500 amostras de teste que também equivale ao mesmo número de neurônios na camada oculta ($\tilde{N} = 500$), apenas 20 amostras de vídeo são utilizadas no treinamento. Por fim, o método NRVQA-ELMtc mostra maior desempenho do que a métrica FR MS-SSIM e o método NRVQA-ELM na categoria "Todos" do experimentos D, independente dos valores de fps, conforme mostra a Tabela 12.

Tabela 10: Mediana da distribuição da acurácia (PLCC) da métrica MS-SSIM (1) e dos métodos NRVQA-LM (2), NRVQA-ELM (3), NRVQA-ELMtc (4) para o experimento A.

Categoria	Métrica	Treinamento-Teste								
		*G*1-*G*1	*G*1-*G*2	*G*1-*S*	*G*2-*G*1	*G*2-*G*2	*G*2-*S*	*S*-*G*1	*S*-*G*2	S-S
H.264	1*	0,7647	**0,7671**	0,7305	**0,7647**	0,7671	0,7305	0,7647	0,7671	0,7305
	2	0,8497	0,6286	0,7080	0,6300	0,8528	0,7110	0,7637	0,7648	0,7451
	3⋆	0,9700	0,4329	0,6566	0,4242	0,9700	0,6518	0,9823	0,9823	0,9819
	4⋆	**0,9994**	0,6963	**0,7977**	0,6992	**0,9996**	**0,7965**	**0,9972**	**0,9968**	**0,9922**
IP	1*	0,7886	**0,7852**	0,7412	**0,7886**	0,7852	0,7412	0,7886	0,7852	0,7412
	2	0,7608	0,4337	0,4621	0,4371	0,7641	0,4612	0,5641	0,5645	0,4715
	3⋆	0,9444	0,4102	0,5682	0,4212	0,9480	0,5662	**1,0000**	**1,0000**	0,9678
	4⋆	**1,0000**	0,7186	**0,7457**	0,7211	**1,0000**	**0,7425**	**1,0000**	**1,0000**	**1,0000**
MPEG-2	1*	0,7357	0,7300	0,6851	0,7357	0,7300	0,6851	0,7357	0,7300	0,6851
	2	0,9506	**0,8862**	**0,9079**	**0,8875**	0,9514	**0,9095**	0,9367	0,9380	0,9317
	3⋆	0,9798	0,3999	0,6377	0,3987	0,9813	0,6462	0,9811	0,9805	0,9877
	4⋆	**1,0000**	0,7188	0,8091	0,7112	**1,0000**	0,8057	**0,9967**	**0,9967**	**0,9957**
Wireless	1*	0,7653	**0,7642**	**0,7418**	**0,7653**	0,7642	**0,7418**	0,7653	0,7642	0,7418
	2	0,5686	0,3683	0,3135	0,3695	0,5741	0,3139	0,3713	0,3745	0,2886
	3⋆	0,9688	0,3512	0,5362	0,3544	0,9539	0,5351	0,9941	0,9906	0,9929
	4⋆	**1,0000**	0,6230	0,6899	0,6151	**1,0000**	0,6923	**1,0000**	**1,0000**	**0,9991**
Todos	1*	0,7622	**0,7630**	**0,7570**	**0,7622**	0,7630	**0,7570**	0,7622	0,7630	0,7570
	2	0,2858	0,2232	0,1743	0,2250	0,2899	0,1722	0,1857	0,1845	0,1392
	3⋆	0,9740	0,2839	0,6248	0,2823	0,9731	0,6173	0,9919	0,9924	0,9762
	4⋆	**0,9952**	0,4606	0,7130	0,4618	**0,9955**	0,7114	**0,9955**	**0,9977**	**0,9960**

* Não há grupos de treinamento, validação apenas com os grupos de teste *G*1, *G*2 e S.

⋆ Função de ativação seno.

Tabela 11: Mediana da distribuição da acurácia (PLCC) da métrica MS-SSIM (1) e dos métodos NRVQA-LM (2), NRVQA-ELM (3), NRVQA-ELMtc (4) para o experimento B.

Categoria	Métrica	Treinamento-Teste								
		*G*1-*G*1	*G*1-*G*2	*G*1-*S*	*G*2-*G*1	*G*2-*G*2	*G*2-*S*	*S*-*G*1	*S*-*G*2	S-S
Dirac	1*	0,8753	**0,8746**	0,8567	**0,8753**	0,8746	0,8567	0,8753	0,8746	0,8567
	2	0,9727	0,8443	**0,8947**	0,8431	0,9715	**0,8928**	0,9439	0,9439	0,9362
	3⋆	0,9738	0,4556	0,6161	0,4497	**0,9745**	0,6237	0,9805	0,9792	0,9617
	4⋆	**0,9754**	0,7680	0,8046	0,7686	0,9735	0,8046	**0,9926**	**0,9889**	**0,9650**
H.264	1*	0,8517	**0,8504**	0,8347	**0,8517**	0,8504	0,8347	0,8517	0,8504	0,8347
	2	0,9376	0,8133	0,8523	0,8126	0,9349	0,8550	0,8964	0,8931	0,8841
	3⋆	0,9912	0,5361	0,7158	0,5342	0,9900	0,7206	**0,9995**	0,9995	0,9682
	4⋆	**0,9913**	0,8204	**0,8620**	0,8241	**0,9908**	**0,8629**	**0,9995**	**0,9996**	**0,9997**
IP	1*	0,5361	0,5466	0,4664	0,5361	0,5466	0,4664	0,5361	0,5466	0,4664
	2	0,9485	0,6694	0,7398	0,6531	0,9501	0,7359	0,8133	0,8180	0,7860
	3⋆	0,9352	0,4552	0,5887	0,4610	0,9368	0,5939	0,9735	0,9728	0,9116
	4⋆	**0,9654**	**0,7647**	**0,7801**	**0,7637**	**0,9660**	**0,7763**	**0,9741**	**0,9740**	**0,9766**
MPEG-2	1*	0,7812	0,7768	0,7494	0,7812	0,7768	0,7494	0,7812	0,7768	0,7494
	2	0,9737	**0,8782**	**0,9094**	**0,8845**	0,9730	**0,9144**	0,9568	0,9558	0,9473
	3⋆	0,9934	0,4131	0,5312	0,4020	0,9939	0,5362	0,9972	**1,0000**	0,9938
	4⋆	**0,9939**	0,7258	0,7663	0,7264	**0,9940**	0,7662	**0,9989**	0,9981	**1,0000**
Todos	1*	0,6561	**0,6506**	0,6494	**0,6561**	0,6506	0,6494	0,6561	0,6506	0,6494
	2	0,4441	0,3128	0,3467	0,3081	0,4457	0,3491	0,4270	0,4327	0,3909
	3⋆	**0,9681**	0,2994	0,6007	0,2881	0,9656	0,5937	0,9765	0,9752	0,9693
	4⋆	0,9679	0,4919	**0,7140**	0,4883	**0,9664**	**0,7118**	**0,9773**	**0,9760**	**0,9915**

* Não há grupos de treinamento, validação apenas com os grupos de teste *G*1, *G*2 e S.

⋆ Função de ativação seno.

Tabela 12: Mediana da distribuição da acurácia (PLCC) da métrica MS-SSIM (1) e dos métodos NRVQA-ELM (2), NRVQA-ELMtc (3), ambos usando a função de ativação seno nos experimentos D.

Categoria	Métrica	Número de Amostras de Teste e de Neurônios na Camada Oculta (fps)					
		25	30	50	250	300	500
LD	1	0,4067	0,3866	0,3117			
	2	0,8816	0,8904	0,7304			
	3	**0,9967**	**0,9930**	**0,9856**			
SD	1	0,3575	**0,8431**	0,2334	0,1887	0,1839	0,1798
	2	0,4607	0,5222	0,6934	0,9621	0,8863	0,4392
	3	**0,8345**	0,7885	**0,9802**	**0,9849**	**0,9823**	**0,6039**
HDp	1	0,5789	0,5715	0,5362	**0,5099**	**0,5082**	**0,5082**
	2	0,5031	0,5432	0,6368	0,2700	0,1291	0,2778
	3	**0,7907**	**0,7826**	**0,7791**	0,4762	0,2217	0,3825
HDi	1	0,6264	0,5586	0,5074	0,4481	**0,4426**	**0,4403**
	2	0,3806	0,3952	0,4305	0,3817	0,1915	0,1507
	3	**0,6943**	**0,6723**	**0,6002**	**0,4717**	0,3354	0,2335
Todos	1	0,6702	0,4938	0,2957	0,1826	0,1743	0,1668
	2	0,3613	0,3434	0,3362	0,6898	0,6930	0,7874
	3	**0,7140**	**0,6991**	**0,6045**	**0,8451**	**0,8248**	**0,8180**

5.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo apresentou e discutiu os resultados experimentais, comparando o desempenho dos métodos propostos com as métricas de referência completa PSNR, SSIM, MS-SSIM e a métrica sem referência JPEG-NR. O método NRVQA-LM é fortemente dependente dos parâmetros β, *i.e.*, o seu desempenho está relacionado com um treinamento especializado, ou seja, depende do tipo e intensidade dos artefatos (distorções) presentes nas bases de dados usadas. Assim, o método NRVQA-LM implementado pelo modelo sigmoidal da Equação (77), apresenta maior desempenho na predição de qualidade em vídeos contendo distorções provenientes do processo de compressão, *e.g.*, artefatos de compressão Dirac, H.264/AVC e MPEG-2. O método NRVQA-ELMtc apresenta vantagens na predição da qualidade, devido à capacidade de generalização da RNA SLFN, cujo algoritmo de aprendizado ELM associado a um simples critério de parada não-direcionado produz resultados superiores às métricas FR quando são usadas muitas amostras de vídeo com características distintas. O experimento D ilustra essa situação, com destaque para a Figura 45, a qual compara o desempenho da acurácia entre o método proposto NRVQA-ELMtc e a métrica FR MS-SSIM. Dessa forma, para a categoria "Todos", em que todas as resoluções de vídeo (LD, SD, HDi e HDp) são embaralhadas, o método NRVQA-ELMtc, em termos da média e mediana da distribuição da acurácia (PLCC), expressa desempenho superior à métrica de referência completa MS-SSIM. Logo, esse método mostra um grande potencial na tarefa de predição de qualidade de vídeos digitais submetidos a distorções de diversas naturezas, pois ele não requer um treinamento da RNA SLFN sobre tipos específicos de artefatos. O próximo capítulo finaliza esta tese com as conclusões finais e apresenta os trabalhos futuros.

6 CONCLUSÃO E TRABALHOS FUTUROS

Esta tese apresenta dois métodos NR para avaliação de qualidade de vídeo baseados em características espaço-temporais. O primeiro método (NRVQA-LM) é baseado em uma modelagem sigmoidal com solução de mínimos quadrados que usa o algoritmo LM. O segundo método (NRVQA-ELM) utiliza uma RNA SLFN, cujo aprendizado é baseado no algoritmo ELM. Uma versão estendida desse algoritmo também foi proposta no método NRVQA-ELMtc que busca de maneira iterativa os melhores parâmetros da RNA, para que seja obtida a melhor correlação possível entre os escores objetivos e subjetivos, *i.e.*, o método NRVQA-ELMtc por meio de uma busca não direcionada, localiza um ponto de mínimo, em k iterações, que está associado à maior correlação entre a predição de qualidade de vídeo e a percepção do SVH.

Os objetivos traçados na Seção 1.3 foram todos atingidos. As contribuições da tese estão enumeradas na Seção 1.5, contudo, destaca-se o desenvolvimento dos métodos propostos e a metodologia adotada no processo de validação cruzada dos experimentos A, B e D com a análise da distribuição estatística das medidas recomendadas pelo VQEG (descritas na Seção 3.2) com o recurso do *box-plot*. A maioria dos trabalhos sobre abordagens NRVQA utiliza validação cruzada com apenas uma variação, análogo ao experimento C, *i.e.*, sem permutações das amostras de vídeo entre os grupos de treinamento e teste. Os experimentos A-B e D, conforme a Figura 18, foram conduzidos com k igual a 10^3 e 10^2 pares de treinamento-teste distintos e aleatórios.

Tipicamente, quando são utilizadas poucas amostras de vídeo, os mapeamentos cúbico e logístico apresentam diferenças menos expressivas, devido à componente temporal durante o processo de avaliação subjetiva, enquanto o mapeamento cúbico em imagens apresenta maior desempenho, conforme comparação com a DMOS e DMOSp(PSNR) entre as Figuras 9-a e 9-b. Embora os mapeamentos cúbico e logístico, quando aplicados em escores objetivos de vídeos, apresentem diferenças menos expressivas do que em imagens, a comparação entre os resultados obtidos e os reportados na literatura não é justa, devido ao uso de uma função de mapeamento diferente da função polinomial cúbica e as abordagens de validação cruzada (análise estatística da distribuição de dados) adotadas na tese. As recomendações mais recentes do

VQEG (VQEG, 2008, 2009, 2010) sugerem o uso da função de mapeamento não linear polinomial cúbica. Entretanto, a maioria dos trabalhos encontrados na literatura (WANG *et al.*, 2002; SESHADRINATHAN *et al.*, 2010; YAO *et al.*, 2012) utiliza recomendações do VQEG já ultrapassadas (VQEG, 2000, 2003), as quais recomendam que o mapeamento entre os escores objetivos e subjetivos seja realizado por meio de uma função não linear monotônica logística.

Os resultados experimentais apresentados no Capítulo 5 mostram que os métodos propostos NRVQA-LM e NRVQA-ELMtc, nos conjuntos treinamento-teste disjuntos, os quais representam aplicações práticas, possuem acurácia (PLCC) na predição de qualidade de vídeo superior às métricas de referência completa nos conteúdos MPEG-2 e IP. Nos conteúdos H.264 e Dirac, esses métodos apresentam desempenho próximo ao da métrica de referência completa MS-SSIM, conforme descrevem as Figuras 43 e 44, cujas medianas (segundos quartis) estão expressas nas Tabelas 10 e 11, respectivamente. O desempenho do método NRVQA-LM está associado ao treinamento dos parâmetros β. Logo, quando as amostras de treinamento apresentam distorções com intensidades e variações aleatórias, *e.g.*, conteúdos IP e *Wireless* da base de dados LIVE, a correlação entre os escores objetivos do método NRVQA-LM e os escores subjetivos diminui. O mesmo ocorre quando são utilizadas todas as amostras de vídeo (Figuras 43-r e 44-r). Entretanto, esse método apresenta melhor desempenho quando são usadas amostras de vídeo contendo artefatos de compressão, *e.g.*, H.264, MPEG-2 e Dirac, devido ao efeito altamente estruturado causado pelo processo de compressão (WU *et al.*, 2007). O método NRVQA-ELMtc apresenta maior desempenho do que a versão NRVQA-ELM nos experimentos realizados nesta tese.

Logo, a contribuição pela proposta no Algoritmo 1 da Seção 4.2 mostrou-se eficiente em comparação com a versão clássica do algoritmo ELM. Embora, a busca pelos melhores parâmetros da RNA (β, $\mathbf{w}$ e b) não seja direcionada, *i.e.*, limitada ao critério de parada do Algoritmo 1, o método NRVQA-ELMtc apresenta resultados superiores aos das métricas de referência completa PSNR, SSIM, MS-SSIM e sem referência JPEG-NR nos experimentos C e D, conforme a comparação entre as Figuras 28 e 30 (experimento C) e as Figuras 37 e 39 (experimento D), respectivamente. No experimento D, o método NRVQA-ELMtc apresenta um decréscimo na predição da acurácia nos conteúdos em HD (HDi e HDp) com o aumento no número de amostras de teste e de neurônios na camada oculta, devido ao número de amostras de treinamento ser menor do que a quantidade de amostras de teste e de neurônios na camada oculta, conforme as amplitudes interquartílicas da terceira e quarta colunas da Figura 39. Analogamente, o experimento C também exibe um decréscimo na predição da acurácia nos pares de treinamento-teste disjuntos com o crescimento do número de neurônios na camada oculta, *i.e.*, para $\tilde{N} > 400$, conforme as Figuras 30-b e 30-d.

Assim, baseado nos resultados experimentais, os métodos propostos podem ser aplicados no monitoramento de qualidade de vídeo em sistemas de radiodifusão, tais como TVD e IPTV que utilizam os padrões de compressão e codificação Dirac, H.264/AVC e MPEG-2.

Como trabalhos futuros sugere-se a utilização de outras características espaciais e temporais, *e.g.*, parâmetro de informação perceptual espacial SI (ITU-T P.910, 1999), bem como a inclusão das componentes de crominância nos métodos NRVQA. Características dos quadros de vídeo poderiam ser extraídas no domínio da frequência, *e.g.*, no domínio da transformada DCT ou *wavelet*. Uma outra abordagem poderia ser implementada com a inclusão de informações acerca do fluxo de *bits*, *e.g.*, estimadores de movimento, GoP, taxa de compressão, QP, perfis e níveis de codificação. Além disso, as características extraídas das amostras de vídeos poderiam ser pré-processadas ou pós-processadas para realçar as distorções (artefatos). O método proposto NRVQA-LM poderia ser comparado com outras técnicas para solução de mínimos quadrados, *e.g.*, métodos bayesiano (DENISON *et al.*, 2002), DFP (Davidon-Fletcher-Powell *method*) e BFGS (Broyden-Fletcher-Goldfarb-Shanno *method*) (RAO, 2009). Em janeiro de 2013, os autores do algoritmo ELM, disponibilizaram o código da versão OS-ELM (*Online Sequencial Extreme Learning Machine*) (HUANG, 2013), que realiza o treinamento dos dados um a um (*one-by-one*) ou em bloco (*chunk-by-chunk*) com tamanho fixo ou variável (LIANG *et al.*, 2006a), cuja abordagem poderia ser aplicada no experimento D, em que são utilizados blocos de amostras de vídeo (fps). A métrica proposta NRVQA-ELMtc poderia ser comparada com outras abordagens em RNAs, tais como BP, RBF e SVM. A validação dos métodos propostos poderia ser complementada com a utilização de outras métricas FR, tais como FSIM (*Feature Similarity index*) (ZHANG *et al.*, 2011), VQM (*Video Quality Metric*) (PINSON; WOLF, 2004) e MOVIE (SESHADRINATHAN; BOVIK, 2010).

Uma outra abordagem interessante para o desenvolvimento e validação dos métodos propostos poderia ser feita com a elaboração de experimentos subjetivos, a partir de uma base de dados que contemple diversas distorções e variações de TI e SI. Também sugere-se um estudo comparativo dos resultados obtidos com diversas funções de mapeamento, *e.g.*, funções polinomiais e exponenciais, tal qual fora desenvolvido em (ENGELKE *et al.*, 2009). Por fim, como extensão deste trabalho, sugere-se a aplicação dos métodos propostos NRVQA-LM e NRVQA-ELMtc como suporte na correção ou redução de distorções (artefatos) em vídeo digital, cujos algoritmos poderiam ser implementados em dispositivos como *set-top boxes*, *ultrabooks*, *tablets*, *smartphones* e em equipamentos WiDi. Esses métodos poderiam ser usados na identificação de distorções pela indicação do nível de qualidade do vídeo, e, em seguida, técnicas de correção ou suavização de artefatos poderiam ser aplicadas, antes de exibir o conteúdo visual ao usuário final.

