

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

METODOLOGIA PARA MEDIDAS OBJETIVAS DE
QUALIDADE DE VÍDEO EM SISTEMAS DE DIFUSÃO
DE CONTEÚDOS AUDIOVISUAIS

MÁRCIO LUCAS GRACIANO JÚNIOR

ORIENTADOR: JOSÉ CAMARGO DA COSTA

TESE DE DOUTORADO EM
ENGENHARIA ELÉTRICA

PUBLICAÇÃO: ENE/PG.TD - 071/13

BRASÍLIA/DF: MARÇO - 2013.

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

METODOLOGIA PARA MEDIDAS OBJETIVAS DE
QUALIDADE DE VÍDEO EM SISTEMAS DE DIFUSÃO
DE CONTEÚDOS AUDIOVISUAIS

MÁRCIO LUCAS GRACIANO JÚNIOR

TESE DE DOUTORADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA ELÉTRICA.

APROVADA POR:

Prof. José Camargo da Costa, Dr. (ENE-UnB)
(Orientador)

Prof. Mylène Christine Queiroz de Farias, PhD. (ENE-UnB)
(Examinador Interno)

Prof. Flávio Elias de Deus, DsC. (ENE-UnB)
(Examinador Interno)

Prof. Davies William de Lima Monteiro, PhD. (DEE/UFMG)
(Examinador Externo)

Prof. João Souza Neto, DsC. (UCB)
(Examinador Externo)

BRASÍLIA/DF, 25 DE MARÇO DE 2013.

FICHA CATALOGRÁFICA

Graciano Júnior, Márcio Lucas

Metodologia para Medidas Objetivas de Qualidade de Vídeo em
Sistemas de Difusão de Conteúdos Audiovisuais. /

Márcio Lucas Graciano Júnior. – 2013.

172 f. : il.

Tese de Doutorado - Faculdade de Tecnologia da Universidade de Brasília.
Departamento de Engenharia Elétrica, Brasília – 2013.

Orientação: Prof. Dr. José Camargo da Costa.

1. Qualidade de Vídeo 2. Métricas Objetivas 3. Marcas d'água
4. Função de Transferência de Modulação 5. Fator de Qualidade Subjetiva
I. ENE/FT/UnB II. Título.

REFERÊNCIA BIBLIOGRÁFICA

GRACIANO JR., M. L. (2013). Metodologia para Medidas Objetivas de Qualidade de Vídeo em Sistemas de Difusão de Conteúdos Audiovisuais. Tese de Doutorado em Engenharia Elétrica, Publicação ENE/PG.TD - 071/13, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 172p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Márcio Lucas Graciano Júnior.

TÍTULO DA TESE DE DOUTORADO: Metodologia para Medidas Objetivas de Qualidade de Vídeo em Sistemas de Difusão de Conteúdos Audiovisuais.

GRAU / ANO: Doutor / 2013

É concedida à Universidade de Brasília permissão para reproduzir cópias desta tese de doutorado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta tese de doutorado pode ser reproduzida sem a autorização por escrito do autor.

Márcio Lucas Graciano Júnior
SMPW 25 1 9H, Park Way
71.745-501 Brasília - DF - Brasil.

DEDICATÓRIA

Este trabalho é dedicado a
minha família, mãe, filhos, esposa,
irmãos e especialmente ao
Márcio Lucas Graciano,
o primeiro e melhor dos três.

AGRADECIMENTOS

Agradeço à Deus e ao apoio da minha família.

Agradeço ao meu orientador José Camargo da Costa, pela dedicação, estímulo e todos os tipos de apoio possíveis.

Agradeço aos Professores que aceitaram a participação na Qualificação e na Defesa da Tese e pelos valiosos aconselhamentos que enriqueceram o trabalho.

Agradeço ao Professor Luís Fernando Ramos Molinaro, que cedeu gentilmente o espaço do Estúdio do Laboratório NMI (Núcleo de Multimídia e Internet) para a realização dos Testes Subjetivos.

Agradeço a Professora Mylène Christine Queiroz de Farias, que cedeu a licença do software para controle e realização dos Testes Subjetivos, além da valioso aprendizado que obtive com o seu trabalho.

Agradeço ao apoio de todos os voluntários que participaram dos testes subjetivos, especialmente a participação dos colegas de trabalho da Anatel, dos meus filhos e seus colegas de escola.

Agradeço ao apoio dos meus colegas e superiores na Anatel pelo apoio durante o período compartilhado entre o trabalho e a tese e pela concessão da licença, que foi fundamental para a conclusão do trabalho.

Agradeço ao suporte do CNPq (Conselho Nacional de Pesquisa) e do INCT-Namitec.

RESUMO

METODOLOGIA PARA MEDIDAS OBJETIVAS DE QUALIDADE DE VÍDEO EM SISTEMAS DE DIFUSÃO DE CONTEÚDOS AUDIOVISUAIS

Autor: Márcio Lucas Graciano Júnior

Orientador: José Camargo da Costa

Programa de Pós-graduação em Engenharia Elétrica

Brasília, março de 2013

Este trabalho apresenta uma metodologia para a verificação da qualidade de vídeos em sistemas de difusão em massa. A metodologia proposta apresenta uma nova forma de medir a qualidade de vídeo de forma objetiva sem a utilização de referência para comparação entre as sequências de vídeo original e processada. As medidas no ambiente do usuário são feitas independentemente dos processamentos e da tecnologia da rede de distribuição.

A medida da qualidade é feita por meio da inserção de marcas de testes nos vídeos, logo após a produção destes. O processamento e distribuição desses vídeos em um sistema de difusão em massa podem causar degradações que afetam as marcas inseridas. Os vídeos recebidos no ambiente do usuário são captados por meio de uma câmera de vídeo disponível em conjunto com o terminal, que identifica a marca inserida anteriormente à distribuição. Esta captura é feita pelo sistema óptico do ambiente do usuário após uma calibração do sistema. As marcas recebidas passam por um processamento de uma função de qualidade que tem como resultado um valor numérico indicando a qualidade deste vídeo.

Como resultados da metodologia, são apresentadas comparações com sistemas de medição de qualidade de vídeo de referência completa utilizados em testes por organismos de padronização. As comparações com os sistemas VQM (*Video Quality Metrics*) e SSIM (*Structural SIMilarity*) foram feitas utilizando várias sequências de vídeo de testes, onde foi alcançada uma correlação estatística maior que 80% entre as medidas de qualidade resultantes da nova metodologia e as desses sistemas de referência.

ABSTRACT

METHOD FOR OBJECTIVE MEASUREMENTS OF VIDEO QUALITY IN BROADCASTING SYSTEMS

Author: Márcio Lucas Graciano Júnior

Supervisor: José Camargo da Costa

Programa de Pós-graduação em Engenharia Elétrica

Brasília, March of 2013

This work presents a method for verifying the quality of videos on mass broadcasting systems. The proposed method presents a new way of measuring the objective video quality without reference. The measurements in the user's environment are made regardless of the processing and the broadcasting network technology.

A quality measure is made by the insertion of test marks on videos, immediately after the content production. The processing and distribution of these videos on a mass broadcasting system can cause degradations that affect the inserted marks. The video received in the user's environment is captured by a video camera available along with the video terminal. The captured video should identify the mark inserted before the distribution. This capture is made by the optical system in the user environment after a system calibration. The mark is decoded by the processing of a quality function which results in a numerical value indicating the quality of the video.

The results of video quality metrics using this methodology were compared to standardized full reference metrics, VQM (Video Quality Metrics) and SSIM (Structural Similarity), and the linear correlation between this proposed new metrics and the other two metrics was greater than 80%, indicating convergence between these metrics.

SUMÁRIO

1	Introdução	1
1.1	Motivação	1
1.2	Objetivos	3
1.3	Apresentação da Proposta	4
1.4	Organização deste documento	4
2	Referencial Teórico	5
2.1	Introdução	5
2.2	Qualidade de Vídeo	5
2.2.1	Organismos de Padronização	9
2.2.2	Revisão Bibliográfica de Qualidade de Vídeo Objetiva	12
2.2.3	Testes Subjetivos	20
2.3	Marcas d'água Digitais	28
2.3.1	Marcas d'água Invariantes à Rotação, Escala e Translação	33
2.4	Função de Transferência de Modulação	38
2.4.1	Método de estimativa da Resolução – MTF	45
2.5	Acuidade visual humana	49
2.6	Fator de Qualidade Subjetiva – SQF (<i>Subjective Quality Factor</i>)	50
2.6.1	Relação entre o SQF e a MTF	51
3	Metodologia	53
3.1	Passos da Metodologia	53
3.1.1	Etapa 1 - Seleção dos Vídeos Originais	55
3.1.2	Etapa 2 - Marcação dos Vídeos Originais	56
3.1.3	Etapa 3 - Vídeos Originais Marcados	57
3.1.4	Etapa 4 - Processamento e Degradações dos Vídeos	59
3.1.5	Etapa 5 - Vídeos Marcados Processados	59
3.1.6	Etapa 6 - Testes Subjetivos	60
3.1.7	Etapa 7 - Resultados DMOS	61
3.1.8	Etapa 8 - Detecção da Marca pela câmera de vídeo	61

3.1.9	Etapa 9 - Parâmetros Ópticos	64
3.1.10	Etapa 10 - Cálculo da MTF e do SQF	65
3.1.11	Etapa 11 - Fator de Qualidade Subjetivo	66
3.1.12	Etapa 12 - Ajuste da Função de Qualidade	67
3.1.13	Etapa 13 - Função de Qualidade	68
3.2	Calibração e Registro do sistema	68
3.3	Verificação da Qualidade em Operação	72
4	Resultados e Discussões	74
4.1	Introdução	74
4.2	Testes Subjetivos	74
4.2.1	Aparato Utilizado	74
4.2.2	Procedimento de testes	77
4.2.3	Resultados DMOS dos Testes Subjetivos	81
4.3	Função de Qualidade - Métrica de Qualidade	85
4.4	Cálculo da MTF	85
4.5	Cálculo do SQF	88
4.6	Testes subjetivos e Métricas de referência	91
4.7	Comparação das Métricas de Qualidade	94
5	Conclusões	97
5.1	Contribuições	98
5.2	Próximos Passos	99
	REFERÊNCIAS BIBLIOGRÁFICAS	101
	APÊNDICES	114
A	Avaliação de Marcas de Testes para Inserção nos Vídeos	115
A.1	Utilização de Super amostragem	118
A.2	Utilização de Modulação de linhas horizontais do vídeo	118
A.3	Utilização de Marcas d'água digitais no vídeo	119
A.4	Resultados Obtidos	119
B	Testes APS-UFRJ NAMITEC	122
C	Trabalho IEEE-CNCC 2010	146
D	Trabalho IJCA-2013	149

LISTA DE TABELAS

2.1	Principais degradações presentes em sinais de vídeo	10
2.2	Métricas de Qualidade de Vídeo	21
3.1	Características dos Circuitos de Referência - HRC	59
3.2	Parâmetros da Câmera e do Monitor	65
3.3	Sequências utilizadas no vídeo de calibração	71
4.1	Lista de Equipamentos utilizados	76
4.2	Sequências de Vídeo utilizadas	77
4.3	Resultados dos Testes Subjetivos e Função de Qualidade	81
4.4	Resultados da Função de Qualidade	90
A.1	Novos Tamanhos dos Vídeos com Super Amostragem	120

LISTA DE FIGURAS

2.1	Modelo de Referência completa (FR).	9
2.2	Modelo de Referência Reduzida (RR).	11
2.3	Modelo Sem Referência (NR).	11
2.4	Posição do Observador	25
2.5	Método DSCQS.	25
2.6	Gráfico característico da função logística	28
2.7	Esteganografia e marcas d'água.	29
2.8	Tipos de Marcas d'água.	30
2.9	Algoritmo de inserção da marca d'água	34
2.10	Característica de Transferência do Sistema de Imagens	39
2.11	Resolução Espacial.	40
2.12	Função de Espalhamento de Ponto - PSF.	41
2.13	Função de Espalhamento de Linha - LSF.	42
2.14	Função de Espalhamento de Borda - ESF.	43
2.15	MTF em função da frequência espacial	44
2.16	MTF e contraste	45
2.17	Imagem binária e método para o cálculo do ângulo	46
2.18	Representação dos passos para o cálculo da MTF	47
2.19	Diagrama de passos para o cálculo da MTF	48
2.20	MTF do olho humano.	49
2.21	Função de sensibilidade ao contraste do olho humano.	50
3.1	Etapas da Metodologia Proposta.	54
3.2	Método para geração da função de qualidade.	54
3.3	Quadro da Sequência de Vídeo <i>Park Joy</i> .	56
3.4	Sequência de Vídeo <i>Park Joy</i> Normalizada.	57
3.5	Áreas da Sequência de Vídeo <i>Park Joy</i> .	58
3.6	Quadro da Sequência de Vídeo <i>Park Joy</i> Marcada.	58
3.7	Quadro da Sequência de Vídeo <i>Park Joy</i> processada.	60
3.8	Valores DMOS para a Sequência de Vídeo <i>Park Joy</i> .	61
3.9	Detecção da Marca.	62

3.10	Detecção de Marcas na Sequência de Vídeo <i>Park Joy</i>	64
3.11	Diagrama da MTF do sistema óptico	65
3.12	MTF da Sequência de Vídeo <i>Park Joy</i>	67
3.13	SQF da Sequência de Vídeo <i>Park Joy</i>	67
3.14	Calibração do sistema no ambiente do usuário	69
3.15	Diagrama ISO12233 utilizado como referência para o vídeo de calibração.	70
3.16	Procedimento de operacional para verificação da qualidade de vídeo	73
4.1	Diagrama do aparato experimental.	75
4.2	Fotos do espaço físico do Laboratório - Foto 1	80
4.3	Fotos do espaço físico do Laboratório - Foto 2	80
4.4	Resultados dos Testes Subjetivos	83
4.5	Correlação entre os testes subjetivos e a métrica proposta	84
4.6	MTF Total e MTF dos dispositivos Eletrônicos	86
4.7	Função de Transferência de Modulação - MTF	87
4.8	Fator de Qualidade Subjetivo - SQF	89
4.9	Correlação entre o teste subjetivo e a métrica proposta	91
4.10	Correlação entre o teste subjetivo e a métrica de referência VQM	92
4.11	Correlação entre o teste subjetivo e a métrica de referência SSIM	92
4.12	Correlação entre o teste subjetivo e a métrica de referência PSNR	93
4.13	Correlação entre a métrica proposta e o sistema de Referência VQM	95
4.14	Correlação entre a métrica proposta e a Referência SSIM	96
4.15	Correlação entre a métrica proposta e a Referência PSNR	96
A.1	Resultados da aplicação com relação ao VQM	121

LISTA DE SÍMBOLOS, NOMENCLATURA E ABREVIACÕES

ANSI: American National Standard Institute, Instituto Nacional Americano de Padronização.

APS: Active Pixel Sensor, Sensor de Imagem CMOS de pixel ativo.

CB: Component Blue, Componente de diferença de cor azul.

CCNC: Consumer Communications & Networking Conference, Conferência de Comunicações e Redes de Consumidores.

CMOS: Complementary Metal Oxide Semiconductor, Tecnologia Semicondutor Metal Óxido Complementar.

CNPq: Conselho Nacional de Pesquisa.

cpd: cycles per degree, ciclos por grau.

cpp: cycles per pixel, ciclos por pixel.

CR: Component Red, Componente de diferença de cor vermelho.

CSF: Contrast Sensitivity Function, Função de Sensibilidade ao Contraste do Olho Humano.

DCT: Discrete Cosine Transform, Transformada Discreta de Cossenos.

DEE: Departamento de Engenharia Elétrica da UFMG.

DFT: Discrete Fourier Transform, Transformada Discreta de Fourier.

DMOS: Difference Mean Opinion Scores, Diferenças de Médias de Respostas dos Observadores.

DSCQS: Double Stimulus Continuos Quality Scale, Escala de Qualidade Contínua de Estímulo Duplo.

DWT: Discrete Wavelet Transform, Transformada Discreta Wavelet.

ENE: Departamento de Engenharia Elétrica da UnB.

ESF: Edge Spread Function, Função de Espalhamento de Borda.

FQ: Função de Qualidade.

FR: Full Reference, Referência Completa.

HD: High Definition, Alta Definição.

HDTV: High Definition Television, Televisão em Alta Definição.

HRC: Hypothetical Reference Circuit, Circuito Hipotético de Referência.

IEEE: Institute of Electrical and Electronic Engineers, Instituto dos Engenheiros Eletricistas e Eletrônicos.

IJCA: International Journal of Computer Applications, Jornal Internacional de Aplicações de Computadores.

INCT: Instituto Nacional de Ciência e Tecnologia.

ISO: International Organization for Standardization, Organização Internacional para Padronização.

ITU: International Telecommunications Union, União Internacional de Telecomunicações.

LMSP: Locally Most Stable Points, pontos de interesse com maior estabilidade local.

lp/mm: line pair per millimeter, pares de linha por milímetro.

LSF: Line Spread Function, Função de Espalhamento de Linha.

Mbps: Mega bit por segundo.

MOS: Mean Opinion Scores, Média de Respostas dos Observadores.

MPEG: Motion Picture Experts Group, Grupo de Especialistas em Vídeo.

MSE: Mean Squared Error, Erro Quadrático Médio.

MTF: Modulation Transfer Function, Função de Transferência de Modulação.

NC: Normalized Correlation, Correlação Normalizada.

NMI: Núcleo de Multimídia e Internet da UnB.

NR: No Reference, Sem Referência.

NTSC: National Television System Comitee, Comitê Nacional Americano de Padronização do Sistema de Televisão.

PSF: Point Spread Function, Função de Espalhamento de Ponto.

PSNR: Peak Signal to Noise Ratio, Razão Sinal Ruído de Pico.

PVS: Processed Video Sequence, Sequência de Vídeo Processada.

RAM: Random Access Memory, Memória de leitura e escrita de acesso aleatório.

RR: Reduced Reference, Referência Reduzida.

RST: Rotation, Scale, Translation Invariance, Invariante à rotação, escala e translação.

SDTV: Standard Definition Television, Televisão de definição padrão.

SoC: System on Chip, Sistema em Chip.

SQF: Subjective Quality Factor, Fator de Qualidade Subjetivo.

SRC: Source Sequence, Sequência de Vídeo de referência.

SSD: Solid State Disk, Disco rígido utilizando memória de estado sólido.

SSIM: Structural Similarity, Métrica Similaridade Estrutural.

TV: Television, Televisão.

UCB: Universidade Católica de Brasília.

UFMG: Universidade Federal de Minas Gerais.

UFRJ: Universidade Federal do Rio de Janeiro.

UnB: Universidade de Brasília.

VQEG: Video Quality Experts Group, Grupo de Especialistas em Qualidade de Vídeo.

VQM: Video Quality Metrics, Métrica de Qualidade de Vídeo.

Y: Componente de Luminância.

1 Introdução

Os sistemas de medição da qualidade de vídeo encontram-se em fase de desenvolvimento pela indústria e por organizações de padronização e podem ser encontrados principalmente em laboratórios de testes e na academia, não estando disponíveis em aplicações em larga escala para os usuários [1]. Estes sistemas, quando aplicados às redes de distribuição de vídeos, fazem medidas de qualidade com o foco na rede e não na percepção do usuário em seu ambiente de visualização destes conteúdos audiovisuais.