REFERÊNCIAS

ABRAMOWITZ, M.; STEGUN, I. A. **Handbook of mathematical functions with formulas, graphs, and mathematical tables**. 5. ed. Washington: U.S. Government Printing Office, 1964. xiv+1046 p. (National Bureau of Standards Applied Mathematics Series, v. 55).

BABU, R. V.; SURESH, S.; PERKIS, A. No-reference JPEG-image quality assessment using GAP-RBF. **Signal Processing**, v. 87, n. 6, p. 1493–1503, 2007.

BARKOWSKY, M. *et al.* Analysis of freely available dataset for HDTV including coding and transmission distortions. In: **Analysis of Freely Available Dataset for HDTV including Coding and Transmission Distortions**. Scottsdale, United States: [s.n.], 2010. Disponível em: <http://hal.archives-ouvertes.fr/hal-00463568>.

BARLAND, R.; SAADANE, A. A new reference free approach for the quality assessment of MPEG coded videos. In: **Proceedings of the 7th International Conference on Advanced Concepts for Intelligent Vision Systems**. Berlin, Heidelberg: Springer-Verlag, 2005. (ACIVS'05), p. 364–371.

BRANDÃO, T.; QUELUZ, M. P. No-reference quality assessment of H.264/AVC encoded video. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 20, n. 11, p. 1437–1447, November 2010.

BRUNNSTROM, K. *et al.* VQEG validation and ITU standardization of objective perceptual video quality metrics [standards in a nutshell]. **IEEE Signal Processing Magazine**, v. 26, n. 3, p. 96–101, 2009.

CALLET, P. L.; BARBA, D. Image quality assessment: from sites errors to a global appreciation of quality. In: **Proceedings of the Picture Coding Sympsium**. [S.l.: s.n.], 2001. p. 105–108.

CALLET, P. L.; VIARD-GAUDIN, C.; BARBA, D. A convolutional neural network approach for objective video quality assessment. **IEEE Transactions on Neural Networks**, v. 17, n. 5, p. 1316–1327, 2006.

CALYAM, P. *et al.* Multi-resolution multimedia QoE models for IPTV applications. **International Journal of Digital Multimedia Broadcasting**, v. 2012, n. ID 904072, p. 13 pages, 2012.

CANNY, J. A computational approach to edge detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-8, n. 6, p. 679–698, 1986.

CHOE, J.; LEE, K.; LEE, C. No-reference video quality measurement using neural networks. In: **Proceedings of the 16th International Conference on Digital Signal Processing**. Piscataway, NJ, USA: IEEE Press, 2009. (DSP'09), p. 1197–1200.

CHOI, H.; LEE, C. No-reference image quality metric based on image classification. **EURASIP Journal on Advances in Signal Processing**, v. 2011, p. 65, 2011.

CIANCIO, A. *et al.* No-reference blur assessment of digital pictures based on multifeature classifiers. **IEEE Transactions on Image Processing**, v. 20, n. 1, p. 64–75, January 2011.

COVER, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. **Electronic Computers, IEEE Transactions on**, EC-14, n. 3, p. 326–334, 1965.

DAUBECHIES, I. **Ten Lectures on Wavelets**. Philadelphia, USA: SIAM, 1992. (CBMS-NSF Regional Conference Series in Applied Mathematics, v. 61).

DEBIAN. **Debian Wheezy GNU/Linux 7.0**. 2012. Disponível em: <http://debian.org>.

DECHERCHI, S. *et al.* Circular-ELM for the reduced-reference assessment of perceived image quality. **Neurocomputing**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, v. 102, p. 78–89, February 2013.

DENISON, D. *et al.* **Bayesian Methods for Nonlinear Classification and Regression**. [S.l.]: John Wiley & Sons, 2002. (Wiley Series in Probability and Statistics, v. 386).

DING, W. *et al.* Image and video quality assessment using neural network and SVM. **Tsinghua Science and Technology Journal**, v. 13, n. 1, p. 112–116, February 2008.

DOERMANN, D. Unsupervised feature learning framework for no-reference image quality assessment. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12)**. Washington, USA: IEEE Computer Society, 2012. (CVPR'12), p. 1098–1105.

EDEN, A. No-reference image quality analysis for compressed video sequences. **IEEE Transactions on Broadcasting**, v. 54, n. 3, p. 691–697, 2008.

ENGELKE, U. *et al.* Reduced-reference metric design for objective perceptual quality assessment in wireless imaging. **Signal Processing: Image Communication**, v. 24, n. 7, p. 525–547, 2009.

ENGELKE, U.; ZEPERNICK, H.-J. An artificial neural network for quality assessment in wireless imaging based on extraction of structural information. In: **Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'07)**. Honolulu, USA: IEEE, 2007. p. 1249–1252.

ENGELKE, U.; ZEPERNICK, H.-J. Perceptual-based quality metrics for image and video services: a survey. In: **Proceedings EuroNGI Conference Next Generation Internet Networks**. [S.l.: s.n.], 2007. p. 190–197.

ESKICIOGLU, A. M.; FISHER, P. S. Image quality measures and their performance. **IEEE Transactions on Communications**, v. 43, n. 12, p. 2959–2965, 1995.

FERRARI, S.; STENGEL, R. Smooth function approximation using neural networks. **IEEE Transactions on Neural Networks**, v. 16, n. 1, p. 24 –38, January 2005.

GASTALDO, P.; ZUNINO, R. Neural networks for the no-reference assessment of perceived quality. **Journal of Electronic Imaging**, v. 14, n. 3, p. 033004, September 2005.

GASTALDO, P. *et al.* Objective quality assessment of displayed images by using neural networks. **Signal Processing: Image Communication**, v. 20, n. 7, p. 643–661, 2005.

GOLUB, G.; KAHAN, W. Calculating the singular values and pseudo-inverse of a matrix. **Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis**, v. 2, p. 205–224, 1965.

GOLUB, G. H.; REINSCH, C. Singular value decomposition and least squares solutions. **Numerische Mathematik**, v. 14, p. 403–420, 1970.

GONZáLEZ, A. R. *et al.* Cast: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator. **Expert Systems with Applications**, v. 38, n. 9, p. 11489–11500, 2011.

GU, K. *et al.* No-reference stereoscopic IQA approach: from nonlinear effect to parallax compensation. **Journal of Electrical and Computer Engineering**, v. 2012, n. 436031, p. 12, 2012.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. [S.l.]: Springer, 2009. (Springer Series in Statistics).

HAYKIN, S. **Redes Neurais, Princípios e Prática**. 2ª. [S.l.]: Bookman, 1999.

HEMAMI, S. S.; REIBMAN, A. R. No-reference image and video quality estimation: applications and human-motivated design. **Signal Processing: Image Communication**, v. 25, p. 469–481, August 2010.

HERZOG, R. *et al.* NoRM: no-reference image quality metric for realistic image synthesis. **Computer Graphics Forum**, Wiley, v. 31, n. 2, p. 545–554, 2012.

HOFFMAN, J. D. **Numerical Methods for Engineers and Scientists**. 2rd. ed. [S.l.]: Taylor & Francis, 2001.

HUANG, G.; ZHU, Q.; SIEW, C. Extreme learning machine: theory and applications. **Neurocomputing**, v. 70, n. 1-3, p. 489–501, December 2006.

HUANG, G.-B. Learning capability and storage capacity of two-hidden-layer feedforward networks. **IEEE Transactions on Neural Networks**, v. 14, n. 2, p. 274–281, 2003.

HUANG, G. B. **Extreme Learning Machines**. 2013. Disponível em: <http://www3.ntu.edu.sg/home/egbhuang/>.

HUANG, G.-B.; BABRI, H. A. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. **IEEE Transactions on Neural Networks**, v. 9, n. 1, p. 224–229, 1998.

HUANG, G.-B.; CHEN, L. Enhanced random search based incremental extreme learning machine. **Neurocomputing**, v. 71, n. 16-18, p. 3460–3468, 2008.

HUANG, G.-B.; DING, X.; ZHOU, H. Optimization method based extreme learning machine for classification. **Neurocomputing**, v. 74, p. 155–163, December 2010.

HUANG, G.-B. *et al.* Incremental extreme learning machine with fully complex hidden nodes. **Neurocomputing**, v. 71, p. 576–583, January 2008.

HUANG, G.-B.; SIEW, C.-K. Extreme learning machine: RBF network case. In: **Proceedings of the Eighth International Conference on Control, Automation, Robotics and Vision (ICARCV 2004)**. Kunming, China: [s.n.], 2004. v. 2, p. 1029–1036.

HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: a new learning scheme of feedforward neural networks. In: **Proceedings 2004 International Joint Conference on Neural Networks**. [S.l.: s.n.], 2004. v. 2, p. 985–990.

ITU-R. **Recommendation BT-500: methodology for the subjective assessment of the quality for television pictures, Rev. 11**. Geneva, Switzerland, 2004.

ITU-T P.910. **Subjective video quality assessment methods for multimedia applications**. Standardization Sector of ITU, Geneva, Switzerland, 1999.

JANOWSKI, L.; ROMANIAK, P. QoE as a function of frame rate and resolution changes. In: **Proceedings of the Third International Conference on Future Multimedia Networking**. Berlin, Heidelberg: Springer-Verlag, 2010. (FMN'10), p. 34–45.

JIANG, X. *et al.* No-reference perceptual video quality measurement for high definition videos based on an artificial neural network. In: **Proceedings of the 2008 International Conference on Computer and Electrical Engineering**. Washington, USA: IEEE Computer Society, 2008. (ICCEE '08), p. 424–427.

KARAYIANNIS, N. B.; VENETSANOPOULOS, A. N. **Artificial Neural Networks: Learning Algorithms, Performance Evaluation, and Applications**. Norwell, USA: Kluwer Academic Publishers, 1992.

KARUNASEKERA, S. A.; KINGSBURY, N. G. A distortion measure for blocking artifacts in images based on human visual sensitivity. **IEEE Transactions on Image Processing**, v. 4, n. 6, p. 713–724, 1995.

KAWANO, T. *et al.* No reference video-quality-assessment model for video streaming services. In: **18th International Packet Video Workshop (PV'10)**. [S.l.: s.n.], 2010. p. 158–164.

KAWAYOKE, Y.; HORITA, Y. NR objective continuous video quality assessment model based on frame quality measure. In: **Proceedings of 15th IEEE International Conference on Image Processing (ICIP'08)**. [S.l.: s.n.], 2008. p. 385–388.

KEIMEL, C. *et al.* Visual quality of current coding technologies at high definition IPTV bitrates. In: **IEEE International Workshop on Multimedia Signal Processing (MMSP'10)**. [S.l.: s.n.], 2010. p. 390–393.

KEIMEL, C. *et al.* Design of no-reference video quality metrics with multiway partial least squares regression. In: **Proceedings of the Third International Workshop on Quality of Multimedia Experience (QoMEX)**. [S.l.: s.n.], 2011. p. 49–54.

KEIMEL, C. *et al.* No-reference video quality metric for HDTV based on H.264/AVC bitstream features. In: **Proceedings of the 18th IEEE International Conference on Image Processing (ICIP'11)**. [S.l.: s.n.], 2011. p. 3325–3328.

KEIMEL, C.; OELBAUM, T.; DIEPOLD, K. No-reference video quality evaluation for high-definition video. In: **Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'09)**. [S.l.: s.n.], 2009. p. 1145–1148.

KHAN, A. *et al.* Video quality prediction models based on video content dynamics for H.264 video over UMTS networks. **International Journal of Digital Multimedia Broadcasting**, v. 2010, n. 608138, p. 17, 2010.

KIPLI, K. *et al.* Performance of Levenberg-Marquardt backpropagation for full reference hybrid image quality metrics. **Lecture Notes in Engineering and Computer Science**, v. 2195, n. 1, p. 704–707, 2012.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2**. San Francisco, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI'95), p. 1137–1143.

LAHOUHOU, A.; VIENNET, E.; BEGHDADI, A. Selecting low-level features for image quality assessment by statistical methods. **Journal of Computing and Information Technology**, v. 18, n. 2, p. 183–189, 2010.

LEVENBERG, K. A method for the solution of certain problems in least squares. **Quarterly Applied Math**, v. 2, p. 164–168, 1944.

LI, M.-B. *et al.* Fully complex extreme learning machine. **Neurocomputing**, v. 68, p. 306–314, 2005.

LI, S.; MA, L. Full-reference video quality assessment by decoupling detail losses and additive impairments. **IEEE Transactions on Circuits and Systems for Video Technology**, n. 99, 2012.

LIANG, N.-Y. *et al.* A fast and accurate online sequential learning algorithm for feedforward networks. **IEEE Transactions on Neural Networks**, v. 17, n. 6, p. 1411–1423, 2006.

LIANG, N.-Y. *et al.* Classification of mental tasks from EEG signals using extreme learning machine. **International Journal of Neural Systems**, p. 29–38, 2006.

LIAO, N.; CHEN, Z. A packet-layer video quality assessment model with spatiotemporal complexity estimation. **EURASIP Journal on Image and Video Processing**, v. 2011, p. 5, 2011.

LIN, W.; KUO, C. C. J. Perceptual visual quality metrics: a survey. **Journal of Visual Communication and Image Representation**, v. 22, n. 4, p. 297–312, 2011.

LIU, H.; HEYNDERICKX, I. A perceptually relevant no-reference blockiness metric based on local image characteristics. **EURASIP Journal on Advances in Signal Processing**, Hindawi Publishing Corp., New York, United States, v. 2009, p. 2:1–2:14, January 2009.

LIU, H. *et al.* No-reference image quality assessment based on localized gradient statistics: application to JPEG and JPEG2000. In: ROGOWITZ, B. E.; PAPPAS, T. N. (Ed.). **Human Vision and Electronic Imaging**. [S.l.]: SPIE, 2010. (SPIE Proceedings, v. 7527), p. 75271.

LIU, H. *et al.* An efficient no-reference metric for perceived blur. In: **3rd European Workshop on Visual Information Processing (EUVIP'11)**. [S.l.: s.n.], 2011. p. 174 –179.

LU, S.; ZHANG, G.; WANG, X. A rank reduced matrix method in extreme learning machine. In: WANG, J.; YEN, G.; POLYCARPOU, M. (Ed.). **Advances in Neural Networks – ISNN 2012**. Heidelberg, Berlin: Springer, 2012. v. 7367, p. 72–79.

MAO, W. *et al.* Model selection of extreme learning machine based on multi-objective optimization. **Neural Computing & Applications**, Springer London, p. 1–9, 2012. Disponível em: <http://dx.doi.org/10.1007/s00521-011-0804-2>.

MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. **SIAM Journal on Applied Mathematics**, JSTOR, v. 11, n. 2, p. 431–441, 1963.

MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento Digital de Imagens**. Rio de Janeiro: Editora Brasport, 1999.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysic**, v. 5, p. 115–133, 1943.

MIYAHARA, M.; KOTANI, K.; ALGAZI, V. R. Objective picture quality scale (PQS) for image coding. **IEEE Transactions on Communications**, v. 46, n. 9, p. 1215–1226, 1998.

MOHAMED, S.; MEMBER, S.; RUBINO, G. A study of real-time packet video quality using random neural networks. **IEEE Transactions On Circuits and Systems for Video Technology**, v. 12, p. 1071–1083, 2002.

MORÉ, J. The Levenberg-Marquardt algorithm: implementation and theory. In: WATSON, G. A. (Ed.). **Numerical Analysis**. Berlin: Springer, 1977, (Lecture Notes in Mathematics, x). cap. 10, p. 105–116.

OELBAUM, T.; KEIMEL, C.; DIEPOLD, K. Rule-based no-reference video quality evaluation using additionally coded videos. **IEEE Journal of Selected Topics in Signal Processing**, v. 3, n. 2, p. 294–303, April 2009.

OKAMOTO, J. *et al.* Proposal for an objective video quality assessment method that takes temporal and spatial information into consideration. **Electronics and Communications in Japan (Part I: Communications)**, v. 89, n. 12, p. 97–108, 2006.

ORTEGA, J. M. **Matrix Theory**. New York and London: Plenum Press, 1987.

PANCHAL, G. *et al.* Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. **International Journal of Computer Theory and Engineering**, v. 3, n. 2, p. 332–337, 2011.

PAPPAS, T. N.; SAFRANEK, R. J. Perceptual criteria for image quality evaluation. In: **Handbook of Image and Video Processing**. [S.l.]: Academic Press, 2000. p. 669–684.

PARKER, J. R. **Algorithms for Image Processing and Computer Vision**. New York, NY, USA: John Wiley & Sons, Inc., 1997. 23-29 p.

PINSON, M.; WOLF, S. A new standardized method for objectively measuring video quality. **IEEE Transactions on Broadcasting**, IEEE, v. 50, n. 3, p. 312–322, September 2004.

PITREY, Y. *et al.* Subjective quality assessment of MPEG-4 scalable video coding in a mobile scenario. In: **Second European Workshop on Visual Information Processing (EUVIP'10)**. Paris, France: [s.n.], 2010. Disponível em: <http://hal.archives-ouvertes.fr/hal-00608333>.

PITREY, Y. *et al.* Subjective quality evaluation of H.264 high-definition video coding versus spatial up-scaling and interlacing. In: **QoE for Multimedia Content Sharing**. Tampere, Finland: [s.n.], 2010. IRCCyN contribution. Disponível em: <http://hal.archives-ouvertes.fr/hal-00608327>.

PITREY, Y. *et al.* Influence of the source content and encoding configuration on the perceived quality for scalable video coding. In: **Proceedings of the SPIE Human Vision and Electronic Imaging XVII**. San Franscisco, United States: [s.n.], 2012. v. 8291, n. 54, p. 1–6. Disponível em: <http://hal.archives-ouvertes.fr/hal-00665993>.

PITREY, Y. *et al.* Aligning subjective tests using a low cost common set. In: **QoE for Multimedia Content Sharing**. Lisbon, Portugal: [s.n.], 2011. IRCCyN contribution. Disponível em: <http://hal.archives-ouvertes.fr/hal-00608310>.

PITREY, Y. *et al.* Subjective quality of SVC-coded videos with different error-patterns concealed using spatial scalability. In: **Proceedings of EUVIP'11**. Paris, France: [s.n.], 2011. Paper number 67. Disponível em: <http://hal.archives-ouvertes.fr/hal-00608300>.