Existem várias formas de apresentação de vídeo atualmente, desde o vídeo analógico até os mais diversos tipos de vídeos digitais que podem utilizar vários tipos de codificadores, compressores e transformadas, além de novos desenvolvimentos que surgem a cada dia [2]. O grande avanço dos sistemas de televisão em utilização no mundo [3][4] permite que a maioria dos países desenvolvidos e em desenvolvimento façam a migração para tecnologias digitais de transmissão [5]. No panorama mundial, ainda existe grande diversidade de padrões de tecnologias para a transmissão de televisão digital, bem como cronogramas distintos e grandes diferenças regionais e econômicas [3][4][6]. Nos países mais ricos e principalmente nas áreas de maior atratividade econômica, a digitalização está em estágio adiantado [3][4][5][6].

Na cadeia de distribuição de conteúdo audiovisual, é comum que a parte principal desta rede, que centraliza a produção, geração e difusão, o processamento seja realizado com modernos equipamentos digitais. Na rede de distribuição, mais capilarizada, que atinge a maioria dos espectadores, existem maior quantidade e diversidade de equipamentos terminais e receptores, desde os mais modernos digitais aos mais baratos analógicos, necessitando de conversões digital para analógico ao longo da cadeia. Esta diversidade obriga a transmissão simultânea de sinais analógicos e digitais durante o período de transição da tecnologia [6].

1.1 Motivação

A maior parte dos trabalhos na área de qualidade de vídeo leva em conta a qualidade

voltada para aspectos técnicos da cadeia de distribuição do vídeo, efetuando as medidas de qualidade em equipamentos que podem estar no centro da rede, ou mesmo quando na ponta do usuário, não reflete a qualidade final observada pelo usuário em seu terminal. A verificação de qualidade medida próxima aos usuários necessita utilizar dispositivos de captura de informações (*probes*) em grande número, para cada usuário, o que poderia elevar bastante os custos para o sistema. Em alguns receptores digitais modernos, que contam com capacidade de processamento e memória livre, a verificação da qualidade poderia ser feita diretamente no receptor do cliente, incluindo códigos que utilizariam essa capacidade de processamento para a verificação da qualidade. Considerando dados que a maior parte das reclamações para uma operadora de distribuição de vídeo são relacionados a problemas com equipamentos de usuários [7], devemos nos preocupar em interferir o mínimo nestes equipamentos.

Para que seja verificada a qualidade dos conteúdos audiovisuais em toda a cadeia de vídeo, desde a produção do conteúdo, distribuição nos mais diversos meios e principalmente o consumo pelo espectador, verificando a qualidade pelo seu ponto de vista, é necessário que as medidas de qualidade sejam feitas o mais próximo possível do espectador. A proposta é desenvolvermos dispositivos externos aos equipamentos dos usuários para a verificação da qualidade do vídeo. Desta forma, teremos uma mínima interferência na prestação do serviço de distribuição, evitando a troca de toda a base de equipamentos terminais, incluindo os atuais prestadores com seus sistemas legados analógicos ou digitais de baixa capacidade. Estes dispositivos apresentarão a vantagem de medir a qualidade em condições mais próximas às visualizadas pelo espectador, independentemente do número de conversões múltiplas entre formatos analógicos e digitais que aconteçam na cadeia de distribuição do vídeo.

Do ponto de vista da regulamentação brasileira para a aferição da qualidade das redes de telecomunicações, os órgãos de controle [8] verificaram que os órgãos reguladores focam seu trabalho na regulamentação técnica da qualidade das redes de telecomunicações, mas o foco na qualidade percebida pelo usuário não é considerado na mesma proporção.

No Brasil, os órgãos reguladores das comunicações, no caso da Anatel, que trata da regulamentação de telecomunicações [9] e do Ministério das Comunicações, que trata da regulamentação da radiodifusão[10][11], tratam duas formas de distribuição de conteúdos audiovisuais em sistemas de comunicação em massa com uma pequena diferença do ponto de vista da legislação e da regulamentação [12]. Apesar do trata-

mento regulatório diferente, do ponto de vista tecnológico, podemos considerar estas duas formas como distribuição de conteúdos audiovisuais em redes de distribuição em massa.

A regulamentação para o setor de televisão no Brasil encontra-se defasada em relação à tecnologia, com o foco nas redes analógicas de distribuição [13]. Nos últimos anos, o debate sobre um novo marco para as comunicações vem sendo discutido pela sociedade, resultando na edição de uma nova lei para os serviços de acesso condicionado (Televisão por Assinatura) [14] que se encontra regulamentada pelas agências responsáveis por estes serviços, a Anatel no que diz respeito a distribuição [15] e a Ancine no que diz respeito a produção e empacotamento.

A nova regulamentação da distribuição de Televisão por Assinatura [15] é independente das tecnologias de distribuição de vídeo, permitindo que evoluções tecnológicas sejam aderentes à regulamentação, uma vez que esta estará focada na prestação do serviço e atendimento aos usuários.

Outro aspecto da regulamentação que é dificultado para um bom atendimento aos usuários é a fiscalização insuficiente dos serviços, pela grande quantidade de prestadores de serviços e de usuários dispersos em um país continental, bem como pela falta de recursos. Um sistema de verificação da qualidade distribuído e próximo aos usuários, verificado e certificado pelos órgãos competentes, poderia criar uma nova categoria de “*fiscalmidores*” (fiscais consumidores), o que elevaria a qualidade do serviço para todos, uma vez que o principal interessado poderia atuar em tempo real em caso de degradação de qualidade.

1.2 Objetivos

O objetivo deste trabalho é o desenvolvimento de uma metodologia para aferição da qualidade de vídeo de forma objetiva em sistemas de distribuição em massa de vídeo de forma automática sem consulta ou arbitragem pelo usuário. Deve ser permitida a medida da qualidade de vídeo independente do tipo da rede, do tipo do vídeo e do tipo de dispositivo terminal do usuário. Deve ainda permitir o monitoramento da qualidade em pontos próximos aos usuários que se comportem como eles. A metodologia deve ser implementada com dispositivos simples, de baixo custo e complexidade para a utilização no ambiente do usuário. Os dispositivos devem interferir minimamente

com os equipamentos já em utilização pelo distribuidor de vídeo. Deve ser possível a utilização da metodologia em dispositivos com pequena capacidade de processamento em medidas em tempo real sem referência. O resultado da medida de qualidade objetiva através desta metodologia deve ter tendência de comportamento equivalente aos sistemas de medida utilizando referência completa. Deve ser possível a utilização da metodologia em sistemas em chip (SoC) com sensores de imagem e ainda em aparelhos móveis com sensores de imagem. Com esta metodologia e dispositivos será possível aos consumidores fazer uma fiscalização em tempo real da qualidade dos vídeos em toda a cadeia de distribuição, desta forma alertando aos interessados e melhorando a qualidade do sistema como um todo.

1.3 Apresentação da Proposta

A metodologia proposta apresenta uma métrica de qualidade de vídeo que apresenta como principal vantagem o fato de ser uma métrica sem referência, além de contar com características do sistema visual humano e também características estatísticas das principais formas de degradação presentes nos sistemas de vídeo.

A Seção 2.2.2 apresenta o histórico do desenvolvimento de métricas objetivas de qualidade de vídeo, suas características e o seu estágio de desenvolvimento.

1.4 Organização deste documento

No capítulo 2 são apresentados os principais conceitos teóricos utilizados para o desenvolvimento desta pesquisa e uma revisão bibliográfica das principais métricas objetivas de qualidade de vídeo com características da métrica apresentada nesta pesquisa. Em seguida, o capítulo 3 descreve os passos da metodologia empregada no desenvolvimento da pesquisa. Resultados experimentais são discutidos no capítulo 4, seguidos pelas conclusões e trabalhos futuros no capítulo 5. Os apêndices apresentam material complementar. O apêndice A apresenta um estudo para a definição do tipo de marca d'água a ser utilizada. O apêndice B reproduz um trabalho realizado no decorrer do desenvolvimento desta tese relacionado à utilização e teste de sensores APS, que poderá ser utilizado em trabalhos futuros. Os apêndices C e D reproduzem as publicações feitas no decorrer da pesquisa.

2 Referencial Teórico

2.1 Introdução

Neste capítulo é apresentado o referencial teórico utilizado para a construção da metodologia que está sendo desenvolvida. Esta metodologia envolve medidas objetivas de qualidade de vídeo e avaliação subjetiva de qualidade feita por avaliadores humanos. Para a verificação da qualidade de vídeo são utilizadas marcas d'água digitais semi-invisíveis inseridas durante a geração dos vídeos, e as marcas d'água recuperadas no ambiente de recepção do usuário são avaliadas através das propriedades ópticas do ambiente e dos equipamentos na recepção.

Nas seções seguintes são apresentados os conceitos utilizados nesta pesquisa referentes a Qualidade de Vídeo, os tipos de marcas d'água digitais e suas propriedades, além da utilização da Função de Transferência de Modulação do sistema óptico do usuário.

2.2 Qualidade de Vídeo

As métricas de qualidade de vídeo podem ser classificadas em métricas de dados ou métricas de imagem [16]. As primeiras medem a fidelidade do sinal sem considerar o seu conteúdo e as últimas estimam a qualidade considerando a informação visual contida nos dados. As métricas de dados são amplamente utilizadas na literatura devido a simplicidade, significado físico e facilidade de implementação. As métricas de imagem podem ser divididas em três categorias de acordo com a disponibilidade do sinal de referência original [16]. As métricas que utilizam o vídeo original como referência, chamadas de métricas com referência completa ou FR (*full reference*), as métricas que utilizam algum detalhe do vídeo original como referência, chamadas de métricas com referência reduzida ou RR (*reduced reference*) e as métricas que utilizam apenas o vídeo recebido e a partir deste vídeo estimam sua qualidade, estas métricas são conhecidas como sem referência ou NR (*no reference*).