PÉCHARD, S.; PÉPION, R.; CALLET, P. L. Suitable methodology in subjective video quality assessment: a resolution dependent paradigm. In: **Proceedings of the Third International Workshop on Image Media Quality and its Applications, IMQA'08**. Kyoto, Japan: [s.n.], 2008. p. 6. Disponível em: <http://hal.archives-ouvertes.fr/hal-00300182>.

RAO, C. R.; MITRA, S. K. **Generalized Inverse of Matrices and its Applications**. New York, NY: John Wiley & Sons, 1971.

RAO, K. R.; YIP, P. **Discrete Cosine Transform: Algorithms, Advantages, Applications**. San Diego, CA, USA: Academic Press Professional, Inc., 1990.

RAO, S. **Engineering Optimization: Theory and Practice**. 4th. ed. [S.l.]: John Wiley & Sons, 2009.

RIES, M.; KUBANEK, J.; RUPP, M. Video quality estimation for mobile streaming applications with neuronal networks. In: **5th International Conference on Measurement of Audio and Video Quality in Networks (MESAQIN'06)**. Prague, Czech Republic: [s.n.], 2006.

ROSENBLATT, F. **Principles of Neurodynamics**. New York: Spartan Book, 1962.

SALOMON, D. **Data Compression: The Complete Reference**. 4th. ed. Northridge, CA: Springer, 2007. I-XXV, 1-1092 p.

SAZZAD, Z. M. P.; KAWAYOKE, Y.; HORITA, Y. No reference image quality assessment for JPEG2000 based on spatial features. **Signal Processing: Image Communication**, v. 23, n. 4, p. 257–268, 2008.

SCHALKOFF, R. J. **Digital Image Processing and Computer Vision**. [S.l.]: Wiley, 1989.

SERRE, D. **Matrices: Theory and Applications**. 1st. ed. New York: Springer, 2002.

SESHADRINATHAN, K.; BOVIK, A. Motion tuned spatio-temporal quality assessment of natural videos. **IEEE Transactions on Image Processing**, v. 19, n. 2, p. 335–350, February 2010.

SESHADRINATHAN, K. *et al.* Study of subjective and objective quality assessment of video. **IEEE Transactions on Image Processing**, v. 19, p. 1427–1441, June 2010.

SHAHID, M.; ROSSHOLM, A.; LOVSTROM, B. A reduced complexity no-reference artificial neural network based video quality predictor. In: **4th International Congress on Image and Signal Processing (CISP'11)**. [S.l.: s.n.], 2011. v. 1, p. 517–521.

SHEIKH, H. R. *et al.* **LIVE Image Quality Assessment Database**. 2003. Disponível em: <http://live.ece.utexas.edu/research/quality>.

SHI, Y. *et al.* Structure and hue similarity for color image quality assessment. In: **2009 International Conference on Electronic Computer Technology**. [S.l.: s.n.], 2009. p. 329–333.

SILVA, E. A. B. **BID – *Blurred Image Database***. 2011. Disponível em: <http://www02.lps.ufrj.br/∼eduardo/eduardo_oficial/ImageDatabase.htm>.

SILVA, W. B.; POHL, A. A. P. No-reference video quality assessment method based on the Levenberg-Marquardt minimization. In: **XXX Brazilian Symposium on Telecommunications (SBrT'12)**. Brasília, Brazil: [s.n.], 2012.

SILVA, W. B.; POHL, A. A. P.; FONSECA, K. V. O. A reduced-reference video quality assessment method based on the activity-difference of DCT coefficients. **IEICE Transactions on Information and Systems**, E96-D, n. 3, p. 708–718, March 2013.

SIMONE, F. *et al.* Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel. In: **First International Workshop on Quality of Multimedia Experience (QoMEX'09)**. San Diego, USA: [s.n.], 2009.

SIMONE, F. *et al.* H.264/AVC video database for the evaluation of quality metrics. In: **35th International Conference on Acoustics, Speech, and Signal Processing (ICASSP'10)**. Dallas, USA: [s.n.], 2010. p. 2430–2433.

SLANINA, M.; RICNY, V.; FORCHHEIMER, R. A novel metric for H.264/AVC no-reference quality assessment. In: **14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services**. [S.l.: s.n.], 2007. p. 114–117.

SONG, X.; YANG, Y. A new no-reference assessment metric of blocking artifacts based on HVS masking effect. In: **Proceedings of the 2nd International Congress on Image and Signal Processing (CISP'09)**. [S.l.: s.n.], 2009. p. 1–6.

SPIEGEL, M. R.; STEPHENS, L. J. **Theory and Problems of Statistics**. 3rd. ed. New York: McGraw-Hill, 1998. (Schaum's Outline Series).

STAELENS, N. *et al.* VIQID: a no-reference bit stream-based visual quality impairment detector. In: **Proceedings of the 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX 2010)**. Piscataway, USA: IEEE, 2010. p. 206–211.

SUGIMOTO, O. *et al.* Objective perceptual video quality measurement method based on hybrid no reference framework. In: **16th IEEE International Conference on Image Processing (ICIP'09)**. [S.l.: s.n.], 2009. p. 2237–2240.

SURESH, S.; BABU, R. V.; KIM, H. J. No-reference image quality assessment using modified extreme learning machine classifier. **Applied Soft Computing**, v. 9, n. 2, p. 541–552, 2009.

TAMURA, S.; TATEISHI, M. Capabilities of a four-layered feedforward neural network: four layers versus three. **IEEE Transactions on Neural Networks**, v. 8, n. 2, p. 251–255, 1997.

TONG, H. *et al.* Learning no-reference quality metric by examples. In: **Proceedings of International Multi-Media Modelling Conference**. [S.l.: s.n.], 2005. p. 247–254.

TUKEY, J. W. **Exploratory Data Analysis**. New York: Addison-Wesley Publishing Company, 1977.

van Marais, I. Z.; STEYN, W.; du Preez, J. Construction of an image quality assessment model for use on board an Leo satellite. In: **IEEE International Geoscience and Remote Sensing Symposium (IGARSS'08)**. [S.l.: s.n.], 2008. v. 2, p. II–1068 –II–1071.

VEERARAJAN, T. **Probability, Statistics and Random Processes**. 3rd. ed. New Delhi, India: Tata McGraw-Hill Education, 2008.

VENKATARAMAN, M. *et al.* Designing a collector overlay architecture for fault diagnosis in video networks. **Computer Communications**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, v. 35, n. 4, p. 418–430, February 2012.

VQEG. **Final report from the video quality experts group on the validation of objective models of video quality assessment**. Video Quality Experts Group (VQEG): [s.n.], 2000. Tech. Rep. Disponível em: <http://www.its.bldrdoc.gov/vqeg/>.

VQEG. **Final report from the video quality experts group on the validation of objective models video quality assessment, Phase II**. Video Quality Experts Group (VQEG): [s.n.], 2003. Tech. Rep. Disponível em: <http://www.its.bldrdoc.gov/vqeg/>.

VQEG. **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): [s.n.], 2008. Tech. Rep. Disponível em: <http://www.its.bldrdoc.gov/vqeg/>.

VQEG. **Final report from the video quality experts group on the validation of reduced-reference and no-reference objective models for standard definition television, Phase I**. Video Quality Experts Group (VQEG): [s.n.], 2009. Tech. Rep. Disponível em: <http://www.its.bldrdoc.gov/vqeg/>.

VQEG. **Report on the validation of video quality models for high definition video content**. Video Quality Experts Group (VQEG): [s.n.], June 2010. Tech. Rep., Version 2.0. Disponível em: <http://www.its.bldrdoc.gov/vqeg/projects/hdtv/>.

WANG, A. *et al.* New no-reference blocking artifacts metric based on human visual system. In: **Proceedings of the International Conference on Wireless Communications Signal Processing (WCSP'09)**. [S.l.: s.n.], 2009. p. 1–5.

WANG, Z.; BOVIK, A. C. A universal image quality index. **IEEE Signal Processing Letters**, v. 9, n. 3, p. 81–84, 2002.

WANG, Z.; BOVIK, A. C. **Modern Image Quality Assessment**. [S.l.]: Morgan & Claypool Publishers, 2006. (Synthesis Lectures on Image, Video, and Multimedia Processing).

WANG, Z.; BOVIK, A. C.; LU, L. Why is image quality assessment so difficult? In: **IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'02)**. [S.l.: s.n.], 2002. v. 4, p. IV–3313–IV–3316.

WANG, Z. *et al.* Image quality assessment: from error visibility to structural similarity. **IEEE Signal Processing Letters**, v. 13, n. 4, p. 600–612, 2004.

WANG, Z.; LU, L.; BOVIK, A. C. Video quality assessment based on structural distortion measurement. **Signal Processing: Image Communication**, v. 19, n. 2, p. 121–132, 2004.

WANG, Z.; SHEIKH, H. R.; BOVIK, A. C. No-reference perceptual quality assessment of JPEG compressed images. In: **Proceedings of the IEEE International Conference on Image Processing (ICIP'02)**. [S.l.: s.n.], 2002. v. 1, p. I–477–I–480.

WANG, Z.; SHEIKH, H. R.; BOVIK, A. C. Objective video quality assessment. In: FURHT, B.; MARQUES, O. (Ed.). **The Handbook of Video Databases: Design and Applications**. Boca Raton, USA: CRC Press, 2003. cap. 41, p. 1041–1078.

WANG, Z.; SIMONCELLI, E. P.; BOVIK, A. C. Multiscale structural similarity for image quality assessment. In: **Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems, and Computers**. Pacific Grove, CA: IEEE Computer Society, 2003. v. 2, p. 1398–1402.

WIDROW, B. The original adaptive neural net broom-balancer. In: **Proceedings of the IEEE International Symposium on Circuits and Systems**. [S.l.: s.n.], 1987. p. 351–357.

WIDROW, B.; HOFF, J. M. Adaptive switching circuits. **IRE WESCON Convention Record**, v. 4, p. 96–104, 1960.

WU, H. R.; RAO, K. R.; KASSIM, A. A. Digital video image quality and perceptual coding. **Journal of Electronic Imaging**, v. 16, n. 3, p. 039901, 2007.

WU, H. R.; YUEN, M. A generalized block-edge impairment metric for video coding. **IEEE Signal Processing Letters**, v. 4, n. 11, p. 317–320, 1997.

YAMAGISHI, K. *et al.* No reference video-quality-assessment model for monitoring video quality of IPTV services. **IEICE Transactions**, v. 95-B, n. 2, p. 435–448, 2012.

YANG, F. *et al.* A novel objective no-reference metric for digital video quality assessment. **IEEE Signal Processing Letters**, v. 12, n. 10, p. 685–688, October 2005.

YANG, F. *et al.* No-reference quality assessment for networked video via primary analysis of bit stream. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 20, n. 11, p. 1544–1554, 2010.

YAO, J. *et al.* No-reference video quality assessment using statistical features along temporal trajectory. **Procedia Engineering**, v. 29, n. 0, p. 947–951, 2012.

YAO, J. *et al.* No-reference objective quality assessment for video communication services based on feature extraction. In: **Proceedings of the 2nd International Congress on Image and Signal Processing (CISP'09)**. [S.l.: s.n.], 2009. p. 1–6.

ZHANG, L. *et al.* FSIM: A feature similarity index for image quality assessment. **IEEE Transactions on Image Processing**, v. 20, n. 8, p. 2378–2386, 2011.

ZHU, Q.-Y. *et al.* Evolutionary extreme learning machine. **Pattern Recognition**, v. 38, n. 10, p. 1759–1763, 2005.

APÊNDICE A – DIAGRAMAS TI *vs.* SI

Além das Figuras 5 e 4 presentes na Seção 3.1, as Figuras 46 a 66 apresentam diagramas com as informações temporais (TI) e espaciais (SI) das bases de dados de vídeos utilizadas na validação dos métodos propostos.

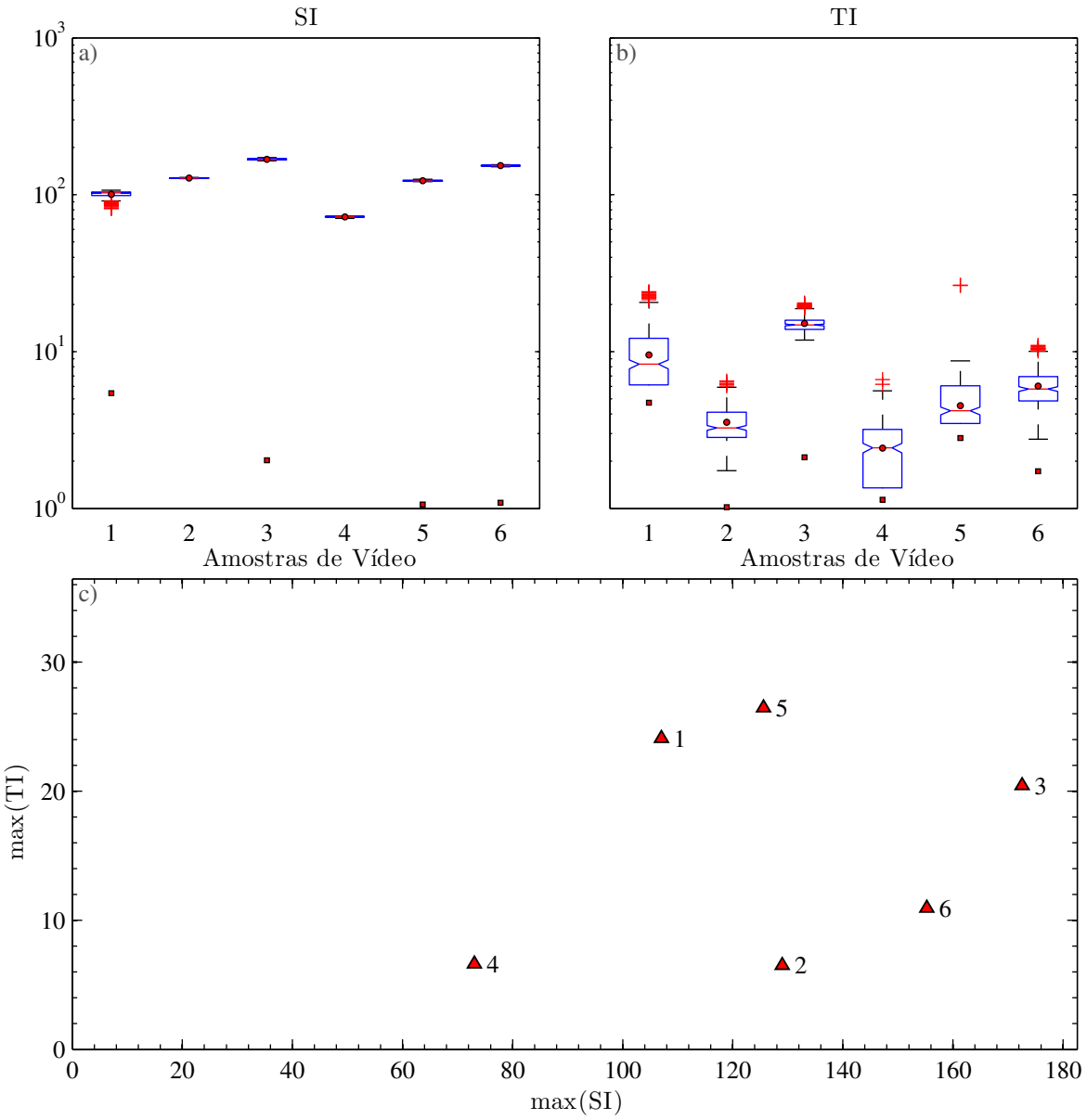

Figura 46: Diagrama TI *vs.* SI para a base de dados EPFL/PoliMi com resolução CIF e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

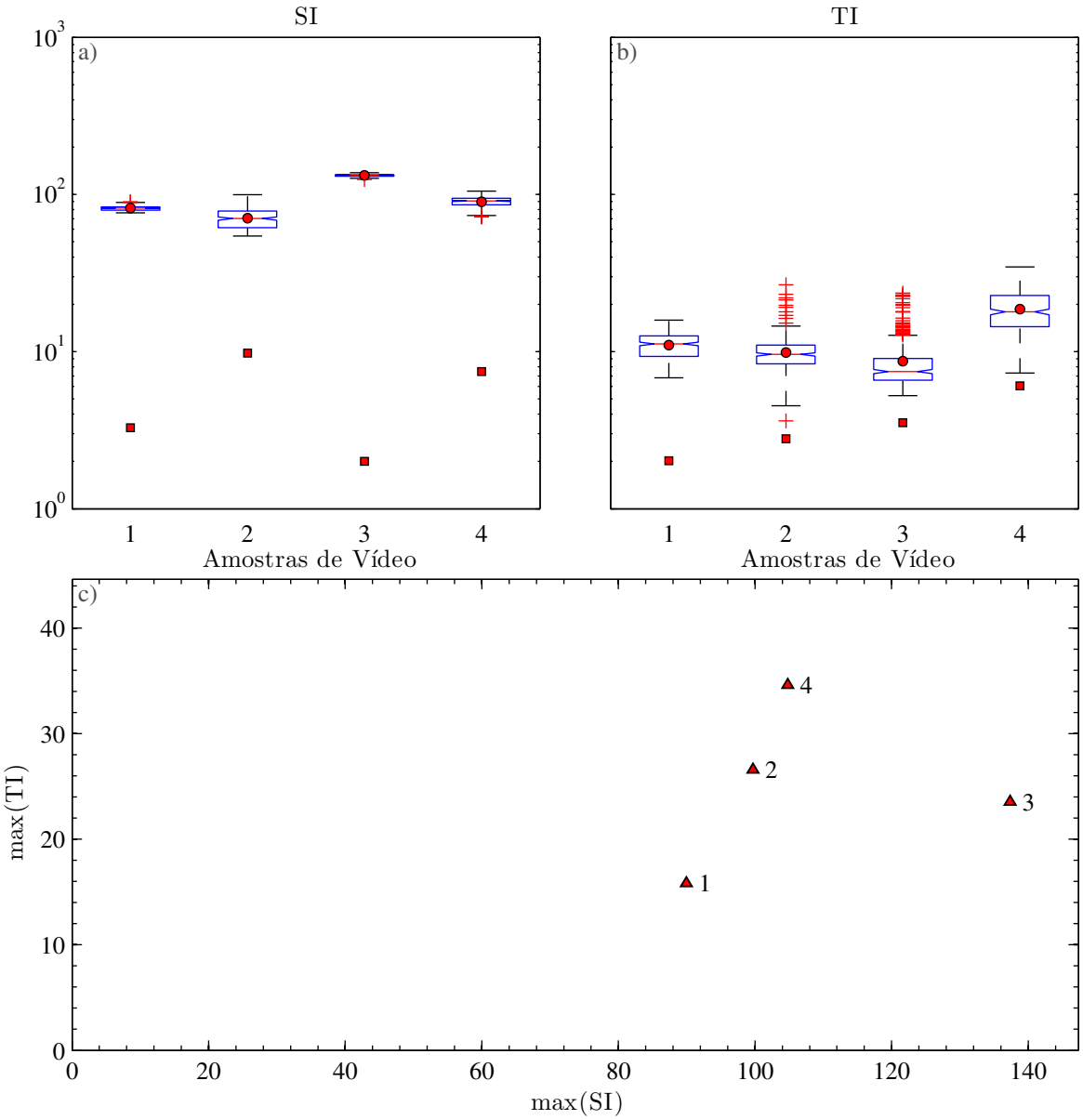

Figura 47: Diagrama TI *vs.* SI para a base de dados IRCCyN/IVC H.264/AVC *vs.* SVC VGA (resolução QVGA) *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