A qualidade da transmissão do vídeo em tempo real pode variar enormemente. Elementos como resolução de *pixels*, tamanho da imagem e as taxas de quadros (*frame rate*) têm grande influência sobre a percepção do usuário.

A tecnologia para disponibilização de vídeo em telas de pequenas dimensões, caso dos receptores móveis portáteis, está em evolução rápida. O valor destes vídeos para os usuários finais também está em rápida ascensão. Existem muitos fatores que influenciam o valor de um serviço aos usuários finais, tais como conteúdos, preço, usabilidade, conveniência, bem como a qualidade da imagem de vídeo em si.

Do ponto de vista da qualidade de vídeo percebida pelo usuário, os fatores que mais causam desconforto são aqueles sobre os quais o operador da rede de distribuição de vídeo pode ter controle, ou pelo menos deseja ter controle:

- taxa de bits de transmissão;
- taxa de perda de pacotes;
- tamanho do formato;
- tipo do codificador;
- complexidade do conteúdo.

Os sistemas de comunicações digitais, especialmente os de alta capacidade como os sistemas de vídeo, evoluíram bastante nos últimos anos. Devido à grande quantidade de informações contidas nestes vídeos e a busca por maior qualidade, como vídeos em alta definição, cada vez mais surgem necessidades de novos desenvolvimentos por parte da comunidade científica e pela indústria. Dentre estes desenvolvimentos, as técnicas de compressão e transmissão são as mais importantes. Muitos dos novos serviços que vêm surgindo foram possíveis graças a estas novas técnicas, que permitem uma grande variedade de transformações entre padrões, tipos de dispositivos para apresentação e meios de transmissão.

Para a verificação da qualidade de uma cadeia de distribuição de vídeo é importante termos uma forma de medir as variações de qualidade em cada etapa desta cadeia. Normalmente, essas medidas são feitas levando em consideração uma comparação direta

ou indireta entre a sequência de vídeo original e a sequência de vídeo sob testes. Nesta comparação, a qualidade pode permanecer a mesma, aumentar ou diminuir [17][18].

Nos sistemas de medição de qualidade de vídeo, a sequência de vídeo original normalmente é um vídeo de alta qualidade supostamente sem degradações e os testes são feitos com sequências de vídeo sob teste onde a qualidade permanece a mesma ou tem a qualidade diminuída através da inserção de degradações [18].

Atualmente, a forma mais confiável para a medição da qualidade de vídeo emprega testes subjetivos, utilizando a percepção dos usuários tabuladas através das médias de observações de vários espectadores. Estas medidas seguem regras padronizadas [19][20]. Nesses sistemas de testes, a melhor forma de verificação da qualidade é a utilização da capacidade dos espectadores de avaliação utilizando experimentos psicofísicos com seres humanos [21].

Os métodos subjetivos são de natureza especificamente não estruturada, incluindo entrevistas abertas ou observação participativa e têm a vantagem de proporcionar riqueza de informações. A desvantagem é que todas essas informações têm de ser exaustivamente analisadas. Qualquer informação que seja proveniente de usuários, peritos ou observadores pode ser considerada subjetiva, ao passo que se a informação pode ser gravada sem viés potencial ou que seja armazenada em arquivos é considerada objetiva [22].

Os métodos objetivos de coleta de dados muitas vezes resultam em informações claras, mas geralmente são bastante limitados no escopo para interpretação. Certas medidas sistemáticas podem ser gravadas objetivamente e automaticamente através de um sistema, mas processos complexos, tais como a interação do usuário são difíceis de registrar desta forma. A análise objetiva dos dados gravados pode frequentemente consumir muito tempo.

O desenvolvimento de metodologias para avaliação objetiva de vídeo tem como meta substituir as avaliações subjetivas, feitas com grupos controlados de observadores e que exige altos recursos, de custo e de tempo, o que os tornam difíceis de serem utilizados em aplicações práticas. Dentro deste contexto, procuram-se métodos de avaliação objetiva que se aproximem da qualidade dos resultados encontrados para a avaliação subjetiva, e que reduzem consideravelmente os custos do processo [18][22].

Métodos objetivos para medir o desempenho de vídeo digital estão sendo cada vez mais solicitados por Governos e pela indústria para especificação de requisitos de desempenho de sistemas, para comparação das ofertas de serviços competidores, acordos de nível de serviço, manutenção de rede, e otimização do uso dos recursos escassos de rede, tais como a largura de banda e espectro [23].

Métodos objetivos que forneçam uma medida física da qualidade de vídeo foram desenvolvidos nos últimos anos graças ao desenvolvimento de novos algoritmos e da facilidade de acesso a recursos computacionais [22][24][25][26][27][28][29]. A maioria dessas métricas objetivas propostas necessitam da sequência de vídeo original para a estimativa de qualidade [18].

Para serem exatas, as medições de qualidade de vídeo devem ser baseadas na qualidade percebida do vídeo atual recebido pelos usuários em vez da qualidade de vídeo medida tradicionalmente por sinais de teste (por exemplo, barras coloridas). Isso ocorre porque o desempenho dos sistemas de vídeo digital é variável e depende das características dinâmicas tanto do vídeo original (por exemplo, detalhes espaciais, movimento) quanto do sistema de transmissão digital (por exemplo, taxa de bits, taxa de erro) [23].

Um dos problemas que surge com a introdução de sistemas de compressão de vídeo digital nos mais diversos serviços é o de avaliação de qualidade. A utilização de sistemas padronizados não garante homogeneidade de desempenho, dada a estrutura desses padrões. Por outro lado, metodologias de avaliação são importantes ferramentas para a configuração adequada desses sistemas numa fase de grande diversidade de aplicações, que podem envolver cascata de sistemas, conversões entre formatos e utilização de configurações escalonáveis.

A dificuldade em se desenvolverem técnicas para estimação de qualidade de sistemas de compressão de vídeo digital deve-se, em parte, ao fato de os algoritmos de compressão introduzirem degradações numa escala que depende fortemente do conteúdo de detalhes e movimento das cenas. A Tabela 2.1 apresenta as principais degradações que diminuem a qualidade de vídeo. Os métodos estatísticos de avaliação são inadequados para quantificar a qualidade desses sistemas.

A avaliação objetiva é utilizada para medir automaticamente a qualidade de vídeo em todo o trajeto da informação em diversas condições, obtendo parâmetros de imagens de testes a partir do conhecimento de resultados padronizados. A escolha das cenas e dos

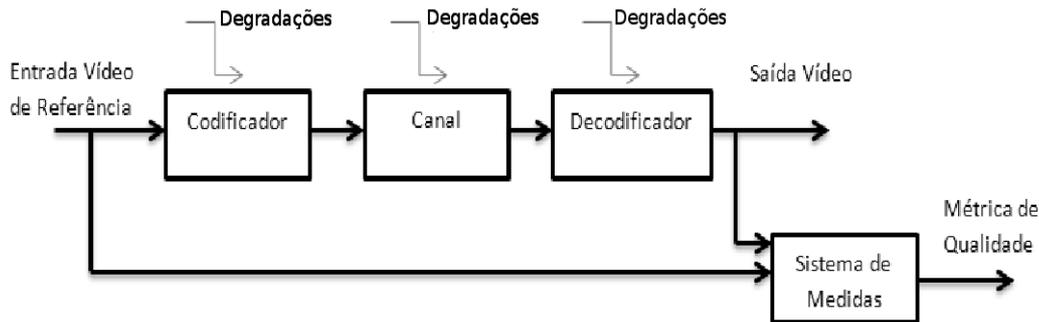


Figura 2.1: Modelo de Referência completa (FR) [19].

parâmetros analisados representa uma característica muito importante da metodologia, pois o nível de detalhes e formas de cada quadro de teste afetará parâmetros de maior ou menor qualidade [1].

2.2.1 Organismos de Padronização

Os atuais modelos de qualidade de vídeo que estão em desenvolvimento para as Recomendações ITU-T preveem sequências de vídeo originais e processadas. Do fluxo de dados transmitidos originalmente, o modelo objetivo pode extrair informações sobre os erros de transmissão (atraso, perda de pacotes) e do codificador (tipo, a taxa de bits, parâmetros do codificador)[31].

Usando informações adicionais do fluxo de vídeo, como os parâmetros do codificador, taxa de bits, quadros por segundo, informações de pacote, informação de erro da transmissão, os modelos objetivos podem se tornar mais rápidos e precisos.

As Recomendações do ITU-T e o Grupo de Especialistas em Qualidade de Vídeo VQEG (*Video Quality Experts Group*) [32][33][19] tratam de implementações de modelos de Referência Completa (FR – *full reference*), Referência Reduzida (RR – *reduced reference*), e Sem Referência (NR – *no reference*) e são ilustrados nas Figuras 2.1, 2.2 e 2.3

A Figura 2.1 apresenta o modelo de referência completa (FR – *full reference*), onde o sistema de medida objetiva calcula a métrica de qualidade na saída do sistema, utilizando uma cópia da entrada do sistema (o vídeo original) para o cálculo da métrica. Os sistemas FR não são utilizados para aplicações práticas em tempo real devido a

Tabela 2.1: Principais problemas que degradam a qualidade de vídeo [30]

Efeito	Definição
Blocagem (Efeito de Bloco)	Degradação da imagem caracterizada pelo aparecimento de padrões retangulares ou tipo tabuleiro não apresentados na imagem original.
<i>Blurring</i> (Borrado)	Degradação global sobre toda a imagem, caracterizada pela perda de resolução de bordas e detalhes.
<i>Edge Busyness</i>	Degradação concentrada perto das bordas dos objetos.
<i>Jerkiness</i>	Degradação do plano original, onde o movimento contínuo é percebido como uma variação brusca dos quadros (degradação temporal).
Perda de Pacotes	Degradação de bloco onde um ou mais blocos da imagem não possui semelhança com a cena corrente. É causado por erros introduzidos pelo canal de transmissão
Ruído Tipo Mosquito	Ruído de quantização caracterizado por artefatos se movendo ao redor das bordas (como um mosquito voando ao redor da cabeça e ombros de uma pessoa).
Degradação de resposta ao movimento	Degradação de movimento de vídeo de forma que imagem tenha uma perda de resolução espaço-temporal.
Persistência de objeto	Degradação caracterizada por um objeto que foi apagado e continua a aparecer no sinal de imagem recebido.
Reposta de corte de cena	Degradação associada ao corte da cena.
<i>Smearing</i>	Degradação localizada acima da sub-região da imagem recebida, caracterizada por redução da definição de bordas e detalhes espaciais.

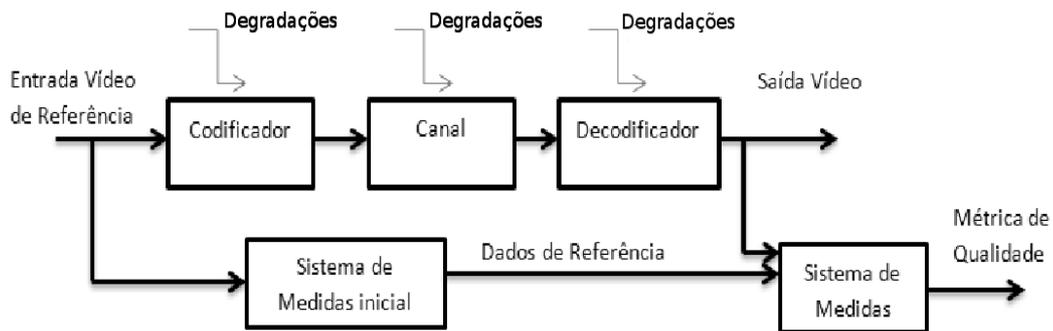


Figura 2.2: Modelo de Referência Reduzida (RR) [19].

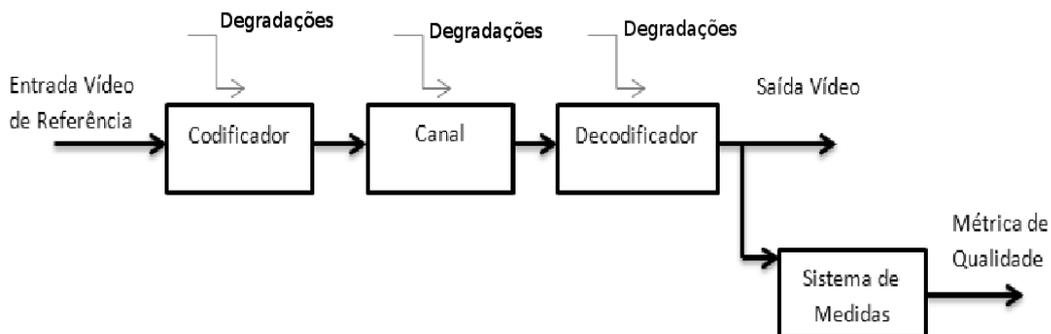


Figura 2.3: Modelo Sem Referência (NR) [19].

dificuldade da obtenção do vídeo original para a comparação. As utilizações principais dos sistemas FR são em laboratórios de testes, onde pode-se ter no ambiente de testes a cópia do vídeo original.

A Figura 2.2 apresenta o modelo de referência reduzida (RR – *reduced reference*), onde o sistema de medida objetiva calcula a métrica de qualidade na saída do sistema, utilizando dados de referência da entrada do sistema para o cálculo da métrica. Os dados de referência são obtidos através de um pré-processamento do vídeo de entrada, onde são extraídas informações para posterior utilização. Os sistemas RR podem ser utilizados para aplicações práticas em tempo real utilizando um canal de controle que carregará um conjunto de características relevantes do vídeo original.

Ao fornecer informação adicional sobre a qualidade do vídeo transmitido, é possível desenvolver métodos rápidos, eficazes e objetivos de medição de qualidade de vídeo. A Figura 2.2 mostra como amostras da qualidade de vídeo são computadas e transmitidas. As amostras de qualidade de vídeo podem ser transmitidas como metadados.

A Figura 2.3 apresenta o modelo sem referência (NR – *no-reference*), onde o sistema de medida objetiva calcula a métrica de qualidade na saída do sistema, sem utilizar informações do vídeo original para o cálculo da métrica. Os sistemas NR são os que mais se adaptam para aplicações práticas em tempo real, um vez que não existe a necessidade de envio de informações na saída do sistema. Esses sistemas são os mais complexos para a geração das métricas de qualidade pois o modelo de cálculo deve prever a maior quantidade possível de informações que deverão ser analisadas na saída do sistema.

2.2.2 Revisão Bibliográfica de Qualidade de Vídeo Objetiva

Nesta seção é apresentado um subconjunto das principais métricas objetivas de qualidade de vídeo estudadas. Essas métricas têm características semelhantes às da métrica apresentada nesta pesquisa. Um resumo dessas métricas é apresentado na Tabela 2.2 no final desta seção.

Os primeiros modelos da visão humana adotaram uma abordagem de canal único [34]. Modelos de canal único consideram o sistema visual humano como um único filtro espacial, cujas características são definidas pela função de sensibilidade ao contraste (CSF - *Contrast Sensitivity Function*). A saída de tal sistema é a versão filtrada do estímulo de entrada e detectabilidade depende de um critério de limiar [34].

O primeiro modelo computacional da visão foi projetado por Schade [35] para prever a sensibilidade padrão para a visão foveal. Ele baseia-se na suposição de que a representação é uma transformação cortical invariante ao deslocamento da imagem da retina e pode assim ser expresso como uma convolução.

A primeira métrica de qualidade de imagem utilizando a luminância foi desenvolvida por Mannos e Sakrison [36]. Eles perceberam que a simples medida de distorção com base em pixel não era capaz de prever com precisão as diferenças de qualidade percebidas pelos observadores. Com base em experiências psicofísicas inferiram algumas propriedades do sistema visual humano e desenvolveram uma expressão para a sensibilidade de contraste como uma função da frequência espacial, o que ainda é largamente utilizado em modelos HVS.

A primeira métrica de qualidade de vídeo foi desenvolvida por Lukas e Budrikis [37].

Baseia-se em um modelo espaço-temporal da função sensibilidade de contraste (CSF) utilizando uma excitação e um caminho inibitório. Os dois caminhos são combinados de uma maneira não linear, permitindo que o modelo possa se adaptar a alterações do nível de fundo de luminância.

Tong et al. [38] propuseram uma métrica de qualidade de vídeo de canal único chamada ST-CIELAB (CIELAB espaço-temporal). ST-CIELAB é uma extensão da métrica de qualidade de imagem CIELAB espacial (S-CIELAB) de Zhang e Wandell [39]. Ambas são compatíveis com o padrão CIELAB [38][39] para campos de cor uniforme.

Watson [40] introduziu a transformação córtex, uma pirâmide multi-resolução que simula a sintonia de frequência espacial e a orientação das células do córtex visual primário. A métrica é atraente por causa de sua flexibilidade. A seletividade de frequência espacial e de orientação são modeladas separadamente, as larguras de banda de filtro podem ser ajustadas dentro de uma ampla faixa, e a transformação é facilmente invertida.

Watson e Ahumada [41] mais tarde propuseram uma pirâmide ortogonalmente orientada operando em uma estrutura treliça hexagonal como uma ferramenta de decomposição alternativa.

Daly [42] propôs a métrica Preditor de Diferenças Visuais (VDP - *Visual Differences Predictor*). O modelo inclui uma não linearidade da amplitude para explicar a adaptação do sistema visual para diferentes níveis de luz, uma CSF bidimensional dependente da orientação, e uma hierarquia de mecanismos de detecção.

Lubin [43] desenvolveu o modelo Discriminação Visual Sarnoff (VDM - *Visual Discrimination Model*) para medir a fidelidade de imagem. Primeiro as imagens de entrada são convolvidas com uma aproximação da função de espalhamento de ponto (PSF) do olho. Em seguida, a amostragem do mosaico de cone na retina é simulada. O estágio de decomposição utiliza uma pirâmide Laplaciana para a separação espacial de frequência, cálculo de contraste local, e filtragem de direção, a partir do qual uma medida de energia de contraste é calculada.

O VDM é um dos poucos modelos que levam em conta a excentricidade das imagens no campo de visão do observador. Ele foi posteriormente modificado pela métrica Sarnoff JND para vídeos coloridos (Lubin e Fibush [44]).

A métrica de distorção para imagens estática apresentada por Teo e Heeger [45][46] baseia-se nas propriedades de resposta de neurônios no córtex visual primário e a psicofísica de detecção de padrões espaciais.

Van den Branden Lambrecht [47] propôs uma série de métricas de qualidade de vídeo (MPQM - *Moving Picture Quality Metric*). As MPQM baseiam-se numa definição de contraste local e filtros Gabor relacionados para a decomposição espacial, dois mecanismos temporais, bem como uma função de sensibilidade ao contraste espaço temporal e um modelo simples intracanal de mascaramento de contraste.

Masry e Hemami [48] projetaram uma métrica para avaliação da qualidade de vídeo contínuo (CVQE - *Continuous Video Quality Evaluation*) de vídeos com baixa taxa de bits. A métrica trabalha apenas com informação de luminância. Ela usa filtros temporais e uma transformada *wavelet* para a decomposição perceptual, seguida pela ponderação das CSF de diferentes bandas, um modelo de controle de ganho.

A métrica DCTune (*Discrete Cosine Transform Tuning*) aperfeiçoa o método de compressão da imagem JPEG, que é baseado na DCT, e foi desenvolvida por Watson [49][50]. DCTune calcula as matrizes de quantização JPEG que atingem a compressão máxima para uma distorção perceptual especificada dada uma determinada imagem e um conjunto particular de condições de visualização. Considera o mascaramento visual, iluminação e técnicas de contraste. DCTune também pode calcular a diferença de percepção entre duas imagens.