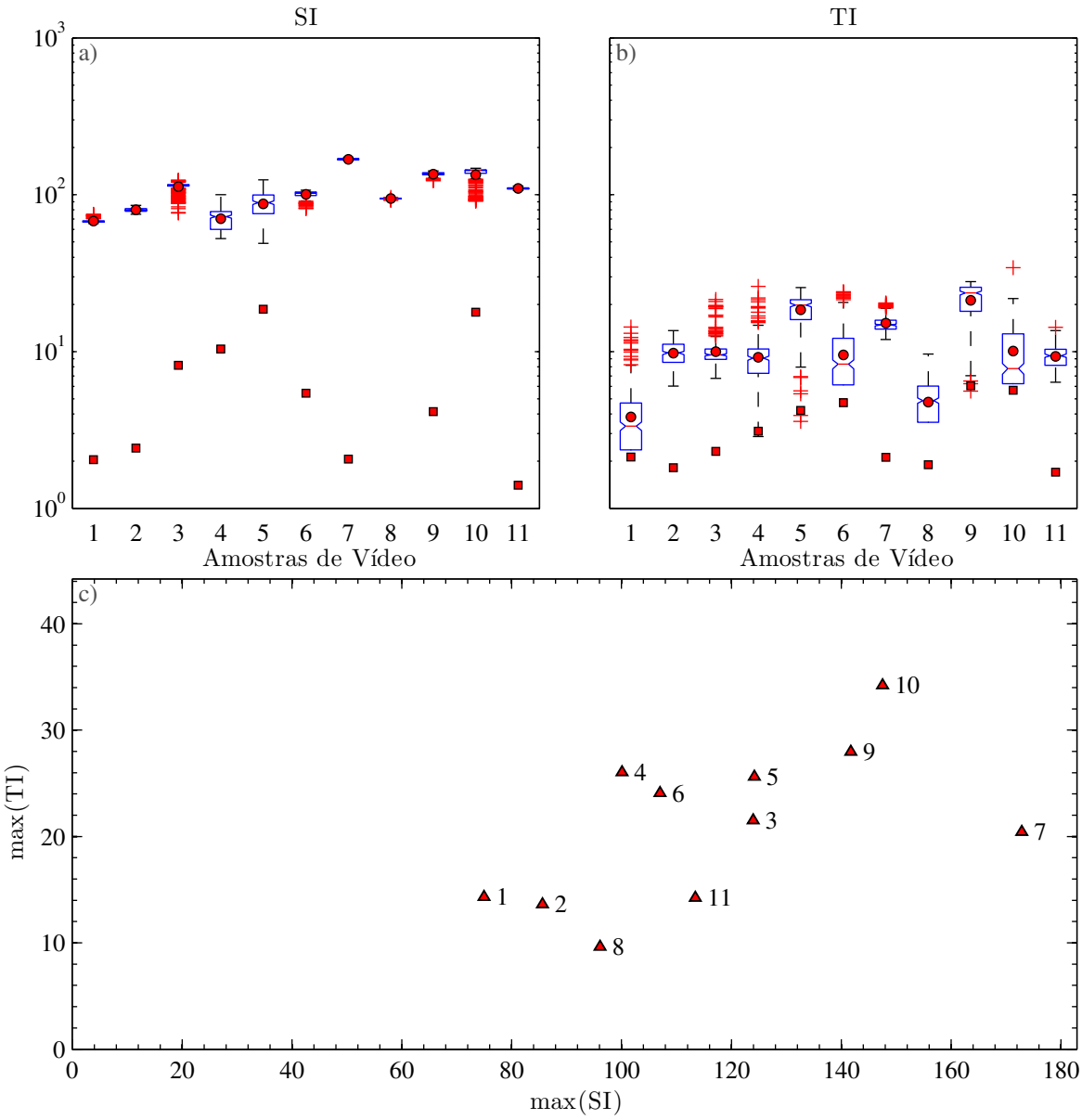

Figura 48: Diagrama TI *vs.* SI para a base de dados IST com resolução CIF e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

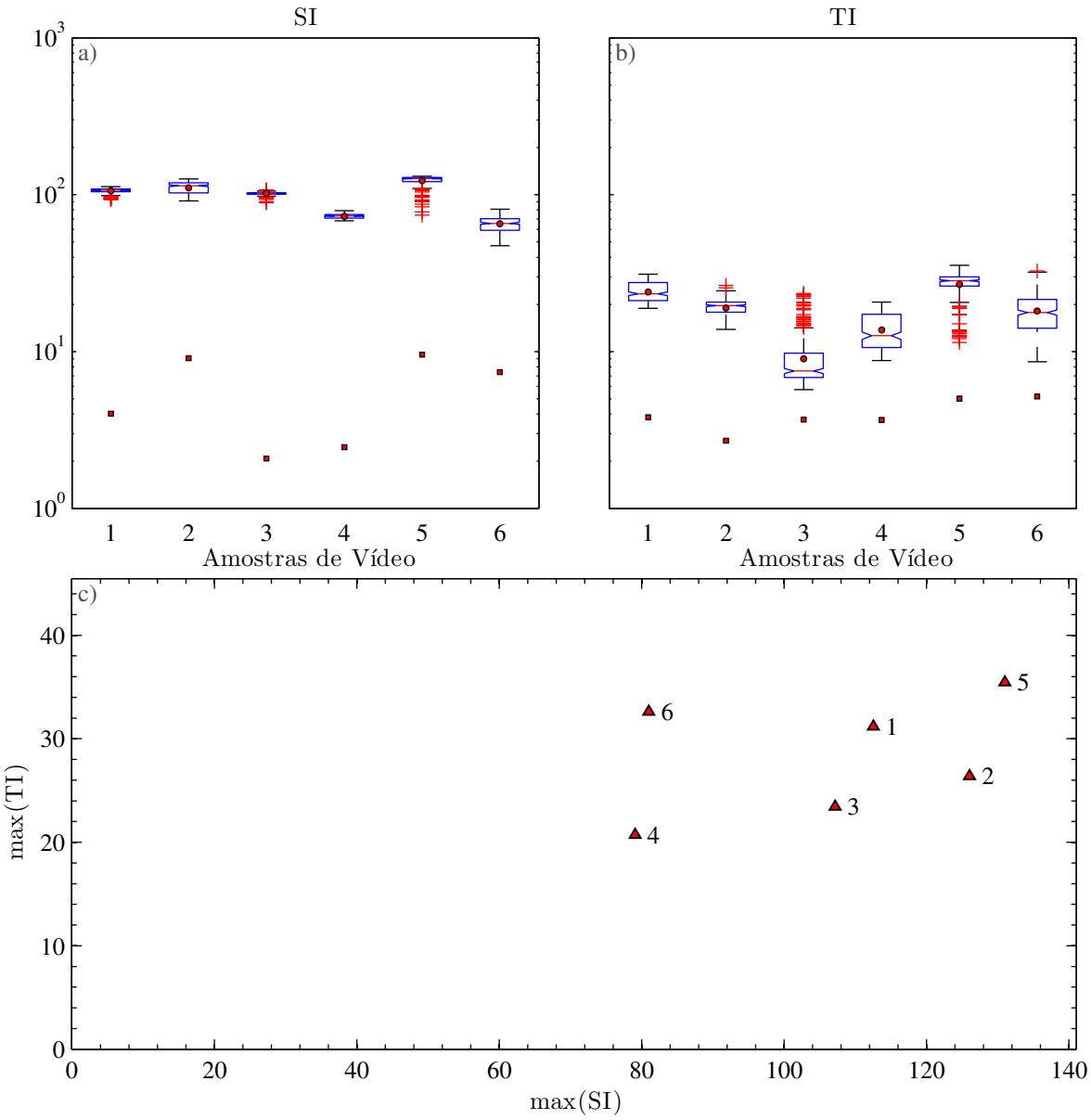

Figura 49: Diagrama TI *vs.* SI para a base de dados EPFL/PoliMi com resolução 4CIF e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

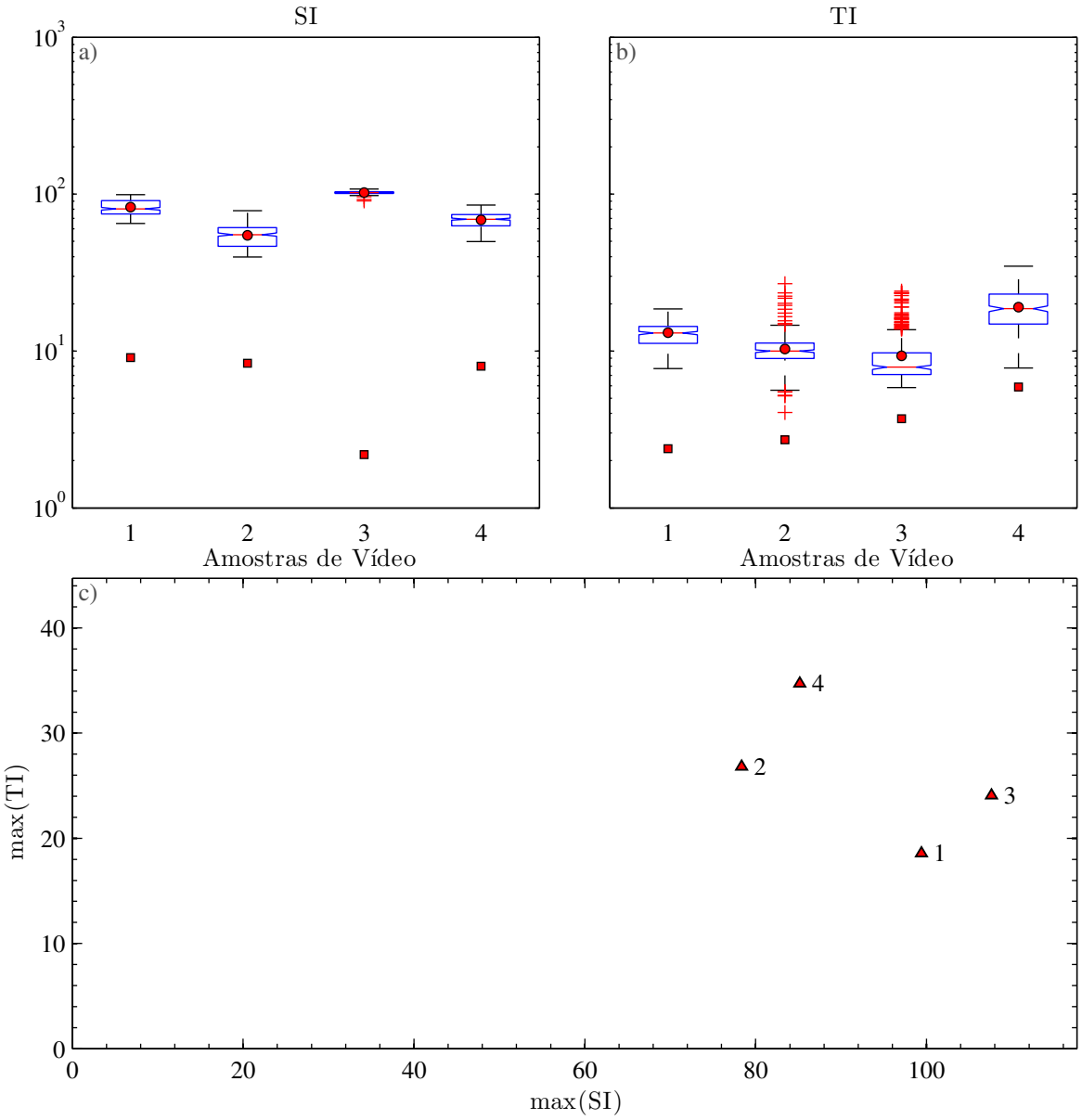

Figura 50: Diagrama TI *vs.* SI para a base de dados IRCCyN/IVC H.264/AVC *vs.* SVC VGA Video Database (resolução VGA) e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

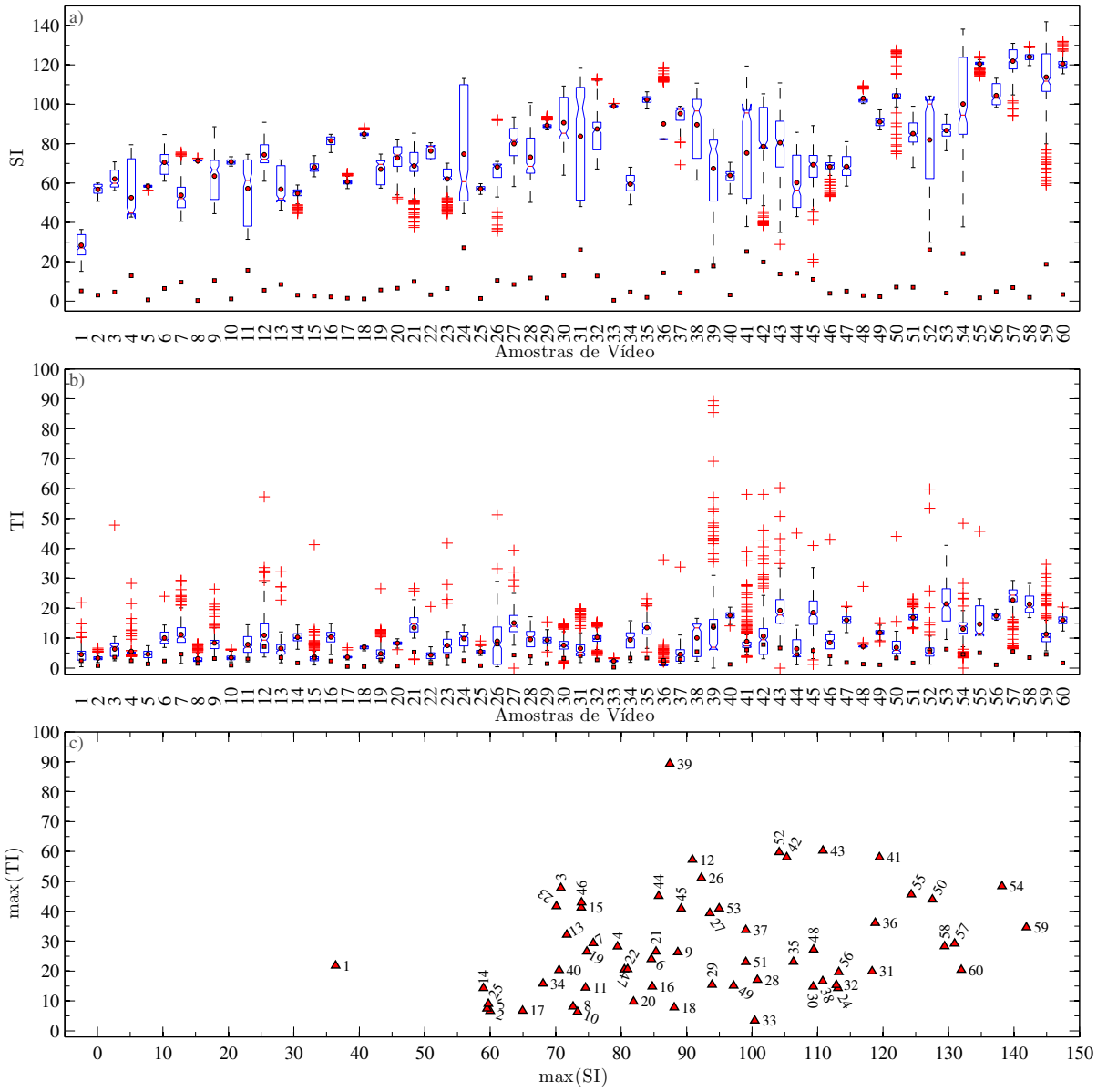

Figura 51: Diagrama TI *vs.* SI para a base de dados IRCCyN/IVC *Influence Content* VGA *Database* e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

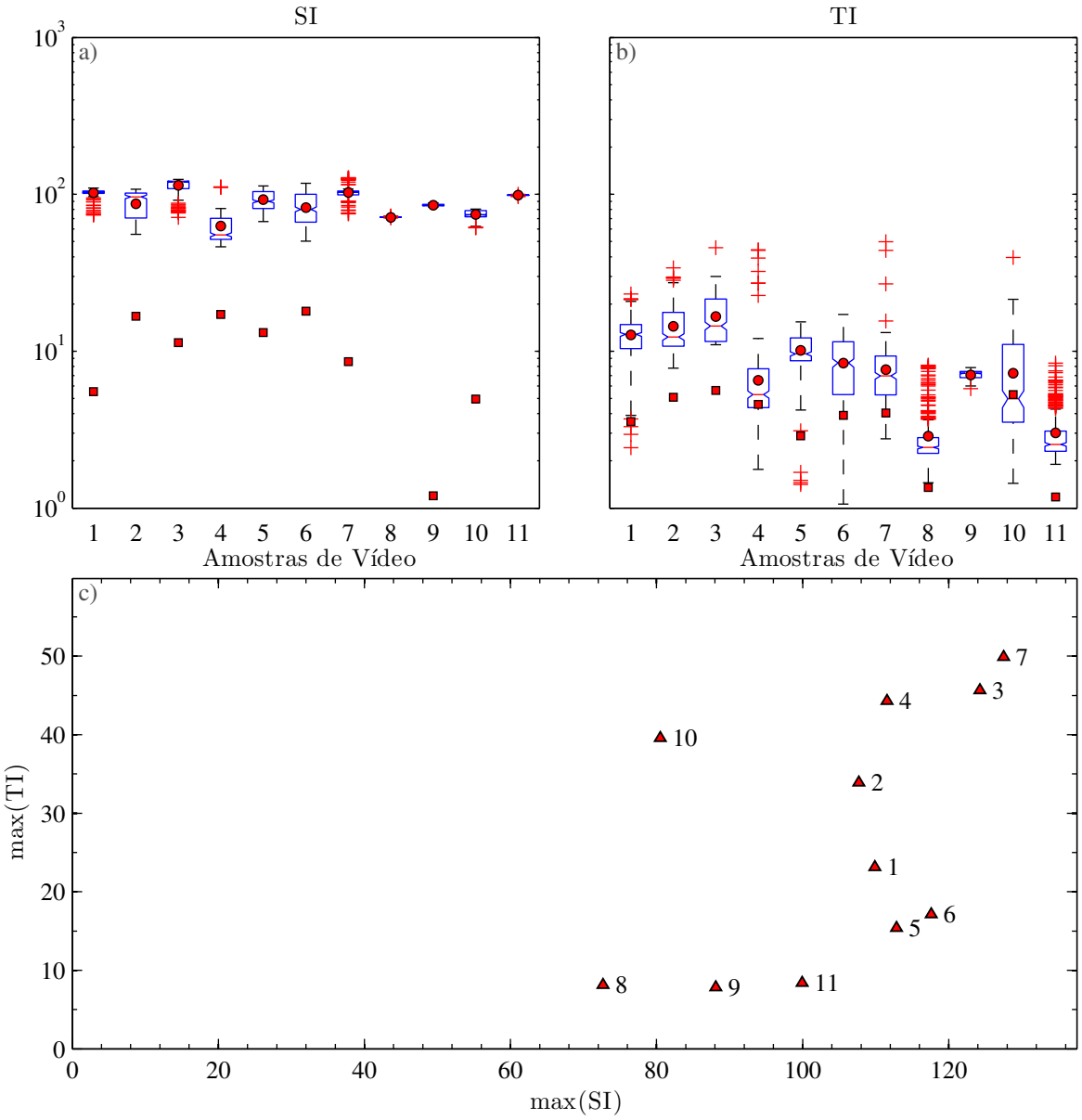

Figura 52: Diagrama TI *vs.* SI para a base de dados IRCCyN/IVC SVC4QoE QP0 QP1 *Video* VGA *Database* e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

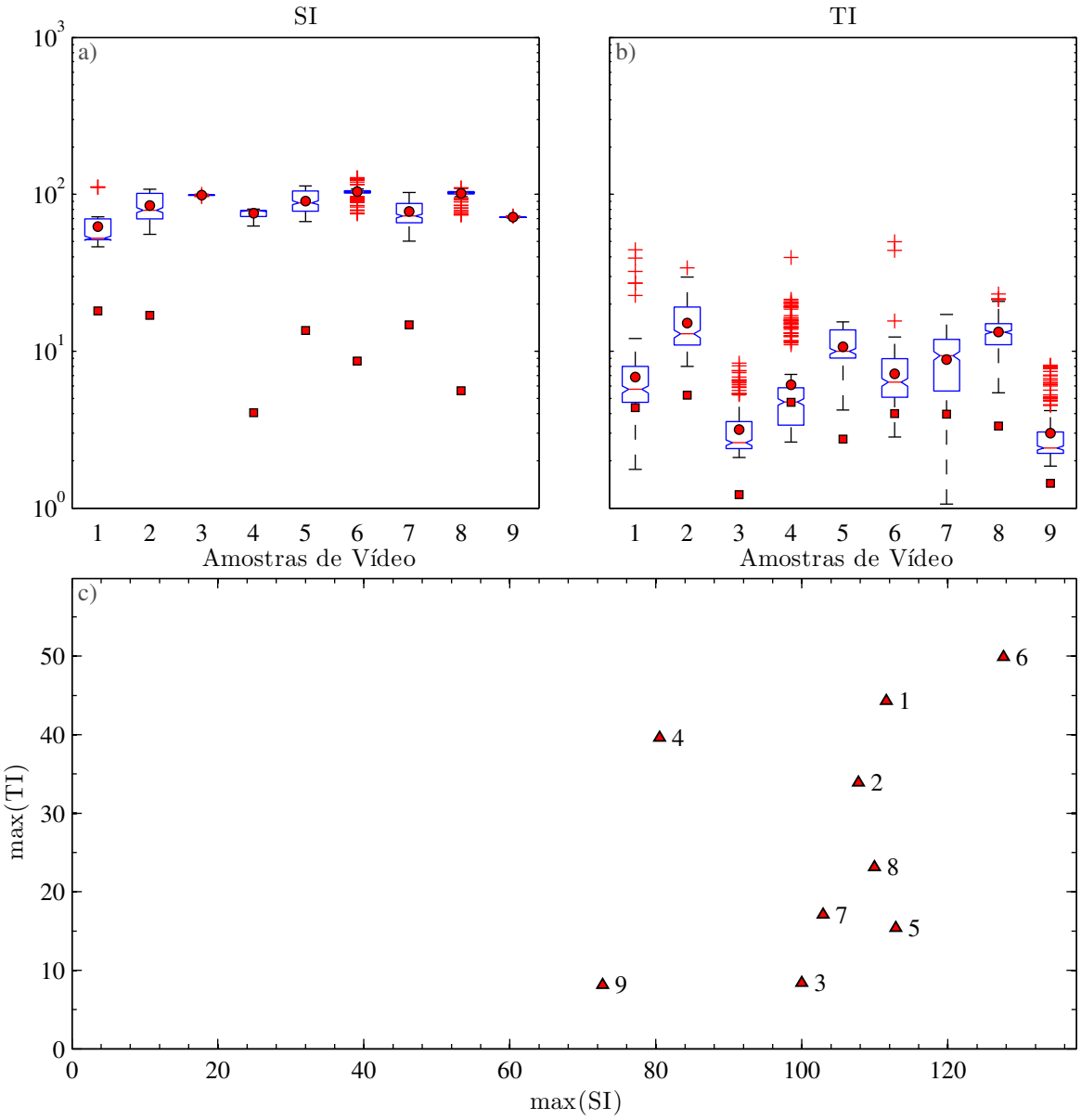

Figura 53: Diagrama TI *vs.* SI para a base de dados IRCCyN/IVC SVC4QoE *Replace Slice* VGA *Database* e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

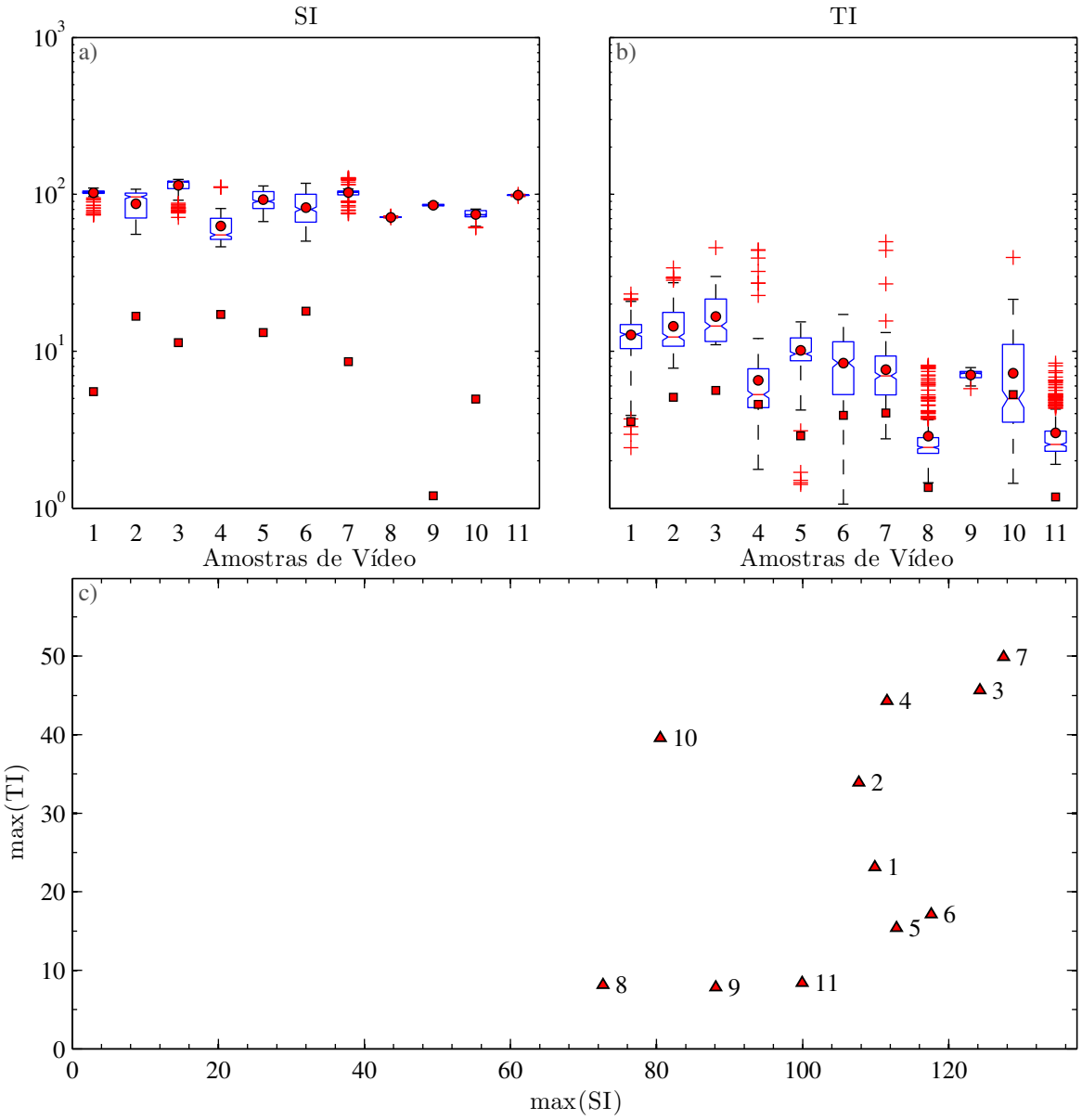

Figura 54: Diagrama TI *vs.* SI para a base de dados IRCCyN/IVC SVC4QoE *Temporal Switch Video* VGA *Database* e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

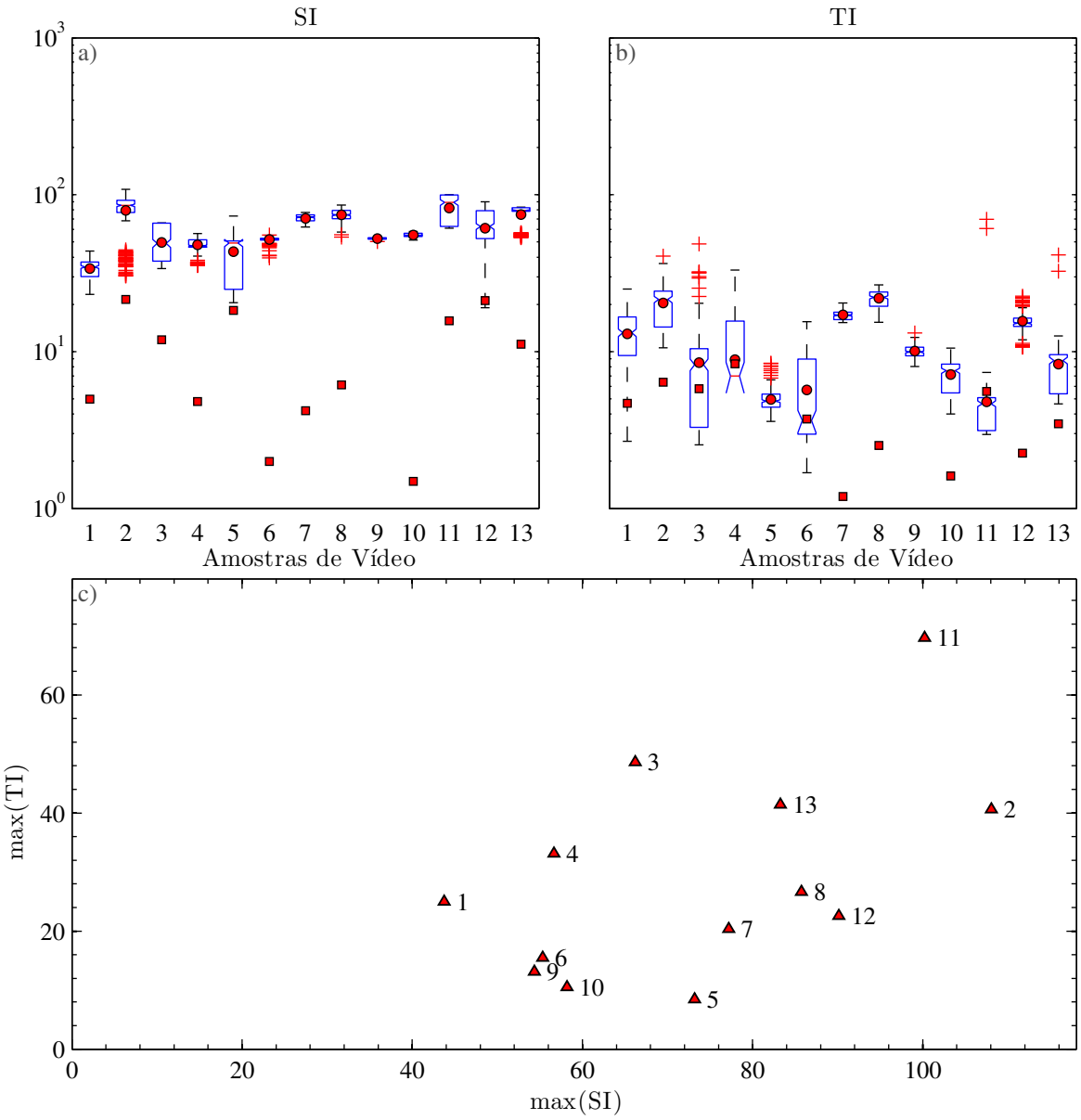

Figura 55: Diagrama TI *vs.* SI para a base de dados HDTV Phase I VQEGHD-4 e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

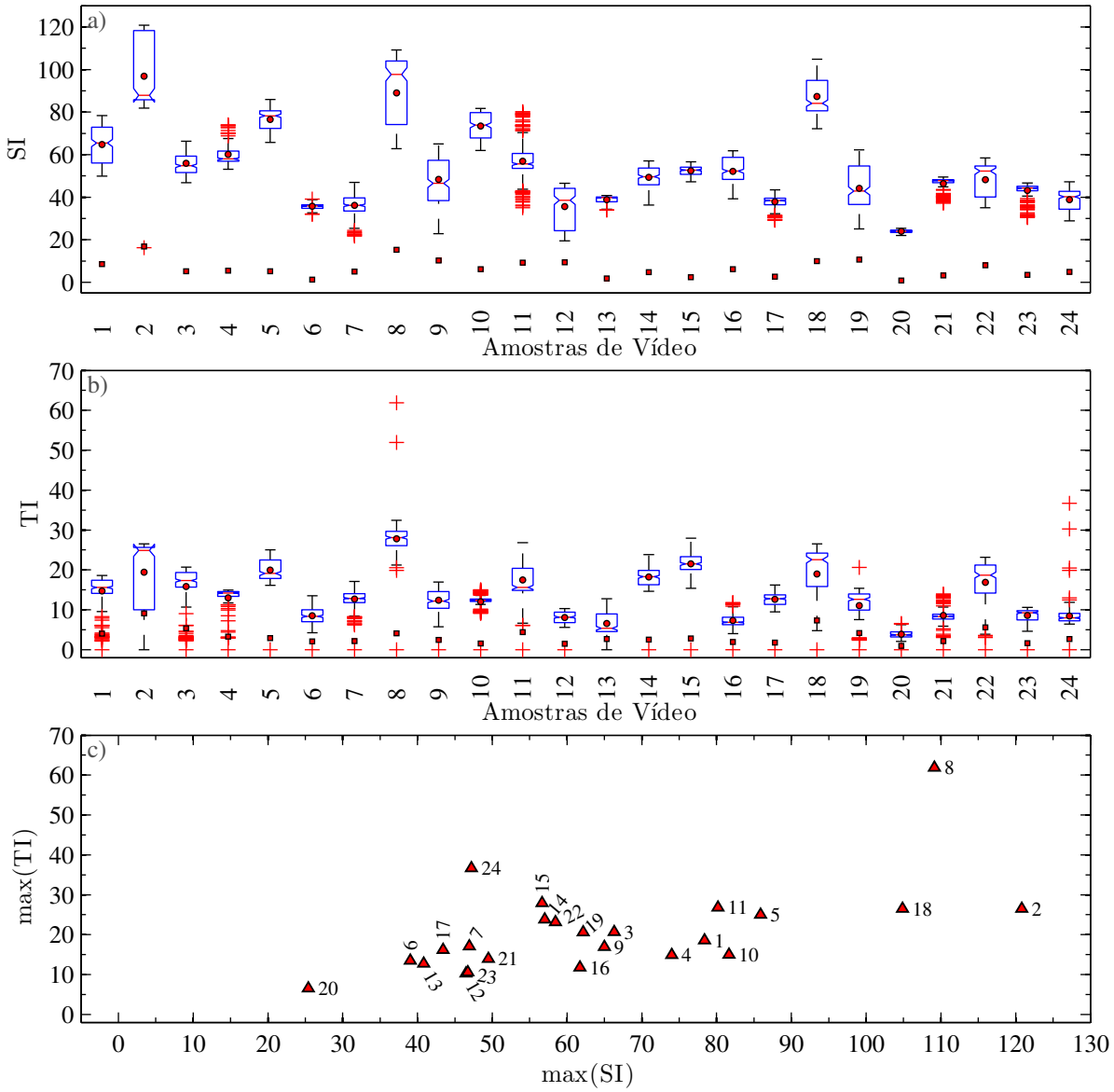

Figura 56: Diagrama TI *vs.* SI para a base de dados IRCCyN/IVC 1080i *Database* e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