Watson [51] estendeu a métrica DCTune para vídeo. Em adição aos efeitos de sensibilidade e de mascaramento espaciais considerados em DCTune. A métrica qualidade de vídeo digital (DVQ - *Digital Video Quality*) leva em conta medições dos limiares de visibilidade para o ruído de quantização da DCT variando temporalmente.

Wolf e Pinson [52] desenvolveram a métrica qualidade de vídeo (VQM - *Video Quality Metrics*), que utiliza informações de referência reduzidas na forma de recursos de baixo nível extraídas de blocos espaço temporais das sequências. Estas características foram selecionadas empiricamente a partir de certo número de candidatos de modo a produzir a melhor correlação com os dados subjetivos.

A métrica VQM [52], fornece métodos padronizados (para aplicações de TV a cabo digital), bem como não padronizados (expandido) para medir a qualidade de vídeo

percebida de sistemas de vídeo digitais. As principais degradações consideradas no modelo geral incluem borramento, distorção de bloco, movimento irregular / anormal, ruído em canais de luminância e crominância, e erros de blocos (por exemplo, erros de transmissão). Uma combinação linear ponderada de todas as métricas de deficiências é usada para chegar à classificação VQM. O Modelo Geral VQM foi padronizado pela ANSI em julho de 2003 (ANSI T1.801.03-2003 [1]) e incluído como modelo normativo do ITU nas Recomendações ITU-T J.144 [32].

Tan et al. [53] apresentaram uma ferramenta de medição de qualidade de vídeo MPEG (*Motion Picture Experts Group*). Ela primeiro calcula a degradação perceptual em cada quadro com base na sensibilidade ao contraste e o mascaramento com a ajuda de filtragem espacial e operadores Sobel, respectivamente. Em seguida, o PSNR do sinal de erro das máscaras é calculado e normalizado. A segunda etapa da métrica é um emulador cognitivo, que simula aspectos de percepção de alto nível.

Wang et al. [54] propôs o índice de similaridade estrutural de Vídeo (VSSIM), que utiliza as distorções estruturais para estimar distorções perceptivas. A técnica VSSIM pretende explorar a forte dependência entre as amostras do sinal. As degradações são consideradas como sendo devida a perda de percepção informação estruturais do sistema visual humano. A base da métrica VSSIM é o índice proposto anteriormente *Structural Similarity* (SSIM) [55], que era específico para avaliação da qualidade de imagens. Para as sequências de vídeo, a métrica VSSIM mede a qualidade do vídeo distorcida em três níveis, ou seja, o nível local região, o nível de quadro, e ao nível da sequência.

Além do SSIM e do VSSIM, a métrica Multiscale-SSIM (MS-SSIM) [56] e a métrica Velocidade SSIM [57] foram propostas.

A métrica MS-SSIM é uma extensão da abordagem de escala única utilizada em SSIM e oferece mais flexibilidade, incorporando as variações da resolução e condições de visualização da imagem. Em cada etapa (ou escala), o método MS-SSIM aplica um filtro passa baixas para as imagens de referência e distorcida e reduz a resolução das imagens filtradas por um fator de dois. O índice MS-SSIM pode ser estendido para vídeo, o aplicando quadro a quadro no componente de luminância do vídeo e o índice MS-SSIM global para o vídeo é calculado como a média dos índices de qualidade de quadro.

A métrica Velocidade SSIM é proposta em [57] e utiliza o índice SSIM em conjunto com os modelos estatísticos de percepção da velocidade de visualização descritos em [58]. Utilização de modelos de percepção de velocidade visual com o índice SSIM foi mostrada em [55], para melhorar o desempenho, em comparação com PSNR e SSIM.

A métrica Fidelidade da Informação Visual (VIF - *Visual Information Fidelity*) [59] é baseada em estatísticas visuais combinadas com o modelo HVS. A métrica VIF modela imagens naturais como realizações de misturas de escala Gaussianas no domínio *wavelet*.

Semelhante ao VSSIM, Lu et al. [60] propuseram uma métrica de avaliação de qualidade de vídeo de referência completa com base em medições de distorção estrutural. A primeira fase avalia a MSSIM escolhendo aleatoriamente zonas localizadas e computando características estatísticas, tais como a média e da variância, para obter a qualidade local e a medida de qualidade de quadro (como em VSSIM).

Shnayderman et al. [61] desenvolveram uma medida de distorção chamada M-SVD (*Metrics Singular Value Decomposition*) para a avaliação da qualidade da imagem com base no conceito de decomposição em valores singulares. M-SVD é uma forma de fatorar matrizes numa série de aproximações lineares que expõem a estrutura subjacente da matriz. A métrica M-SVD mede a distorção como função da distância entre os valores singulares de bloco da imagem original e distorcida.

Uma vez que as medidas de distância são calculadas para todos os blocos, uma medida global é obtida calculando a média das diferenças entre a medida de distância para cada um dos blocos e a mediana de todas as medidas de distância em bloco. Este erro global é usado para derivar a medida M-SVD. Utilizando este conceito de medida de distorção, Tao e Eskicioglu [62] desenvolveram um modelo de qualidade de vídeo objetiva com referência completa.

Pessoa et al. [63] apresentaram um modelo de qualidade de vídeo que segmenta as imagens em regiões planas, de bordas e de texturas. A segmentação em regiões ajuda na captura do grau de distorção percebida. Por exemplo, blocagem é mais perceptível nas regiões planas (*flat*) e manchas são mais evidentes nas regiões de borda e texturizações.

Okamoto et al. [64] propuseram uma métrica de qualidade de vídeo que considera distorções visuais, incluindo borramento das bordas, geração de novas bordas e deterioração na direção temporal. Usando a métrica de Diferenças de Energia de Borda

Média apresentada na norma ANSI T1.801.03 [1], Okamoto et al. investigaram a precisão da previsão da qualidade dessa métrica em relação à previsão de deteriorações em regiões de bordas.

Lee e Sim [65] medem características visuais nas bordas e nas regiões de fronteira do bloco. A métrica KVQM proposta apresenta valores que indicam a degradação visual da imagem, ou seja, efeitos de bordas, blocagem e borramento. O valor final da métrica de qualidade é obtido pela combinação linear ponderada das três métricas individuais.

Bhat et al. [66] apresentaram um método que explora a correlação entre os resultados objetivos e subjetivos. A métrica determina a correlação entre o MOS previsto e o erro quadrático médio (MSE), utilizando o modelo de correlação linear da Equação 2.1:

$$MOS_p = 1 - k(MSE) \quad (2.1)$$

A métrica MPQM de van den Branden Lambrecht e Verscheure [67] simula o modelo espaço temporal do sistema visual humano, com uma abordagem de banco de filtros. A decomposição perceptual dos filtros representa os principais aspectos da sensibilidade ao contraste e mascaramento.

Xiao [68] propôs uma modificação do DVQ de Watson [69] que fez uso do fato de que a sensibilidade do olho humano com os padrões espaço temporal diminui com frequências espaciais e temporais elevados. O método é semelhante ao modelo de Watson, exceto que o contraste local, obtido com os componentes DC é ainda convertido em diferenças visíveis (JND - *Just Noticeable Difference*) utilizando uma matriz de sensibilidade de contraste espacial (SCS) para quadros estáticos e uma matriz que representa a propriedade temporal, para os quadros dinâmicos. Esta métrica de qualidade de vídeo baseada na DCT também chamada VQM (não deve ser confundida com NTIA VQM [52]) foi definida em termos de uma máxima distorção e de uma distorção média ponderada.

Lee e Kwon [70] propuseram uma métrica de qualidade de vídeo objetiva baseada na transformada wavelet. O modelo utiliza uma transformada *wavelet* multinível para calcular as frequências espaciais com base nas subbandas resultantes.

A métrica de qualidade de vídeo de referência completa (MOVIE - *MOViment based*

in Video Integrity Evaluation) Movimento Baseado em Avaliação da Integridade de Vídeo foi proposta por Seshadrinathan e Bovik [71]. O modelo MOVIE captura as características temporais da área visual do meio (MT) do córtex do cérebro humano para análise da qualidade de vídeo. Estudos da neurociência indicam que a área visual MT é fundamental para a percepção de qualidade de vídeo [72].

A sensibilidade do HVS nas bordas e mudanças de locais de luminância é explorado por Hekstra et al. [73] propondo a métrica de qualidade de vídeo objetivo Perceptual Quality Video Metric (PVQM, também conhecido como Swisscom / KPN métrica). O modelo utiliza uma combinação linear de três indicadores de distorção, ou seja, bordas, de correlação temporal e erro de cor para medir a qualidade perceptual.

Lu et al. [74] propuseram uma métrica de referência reduzida e sem referência ponderada por saliências para medir distorções visuais com base na atenção visual, fixação e movimento dos olhos e o caminho da visão e retina, que são considerados os três aspectos da percepção. A métrica estima o mapa perceptual de significado de qualidade (PSQM) para modelar a atenção visual e fixação e movimento dos olhos, enquanto métricas de qualidade visuais existentes são adotadas para simular a retina.

Na métrica de qualidade de vídeo proposta por Ong et al. [75][76], a qualidade perceptual é medida como uma função da distorção, invisibilidade, blocagem e fator de fidelidade dos conteúdos. O limite de visibilidade dá uma medida da quantidade máxima de distorção que um pixel particular pode sofrer e ainda ser imperceptível pela visão humana.

Com base nos trabalhos anteriores de Ong et al. [75][76], Nya et al. [77] propuseram uma métrica de qualidade de vídeo de referência completa melhorada. Uma das modificações inclui o uso de um filtro de Sobel para aproximar o gradiente de luminância local em comparação com equações complexas utilizadas em [75][76].