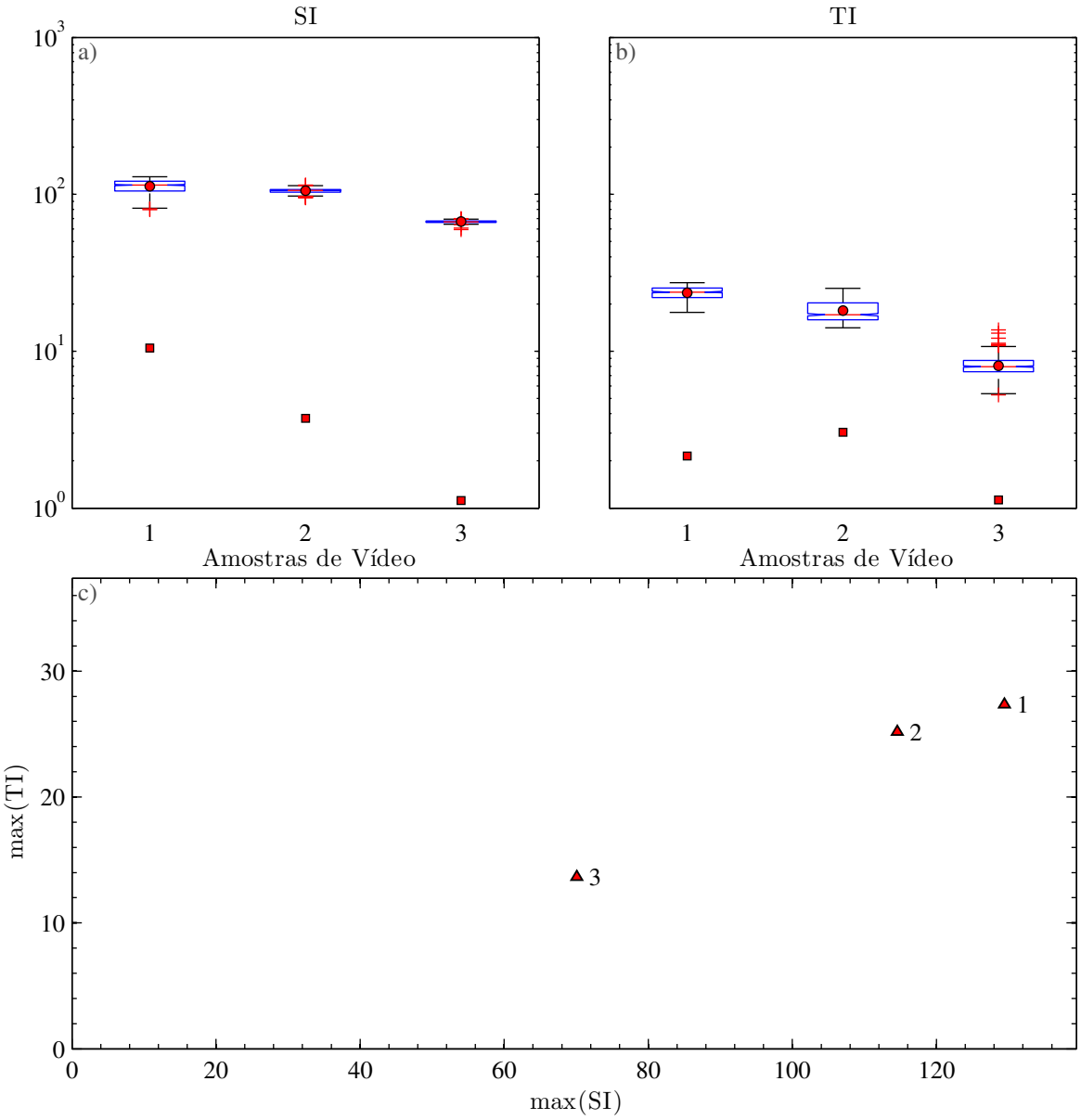

Figura 57: Diagrama TI *vs.* SI para a base de dados IRCCyN/IVC H.264 HD *vs. Upscaling and Interlacing Video Database* e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

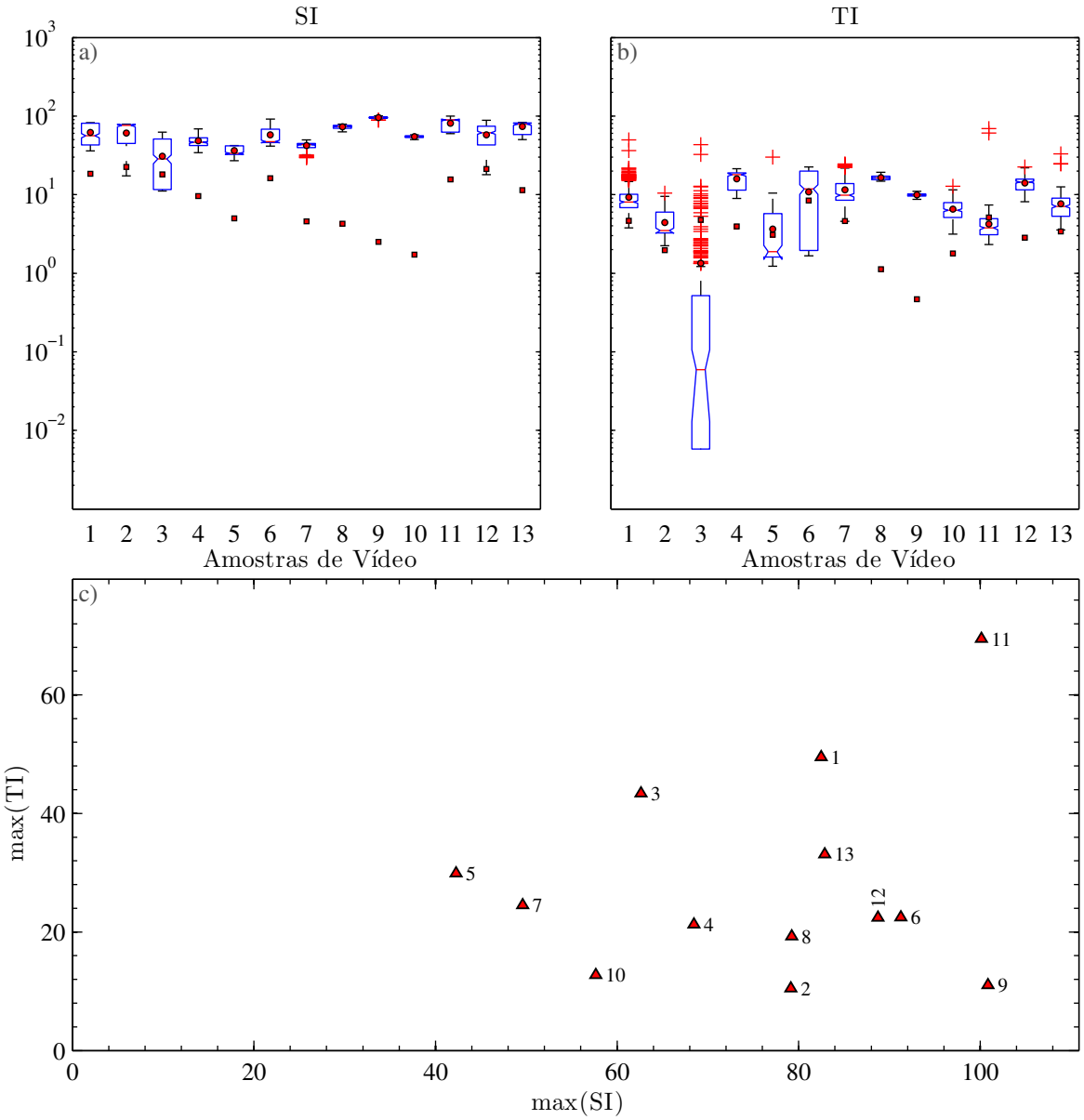

Figura 58: Diagrama TI *vs.* SI para a base de dados VQEG *Pool2* 1080i *Video Database* (HDTV Phase I VQEGHD-2) e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

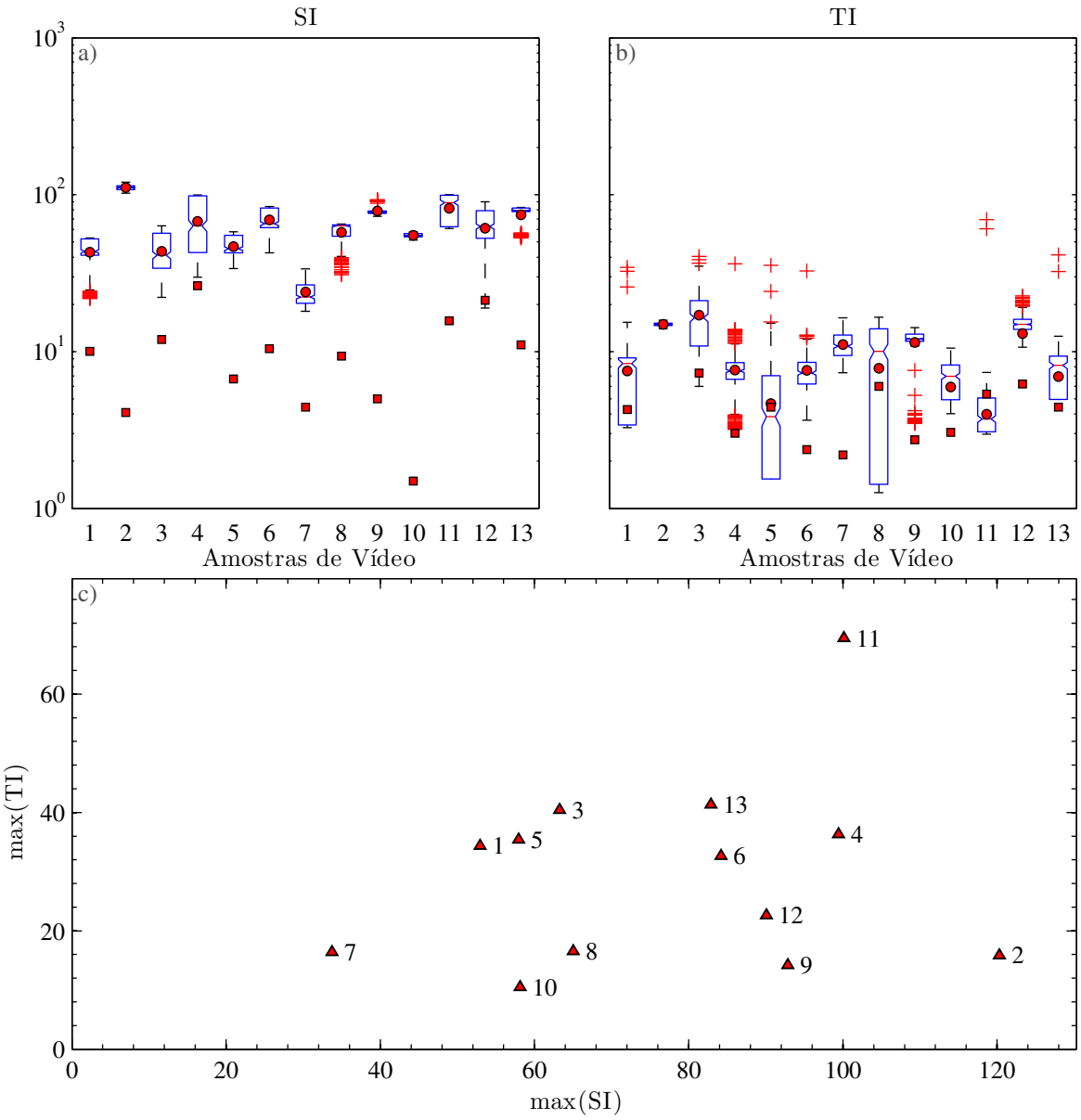

Figura 59: Diagrama TI *vs.* SI para a base de dados HDTV Phase I VQEGHD-2 e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

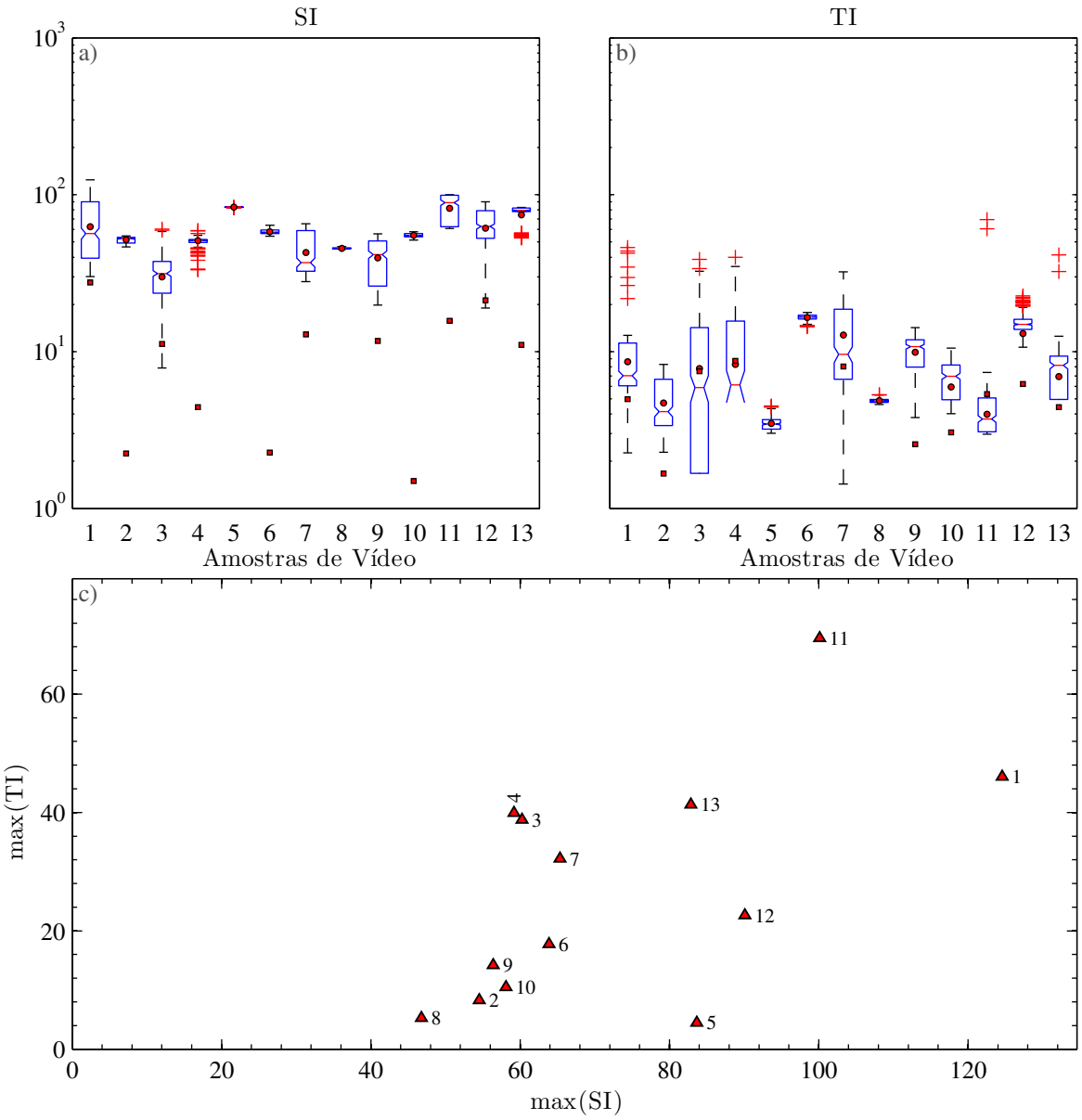

Figura 60: Diagrama TI *vs.* SI para a base de dados HDTV Phase I VQEGHD-3 e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

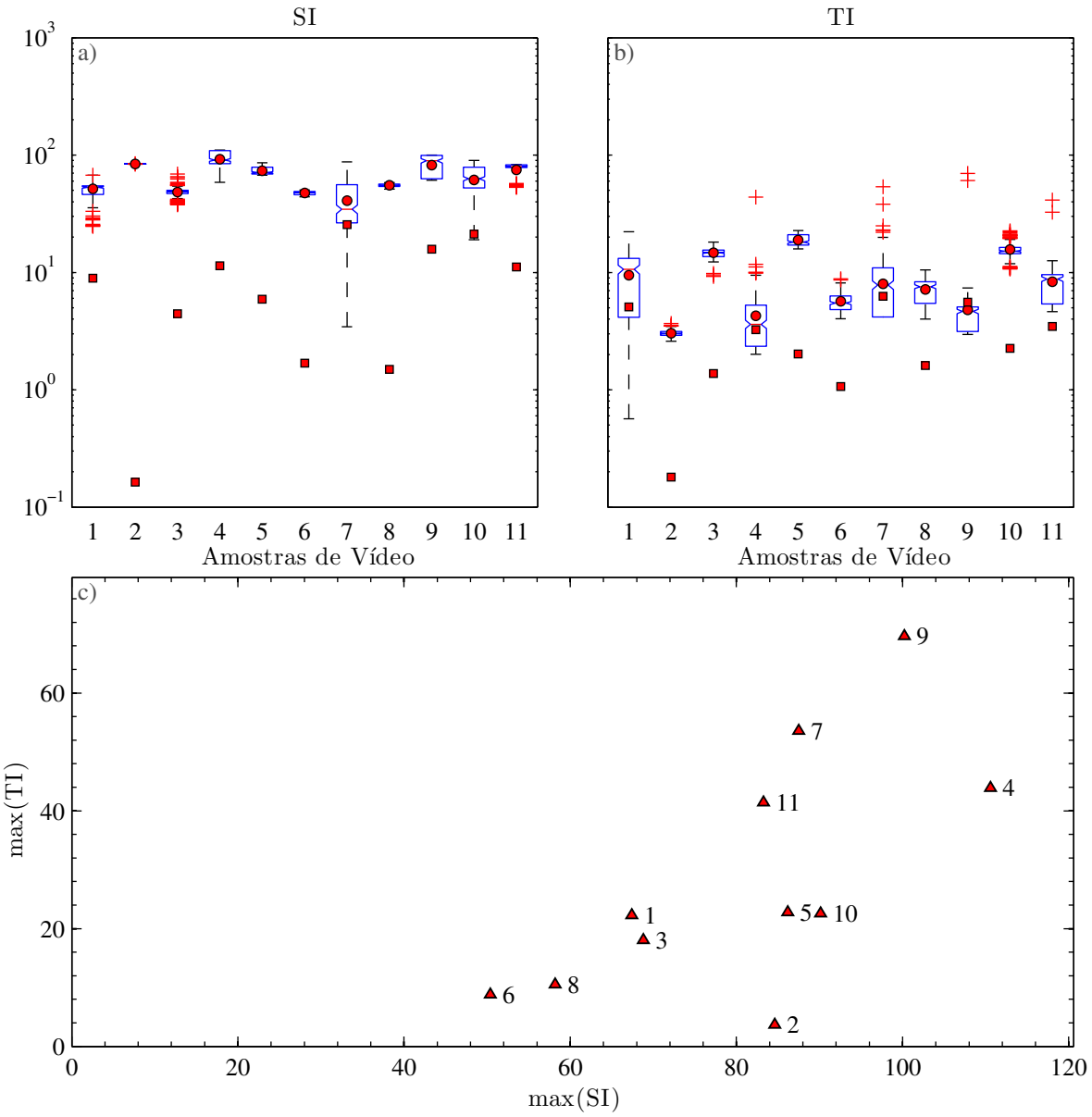

Figura 61: Diagrama TI *vs.* SI para a base de dados HDTV Phase I VQEGHD-5 e *box-plot* em (a) SI, (b) TI e (c) max(TI) *vs.* max(SI).