A métrica VSNR de Chandler e Hemami [78] é uma métrica de referência completa para imagens estáticas, mas também mostrou um desempenho promissor na avaliação da qualidade de vídeo quando aplicada sobre quadro a quadro e, em seguida feito a média. A métrica visa minimizar o problema supra limiar na modelagem HVS. O modelo utiliza o mascaramento visual e conceitos de soma visuais para identificar as distorções percentualmente detectáveis.

Na métrica objetiva de qualidade de vídeo de referência completa avaliação perceptiva da qualidade de vídeo (PEVQ)[79] com base no modelo PVQM discutido anteriormente, a avaliação da qualidade é composta por cinco etapas principais. A primeira fase de pré-processa os dois sinais de vídeo original e distorcido por extração da região de interesse (ROI). O ROI é obtido cortando o quadro atual, com um tamanho de corte definido pelo formato do vídeo. Estes quadros ROI derivados são utilizados em fases posteriores. A segunda fase alinha os sinais de vídeo pré-processados espacialmente e temporalmente. Fases três e quatro calculam quatro medidas de distorção, nomeadamente, bordas em luminância e em crominância e dois indicadores variabilidade temporal, bem como uma medida de distorção temporal. A fase cinco utiliza uma abordagem sigmoide para mapear as distorções à medida de qualidade de vídeo DMOS.

O VQEG desenvolveu testes de validação de métricas objetivas de qualidade aplicáveis ao vídeo HD [80].

Wolf e Pinson [81] fizeram um estudo do desempenho do modelo geral NTIA para sequências de vídeo HDTV, e mediram o grau de precisão através da sua comparação com os resultados do método de avaliação da qualidade subjectiva SSCQE.

Sugimoto et al. [82] propuseram uma métrica para avaliar a qualidade de vídeo percebida de vídeos em HD, considerando distorções, tais como blocagem, a variância de MSE, a degradação temporal do PSNR, potência média de diferença inter quadro na sequência, média do MSE dos blocos com alta variância, a degradação dos componentes de baixa e de alta frequência.

Com base em seu trabalho anterior para PC e serviços celulares [64][79], Okamoto et al. [83] propôs uma métrica de qualidade de vídeo perceptual referência completa para HDTV usando medidas *fuzzy*.

A empresa SwissQual desenvolveu um método de referência completa full-HD chamado VQuad-HD [84]. VQuad-HD consiste de quatro componentes principais, análise da distribuição das semelhanças e diferenças dos pixels locais, análise blocagem, análise de movimentos rápidos, e agregação das características de similaridade, diferença, blocagem e *jerkiness*.

A métrica da Tektronix [80] incorpora componentes adaptativos para alinhamento espacial, a percepção visual humana e a modelagem de cognição [85].

A métrica da Universidade Yonsei da Coréia se baseia na detecção de borda, seguida pela extração de características das áreas de bordas. A degradação de áreas de borda é medida como borda PSNR, que é refinada por recursos adicionais [80].

Na Tabela 2.2 é apresentado um resumo das principais métricas propostas na literatura classificados de acordo com as características de cada métrica e uma comparação de suas desvantagens em relação a métrica de qualidade proposta nesta pesquisa.

2.2.3 Testes Subjetivos

Esses experimentos são valiosos instrumentos de pesquisa para uma melhor compreensão de como os seres humanos julgam a qualidade do vídeo e percebem algumas de suas deficiências. As respostas das avaliações subjetivas obtidas a partir desses experimentos são consideradas importantes referências para o desenvolvimento de métricas de qualidade de vídeo.

Normalmente, as experiências psicofísicas são caras e demoradas [21][87]. A execução de um experimento requer a disponibilidade de materiais, equipamentos e espaço físico. A concepção, execução e análise de dados consomem uma grande quantidade de tempo dos experimentadores. Dessa forma, o número de experimentos que podem ser realizadas é limitado e uma metodologia adequada precisa ser desenvolvida para tirar o máximo proveito dos recursos e garantir que os objetivos pretendidos sejam alcançados [21][87].

2.2.3.1 Geração das Sequências de Teste

Para início do projeto do experimento psicofísico, um conjunto de sequências de vídeos sem degradações é escolhido entre um conjunto de vídeos comumente utilizados em pesquisas de qualidade de vídeo que estão disponíveis para a comunidade [88][28][89][90]. Esses vídeos serão considerados como "vídeos originais".

A abordagem para os testes subjetivos de qualidade de vídeo consiste na utilização de uma quantidade variável de degradações de vídeo e a posterior avaliação dos observadores humanos para que classifiquem a qualidade de cada vídeo apresentado [21].

Tabela 2.2: Métricas de Qualidade de Vídeo

Métrica	Classificação	Categorização
SSIM [55]	Estatísticas Visuais	<i>Full Reference</i>
VSSIM [54]	Estatísticas Visuais	<i>Full Reference</i>
MS-SSIM [56]	Estatísticas Visuais	<i>Full Reference</i>
VIF [59]	Estatísticas Visuais	<i>Full Reference</i>
Lu et al. [60]	Estatísticas Visuais	<i>Full Reference</i>
Tao & Eskicioglu [62]	Estatísticas Visuais	<i>Full Reference</i>
IES2 [63]	Características Visuais	<i>Full Reference</i>
VQM [52]	Características Visuais	<i>Full Reference</i>
Okamoto [64]	Características Visuais	<i>Full Reference</i>
Lee & Sim [65]	Características Visuais	<i>Full Reference</i>
Bhat et al. [66]	Características Visuais	<i>Full Reference</i>
Lukas & Budrikis [37]	Perceptual Dom. Frequência	<i>Full Reference</i>
MPQM [67]	Perceptual Dom. Frequência	<i>Full Reference</i>
DVQ [69]	Perceptual Dom. Frequência	<i>Full Reference</i>
Xiao et al. [68]	Perceptual Dom. Frequência	<i>Full Reference</i>
Lee & Kwon [70]	Perceptual Dom. Frequência	<i>Full Reference</i>
MOVIE [71]	Perceptual Dom. Frequência	<i>Full Reference</i>
PVQM [73]	Perceptual Dom. Pixel	<i>Full Reference</i>
PQSM [74]	Perceptual Dom. Pixel	<i>Full Reference</i>
Ong et al. [75][76]	Perceptual Dom. Pixel	<i>Full Reference</i>
Nya et al. [77]	Perceptual Dom. Pixel	<i>Full Reference</i>
VSNR [78]	Perceptual Dom. Pixel	<i>Full Reference</i>
PEVQ [79]	Perceptual Dom. Pixel	<i>Full Reference</i>
Psytechnics [81]	Perceptual Dom. Pixel	<i>Full Reference</i>
Farias [86]	Perceptual Caract. Visuais	<i>No Reference</i>
Wolf et al. [81]	HD Características Visuais	consumo de recursos (FR)
Sugimoto et al. [82]	HD Características Visuais	consumo de recursos (FR)
Okamoto et al. [83]	HD Características Visuais	consumo de recursos (FR)
VQuad-HD [84]	HD Características Visuais	consumo de recursos (FR)

Na maioria dos experimentos, as degradações são introduzidas utilizando diversos tipos de codificadores de vídeo, diversas taxas de bits e diversas taxas de perda de pacotes [28]. Em alguns experimentos, novas degradações são introduzidas por meio da combinação linear de outras degradações introduzidas [87].

A duração da experiência é limitada a 30 minutos para reduzir os efeitos da fadiga sobre os avaliadores humanos [21]. Um número aproximado de 100 sequências de vídeo pode ser mostrado durante uma sessão de testes de 30 minutos. Utilizando as sequências de vídeos originais, diversos tipos de degradações são introduzidas nos originais para gerar a totalidade das sequências que serão utilizadas durante os testes.

Cada vídeo original gera várias sequências com degradações, e o número total de originais utilizados para cada experiência é então calculado de forma a manter a duração dos testes dentro dos 30 minutos.

2.2.3.2 Avaliação Subjetiva

Os métodos de avaliação subjetiva consistem na realização de experimentos subjetivos (psicofísicos) com participantes (observadores) não especialistas, com o objetivo de obter estimativas para a qualidade de vídeo de um conjunto de sequências.

Nos testes propostos são utilizados os procedimentos normatizados para avaliação subjetiva estabelecidos pela ITU. Os principais documentos da ITU que contêm recomendações de procedimentos para aferição subjetiva da qualidade de vídeo são:

- ITU-R Rec. BT.500-11 [21] destinado a aplicações na área de televisão;
- ITU-T Rec. P.910 [19] destinado a aplicações em multimídia.

Esses documentos fornecem especificações para as condições de visualização dos vídeos, critérios para seleção dos observadores e do material de teste, metodologia empregada nos experimentos e métodos para análise estatística dos dados coletados. A condução efetiva dos testes subjetivos será iniciada após a geração das sequências codificadas.

As condições de visualização, que incluem a distância do observador para o monitor e o controle da luz ambiente, será feita de acordo com a recomendação do ITU-R [21].

O tamanho e tipo dos monitores a serem utilizados nos experimentos são apropriados para a aplicação em questão.

Um mínimo de 15 observadores deve ser utilizado em cada experimento de forma a garantir resultados confiáveis [21]. Os observadores serão escolhidos entre não especialistas na área de processamento de imagens ou vídeo. Instruções orais e escritas do teste devem ser fornecidas aos observadores. Estas instruções incluem uma descrição da aplicação em mente, o tipo de método de aferição utilizado, a escala utilizada e a metodologia escolhida [21].

De acordo com o ITU [21], há duas classes de métodos subjetivos. Na primeira classe, os observadores fornecem julgamentos em uma escala de qualidade (*Quality Scale*). Estes julgamentos ajudam a estabelecer o desempenho do sistema em boas condições de funcionamento. Na segunda classe de experimentos, os observadores fornecem o julgamento em uma escala de degradação (*Impairment Scale*), onde são aferidas a visibilidade e a severidade dos artefatos presentes no vídeo. Esse tipo de experimento avalia o desempenho de um sistema sob condições não-ótimas.