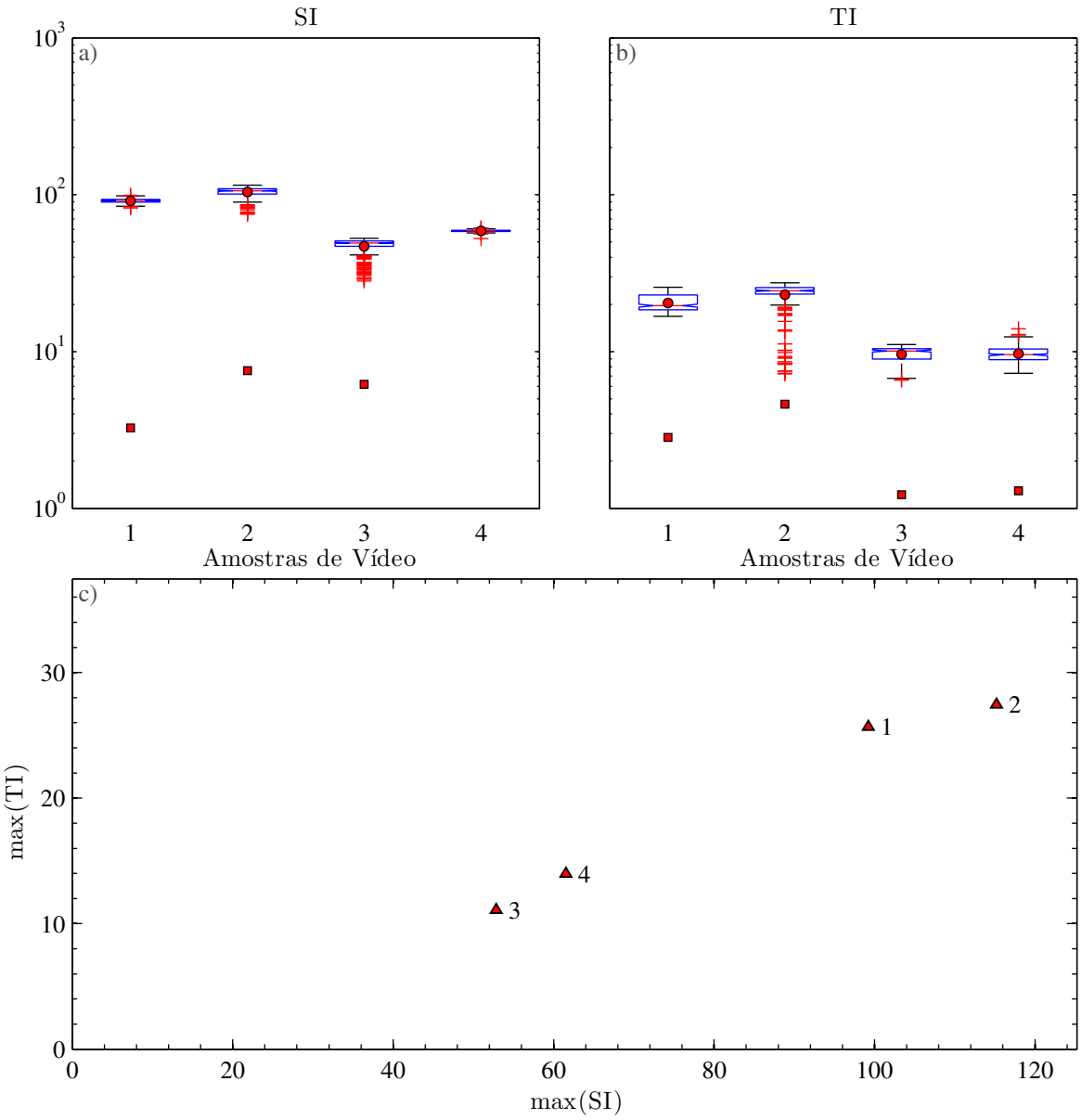

Figura 62: Diagrama TI *vs.* SI para a base de dados TUM 1080p25 *Data Set* e *box-plot* em (a) SI, (b) TI (c) e max(TI) *vs.* max(SI).

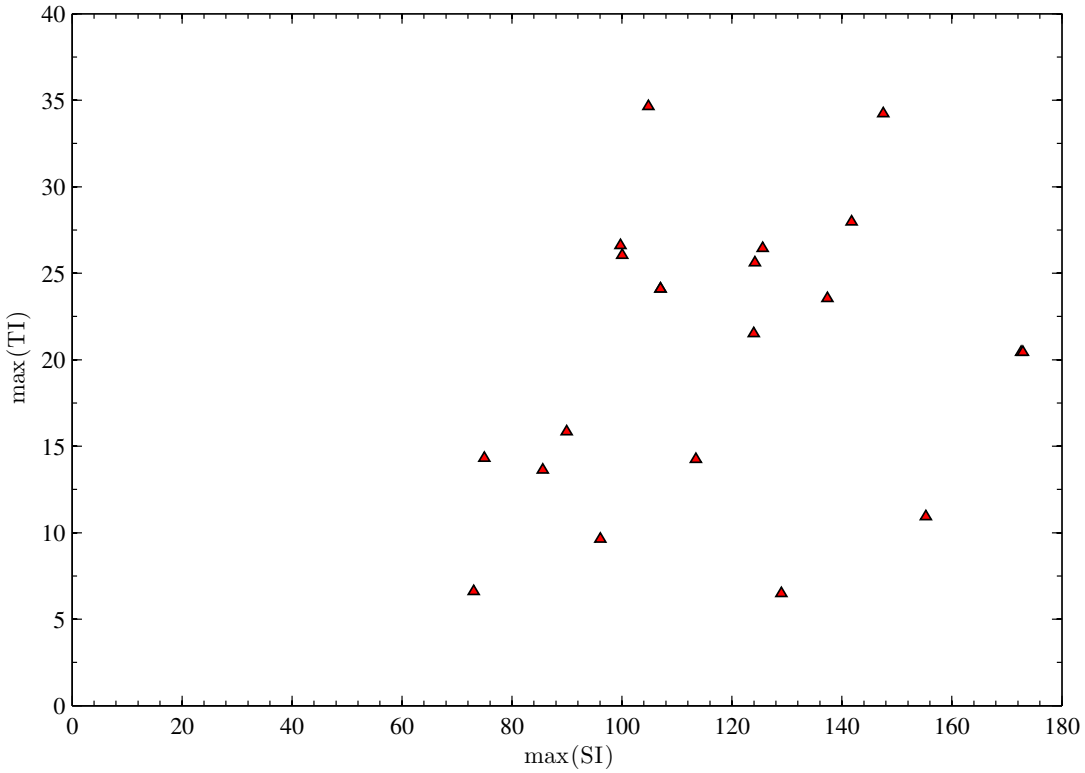

Figura 63: Diagrama max(TI) *vs.* max(SI) para base de dados do superconjunto *S* em LD com 26 vídeos de referência.

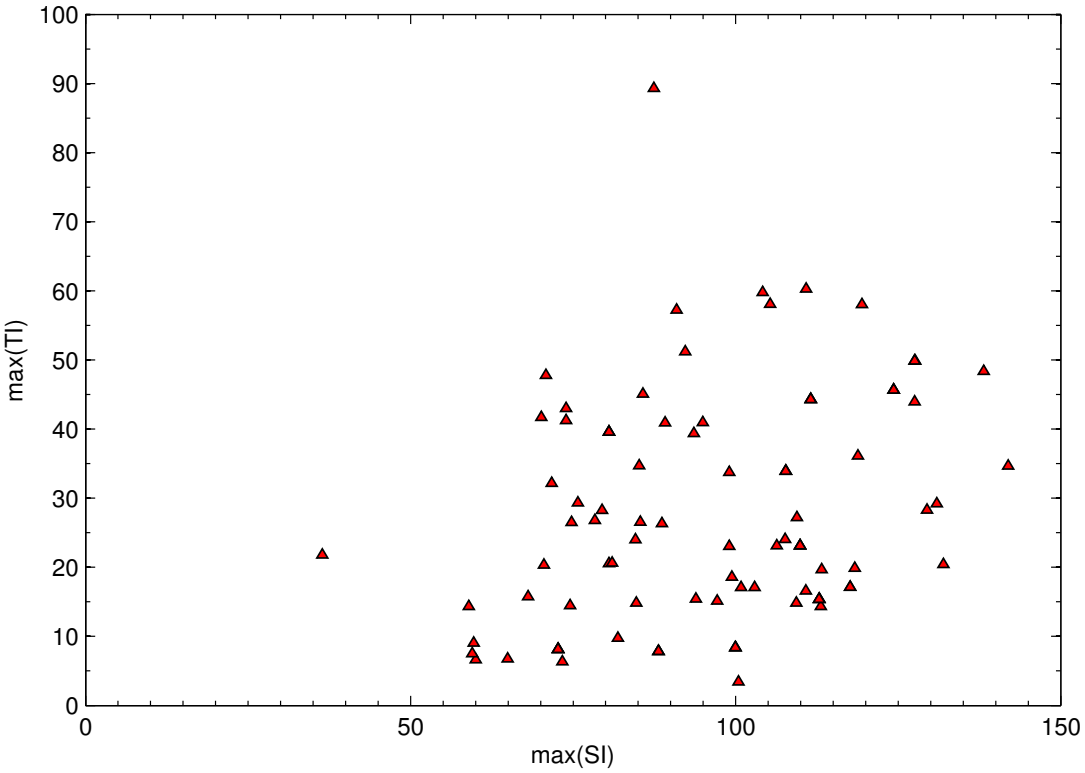

Figura 64: Diagrama max(TI) *vs.* max(SI) para base de dados do superconjunto *S* em SD com 95 vídeos de referência.

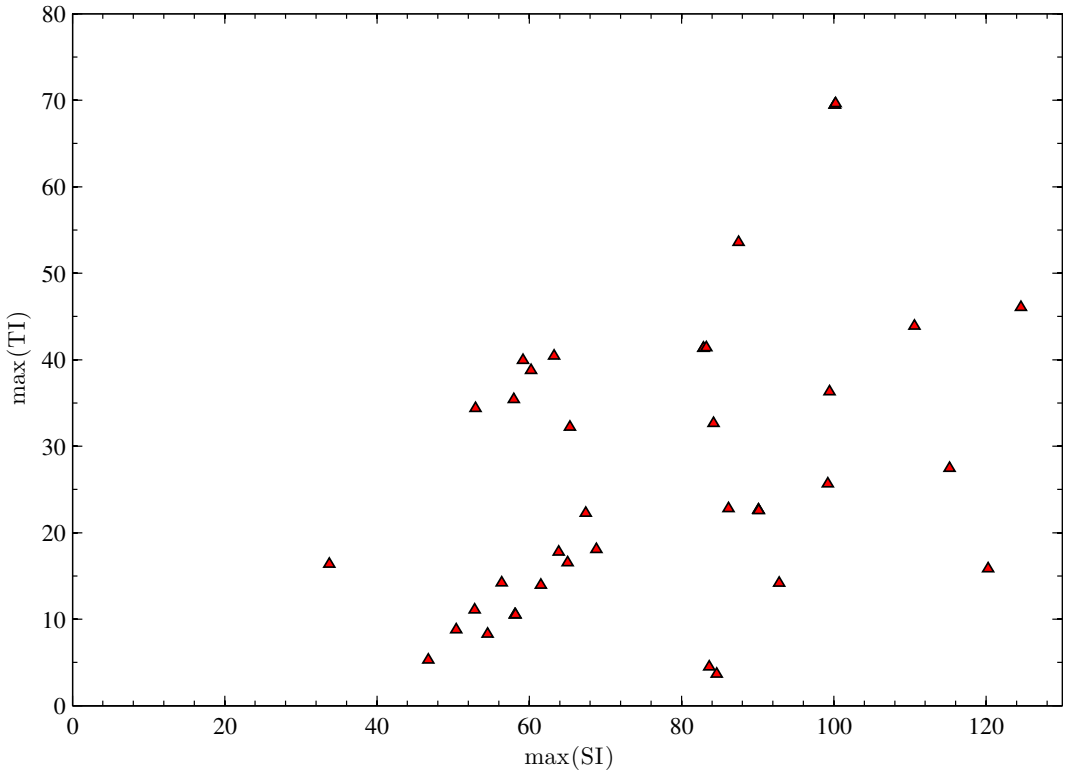

Figura 65: Diagrama max(TI) *vs.* max(SI) para base de dados do superconjunto *S* em HDp com 41 vídeos de referência.

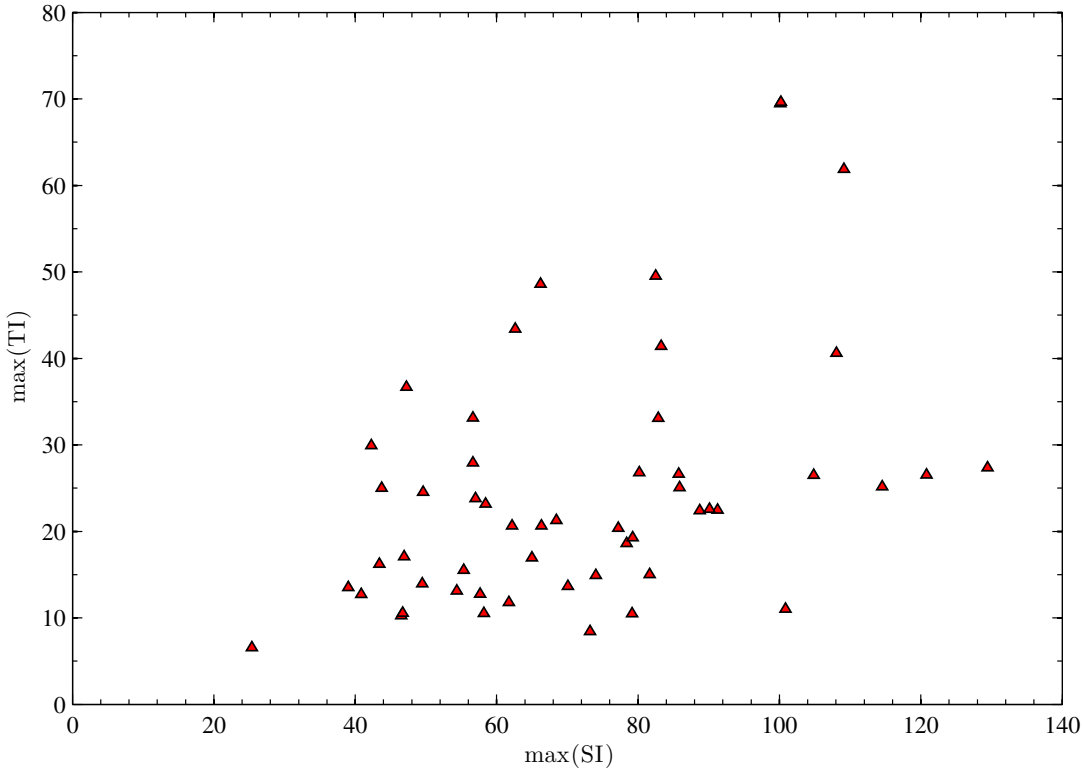

Figura 66: Diagrama max(TI) *vs.* max(SI) para base de dados do superconjunto *S* em HDi com 53 vídeos de referência.

APÊNDICE B – RESULTADOS ADICIONAIS

A seguir são apresentados resultados complementares ao Capítulo 5. Além das medidas de acurácia (PLCC) e da distribuição F percentual (F_p) apresentadas nesse capítulo, outras medidas estatísticas são utilizadas para quantificar o desempenho de uma métrica em relação à MOS ou à outra métrica. Para exemplificar o uso destas medidas, este apêndice apresenta as medidas RMSE, SROCC (monotonicidade), R-quadrado, OR (consistência) e a MAE do método NRVQA-LM para os conteúdos da base de dados IVP (experimento B).

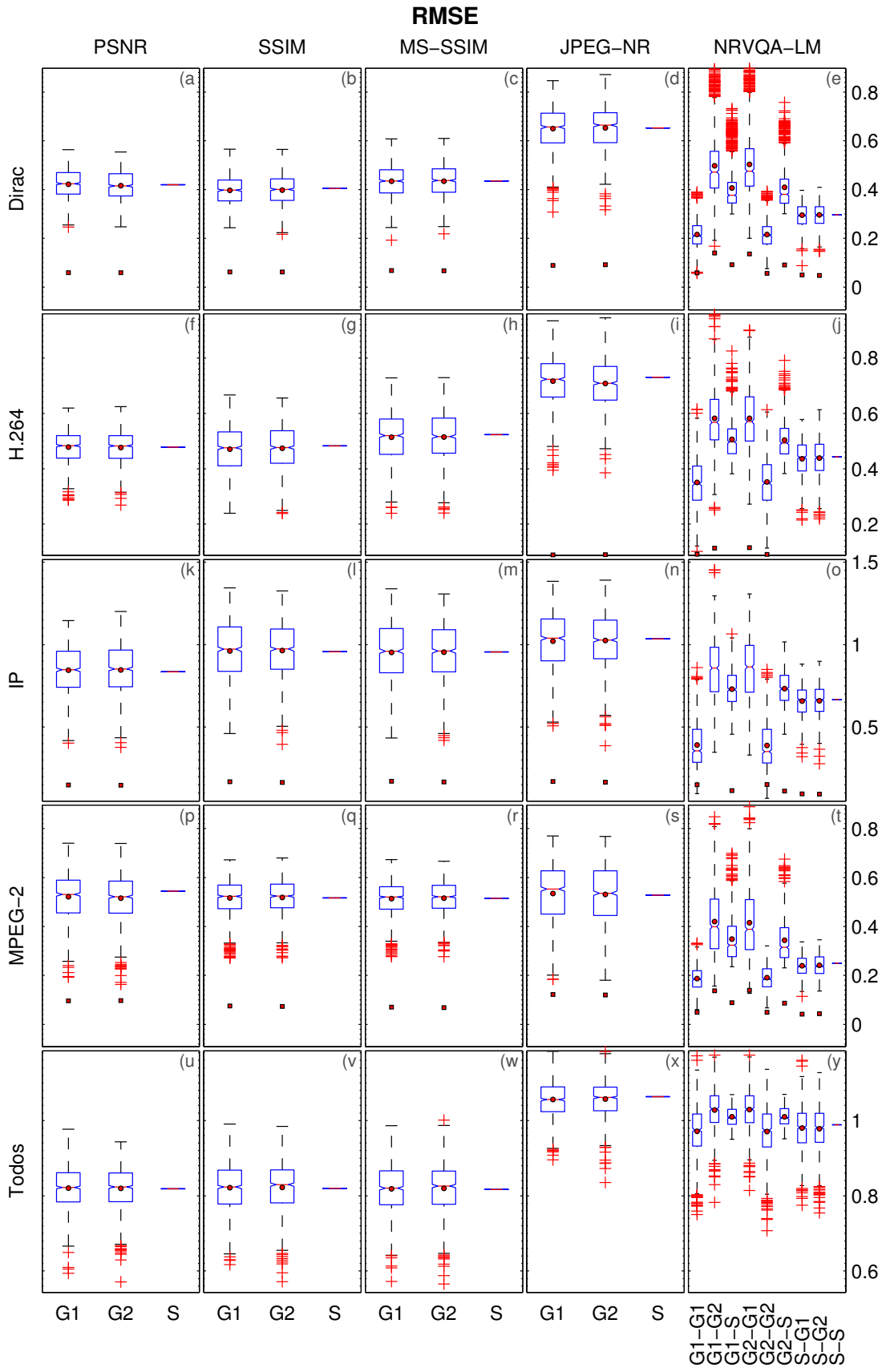

Figura 67: Comparação do RMSE para a base de dados IVP entre as métricas PSNR, SSIM, MS-SSIM, JPEG-NR e NRVQA-LM.

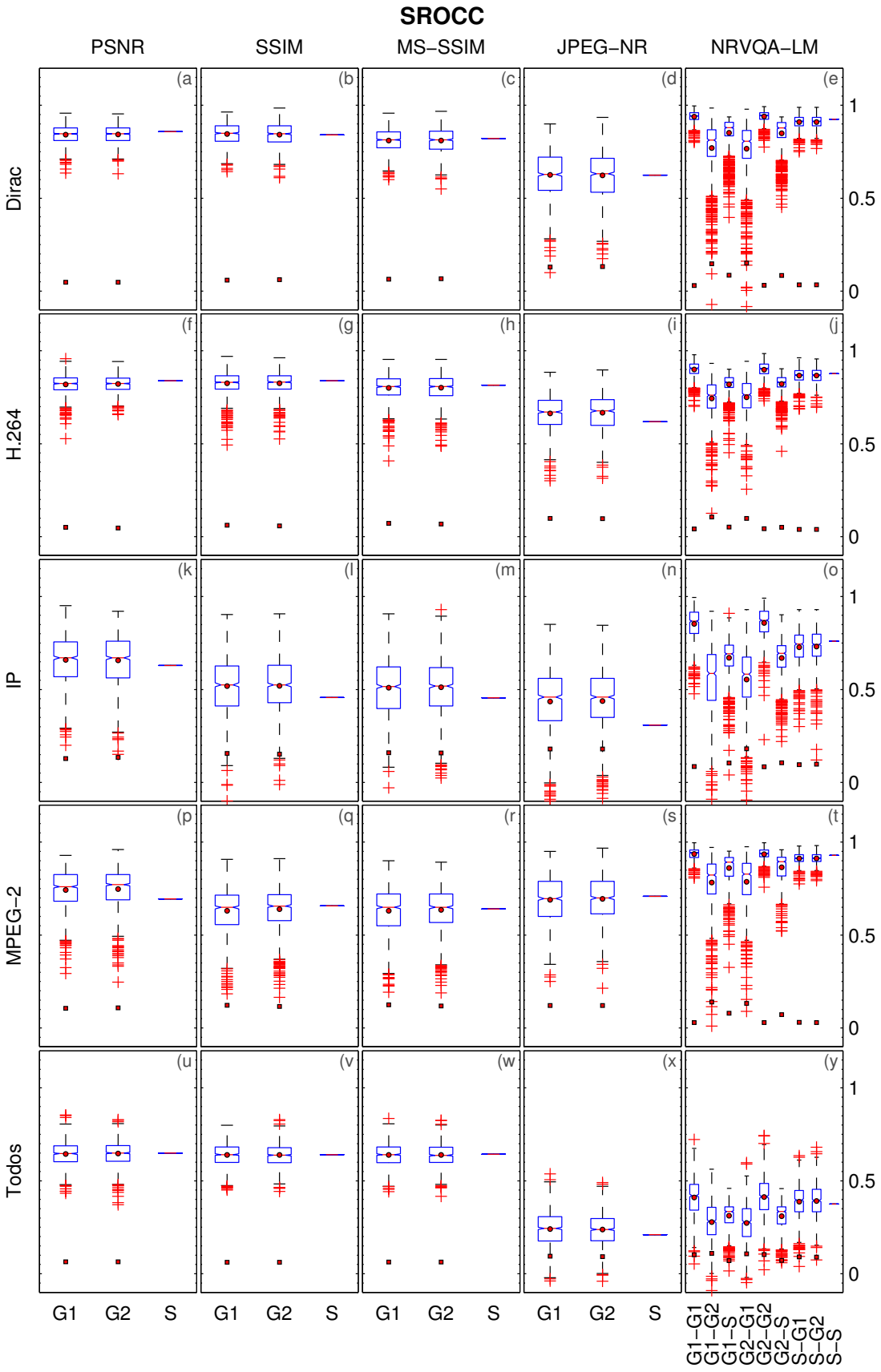

Figura 68: Comparação da monotonicidade (SROCC) para a base de dados IVP entre as métricas PSNR, SSIM, MS-SSIM, JPEG-NR e NRVQA-LM.

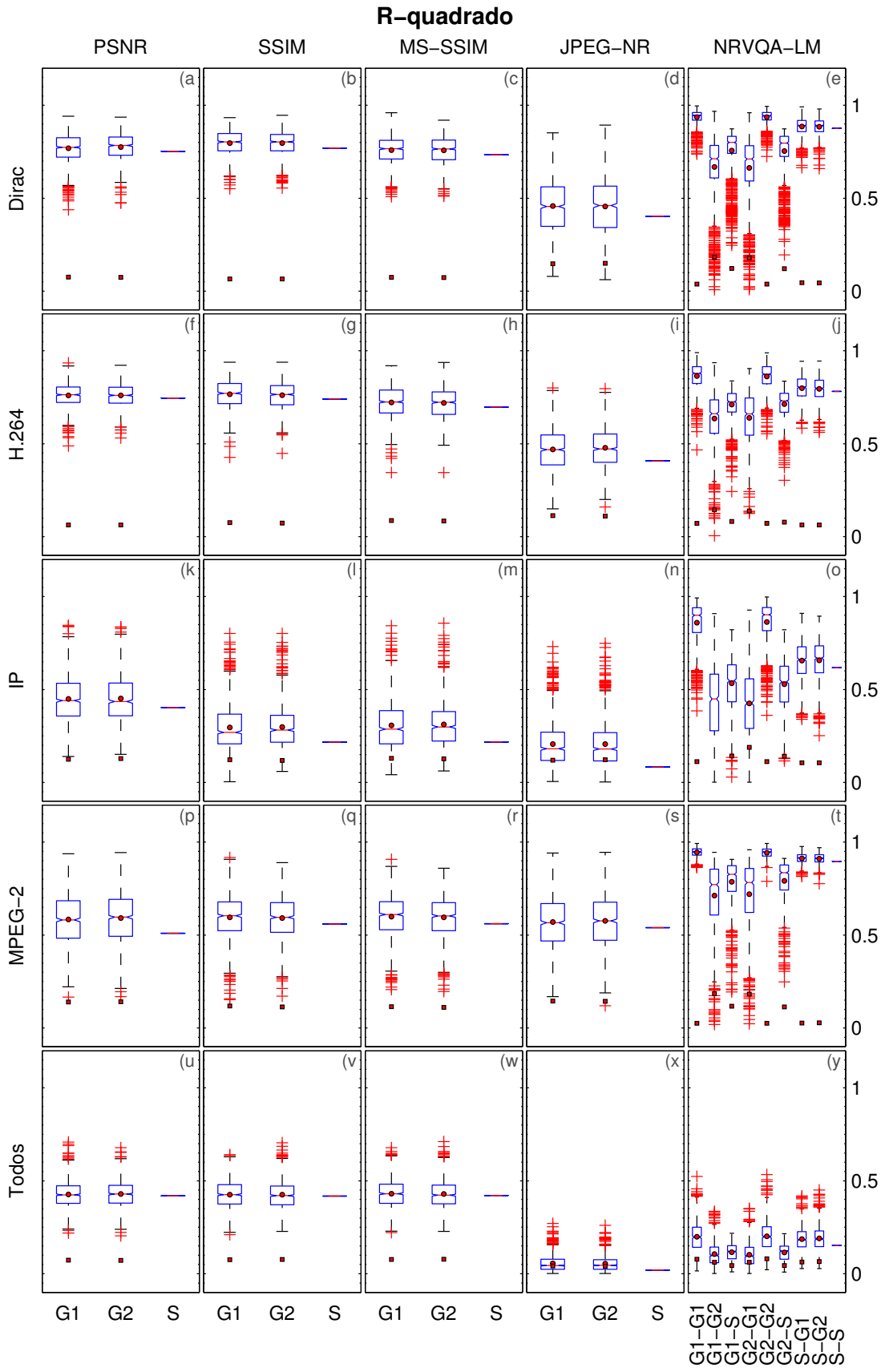

Figura 69: Comparação da medida R-quadrado para a base de dados IVP entre as métricas PSNR, SSIM, MS-SSIM, JPEG-NR e NRVQA-LM.

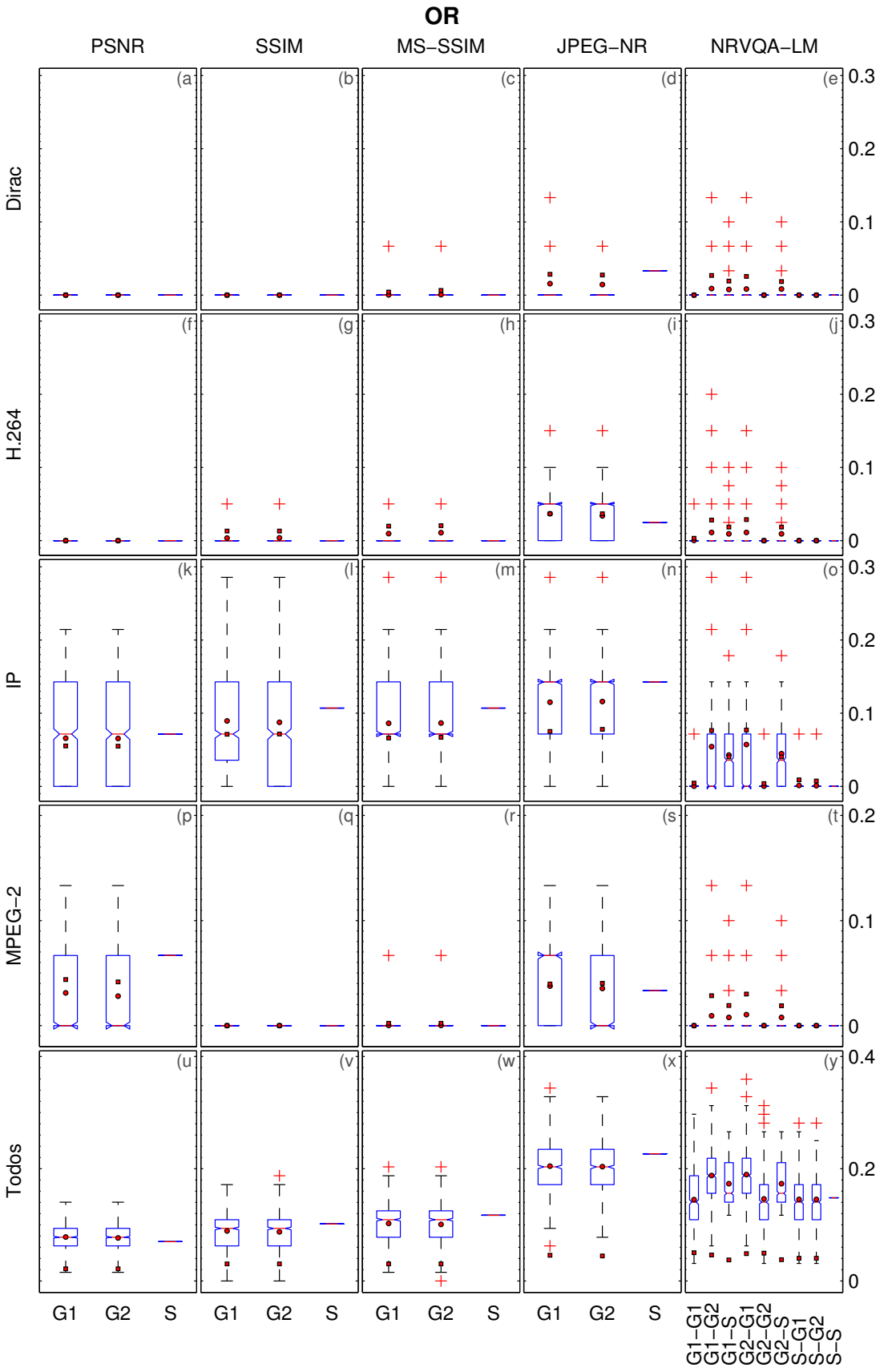

Figura 70: Comparação da consistência (OR) para a base de dados IVP entre as métricas PSNR, SSIM, MS-SSIM, JPEG-NR e NRVQA-LM.

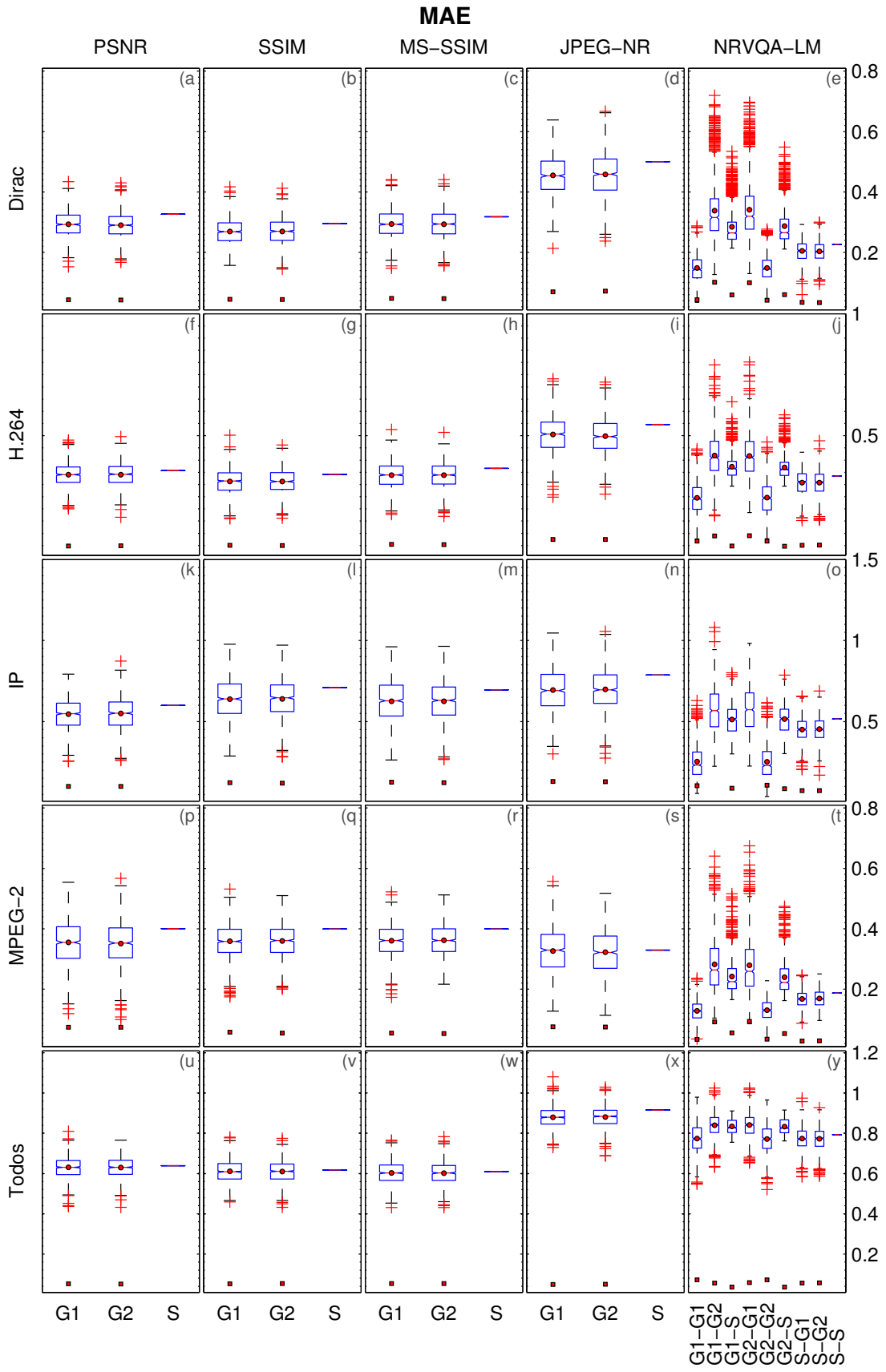

Figura 71: Comparação da medida MAE para a base de dados IVP entre as métricas PSNR, SSIM, MS-SSIM, JPEG-NR e NRVQA-LM.

APÊNDICE C – PRODUÇÃO ACADÊMICA

SILVA, W. B.; FONSECA, K. V. O.; POHL, A. A. P. A Reduced-Reference Video Quality Assessment Method based on the Activity-Difference of DCT Coefficients. **IEICE Transactions on Information and Systems**, vol. E96-D, n. 3, pp. 708-718, *Online* ISSN 1745-1361, *Print* ISSN: 0916-8532, DOI: 10.1587/transinf.E96.D.708, March, 2013.

SILVA, W. B.; POHL, A. A. P. No-Reference Video Quality Assessment Method based on Levenberg-Marquardt Minimization. **XXX Simpósio Brasileiro de Telecomunicações (SBrT'-12)**, Brasília, 2012, ISBN 9788589748070.

ROMANI, E.; SILVA, W. B.; BORBA, M. A. C.; FONSECA, K. V. O.; POHL, A. A. P. Ensaios de Recepção de Sinais de TV Digital em Dispositivo com Diversidade Espacial. **Congresso da Sociedade Brasileira de Engenharia e Televisão (SET'11)**, São Paulo, 2011, vol. 5, p. 32, ISBN 22369619.

SILVA, W. B.; POHL, A. A. P. Framework de Baixo Custo para Ensaios de TV Digital. **XXIX Simpósio Brasileiro de Telecomunicações (SBrT'11)**, Curitiba, 2011, ISBN 9788589748063.

SILVA, W. B.; POHL, A. A. P. Uma Nova Técnica de Detecção de Artefatos de Blocagem em Vídeo Comprimido. **XXIX Simpósio Brasileiro de Telecomunicações (SBrT'11)**, Curitiba, 2011, ISBN 9788589748063.

SILVA, W. B.; FONSECA, K. V. O.; POHL, A. A. P. Quality Performance of LD, SD and HD Video Encoded with Dirac and H.264/AVC. **International Telecommunications Symposium (ITS'10)**, 2010, n. 72687.

SILVA, W. B., POHL, A. A. P.; FONSECA, K. V. O. Um Modelo de Referência Completa Para Avaliação Objetiva da Qualidade de Vídeo em Dispositivos Móveis em Ambientes do Sistema Brasileiro de TV Digital (SBTVD). **Revista de Radiodifusão, Anais do Congresso da Sociedade Brasileira de Engenharia de Televisão (SET'09)**, São Paulo, 2009. vol. 3, pp. 324–337, ISSN 1982–4984.

SILVA, E. S. R.; FONSECA, K. V. O.; POHL, A. A. P.; SILVA, W. B. Sistema de Auxílio à Avaliação Subjetiva de Vídeo Digital. **XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia'09)**, Fortaleza. Artigos Curtos & Workshops. Fortaleza: Biblioteca Central da Universidade de Fortaleza, 2009, vol. 2, pp. 47–50, ISBN 9781605588803.