A metodologia utilizada pode ser classificada de acordo com a frequência com que os julgamentos (notas de qualidade) são coletados. Uma avaliação contínua da qualidade é a metodologia que mais se aproxima das condições reais nas quais os usuários assistem aos vídeos [21]. Neste tipo de metodologia, o observador assiste a trechos de vídeos de média ou longa duração e fornece julgamentos de maneira contínua (sem intervalos), utilizando uma régua deslizante. Neste tipo de avaliação, sequências de até 5 minutos são agregadas de maneira aleatória e sem intervalos, gerando sessões entre 30 e 60 minutos. Os julgamentos coletados neste tipo de experimento correspondem a curvas contínuas que mostram a evolução da qualidade com o tempo. Esta metodologia apresenta a desvantagem de exigir o investimento de uma grande quantidade de tempo, o que limita o número de condições que podem ser testadas. Uma opção mais popular é utilizar sequências de teste de 8 a 10 segundos e coletar os julgamentos de forma discreta, ou seja, ao final da apresentação de cada sequência de teste.

Dependendo da forma de apresentação das sequências de teste (estímulo), a metodologia do experimento pode ser classificada como sendo de estímulo simples ou duplo. No caso de uma metodologia de estímulo simples, a sequência de teste é apresentada isoladamente, enquanto que no caso de uma metodologia de estímulo duplo um par de sequências de teste (uma sequência de teste e o original correspondente) são apre-

sentados em conjunto. A utilização de uma metodologia de estímulo duplo permite uma maior discriminação entre as diversas condições de teste, o que é particularmente interessante quando existem condições de teste com diferenças de qualidade pequenas.

Neste trabalho, foi necessário avaliar uma grande quantidade de condições de teste. Como valores sutis de qualidade precisaram ser discriminados, foi utilizada uma metodologia de estímulo duplo com escala de níveis de qualidade. Sendo assim, optou-se pelo método DSCQS (*Double Stimulus Continuous Quality Scale*) [21]. Além de satisfazer a todos os requisitos especificados na pesquisa, o DSCQS é um método bastante robusto que tem sido utilizado com sucesso nos experimentos conduzidos pelo VQEG [28].

2.2.3.3 Preparação do Ambiente (Condições Físicas)

Com relação a disposição dos observadores em relação ao monitor, a configuração apresentada na Figura 2.4 é utilizada. Devido ao tamanho do monitor, os experimentos foram realizados com um observador de cada vez. O observador foi posicionado em frente ao monitor, a uma distância de 3 vezes a altura do monitor (H). A altura dos olhos do observador deve estar aproximadamente na altura do centro da tela. Não deve haver luz direta sobre a tela do monitor. As luzes das salas devem ser ajustadas de acordo com as recomendações [28].

2.2.3.4 Metodologia do Experimento

A escolha do método experimental mais adequado deve levar em consideração a aplicação e os objetivos do experimento. Faz-se necessário utilizar um método experimental robusto que permita distinguir entre pequenas diferenças de qualidade entre os vídeos, uma vez que para alguns parâmetros as diferenças de qualidade podem ser bem sutis.

Antes das sessões experimentais, teste piloto são realizados utilizando um conjunto pequeno de observadores (especialistas ou não), para avaliar a estabilidade dos parâmetros do experimento como um todo. Uma sessão de treinamento dos observadores deve ser incluída no começo da sessão experimental.

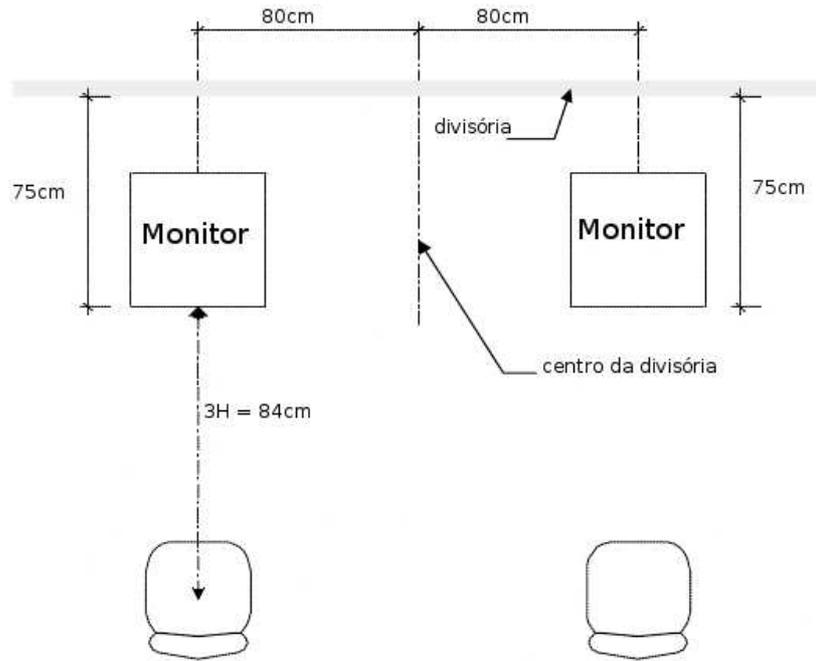


Figura 2.4: Posição do Observador [28].

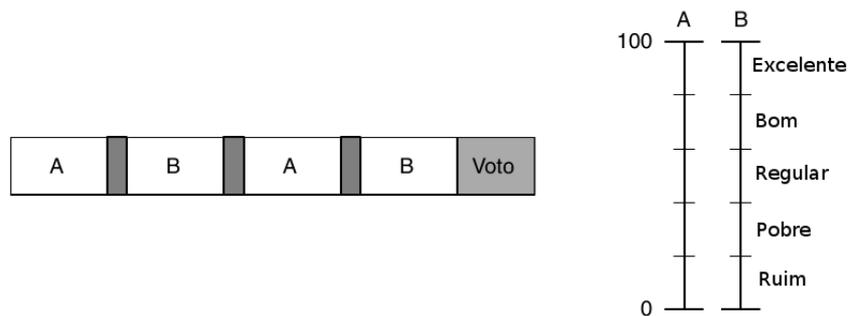


Figura 2.5: Método DSCQS [34].

As sequências de teste são apresentadas em uma ordem (pseudo)aleatória e, preferivelmente, deve-se evitar que sequências de teste referentes ao mesmo original sejam apresentados uma após a outra [21]. No método DSCQS, os observadores assistem a pares de sequências que consistem da sequência de teste e da original correspondente. As sequências são apresentadas duas vezes, alternadamente (original e teste), conforme apresentado na Figura 2.5.

No DSCQS, os observadores não são informados sobre quais as sequências são os originais e quais são os testes. A cada sessão, as posições das sequências originais e das sequências de teste são modificadas aleatoriamente. As sequências são apresentadas em toda a tela do monitor. O tom de cinza a ser mostrado entre sequências apresenta

o valor médio da luminância de 128, em uma escala de tons de cinza variando entre 0 a 255. A varredura das linhas do vídeo deve estar em sincronismo com a varredura do monitor [21].

No período marcado como “voto” na Figura 2.5, o observador pode registrar o seu julgamento, ou seja a sua nota para a qualidade geral de ambas as sequências de vídeo. A escala de qualidade utilizada pelo DSCQS também é apresentada na Figura 2.5 [21]. Observe que a escala é contínua e que notas individuais devem ser atribuídas a cada uma das sequências dos pares.

As notas são registradas no próprio sistema de apresentação dos testes. O usuário atribui uma nota contínua (variando de ruim a excelente) a cada uma das duas sequências, o que é feito inserindo uma marca em cada uma das escalas (A e B). O sistema aguardará que as notas sejam dadas pelo usuário antes de passar à sequência seguinte.

2.2.3.5 Análise dos Resultados

A análise estatística dos dados coletados deve ser executada usando métodos padronizados para análise, conforme especificado em [21] e nos relatórios finais dos experimentos do VQEG Fases I e II [91][92]. Mais especificamente, para cada combinação das variáveis de teste devem ser calculados o valor médio e o desvio padrão das notas de qualidade coletadas. Testes de confiabilidade dos observadores também devem ser estimados.

Os métodos utilizados para analisar os dados resultantes dos avaliadores em um teste subjetivo envolvem análises estatísticas que permitem a descrição precisa de médias, diferenças, relações entre variáveis, ajustes entre funções [93].

Uma das formas de analisar as respostas dos avaliadores subjetivos nas tarefas de detecção é através das médias dos resultados de todos os observadores (MOS - *Mean Opinion Score* da Equação 2.2) em cada sequência de teste.

$$MOS = \frac{1}{L} \sum_{i=1}^L S_i \quad (2.2)$$

onde S_i é o resultado da resposta do i -ésimo avaliador e L é o número total de avaliadores.

Com os dados subjetivos processados, podemos obter maiores informações sobre a características das degradações encontradas nas sequências de vídeo processadas.

A regressão dos valores DMOS (*Differential MOS*), diferença entre os valores MOS em duas medidas (com o vídeo original e o vídeo sob testes) com os resultados do modelo objetivo pode não representar adequadamente o grau de concordância entre estes resultados para as várias sequências de vídeo. Com os dados DMOS relacionados as degradações nos vídeos processados, os resultados da métrica objetiva podem ser ajustados a uma função logística padrão [86][91], conforme Equação 2.3:

$$Y_i^W = w_i \left[\frac{\beta_1 - \beta_2}{1 + e^{-\left(\frac{X_i - \beta_3}{|\beta_4|}\right)}} + \beta_2 \right] + \varepsilon_i^w, i = 1 \dots n \quad (2.3)$$

onde as estimativas iniciais para os parâmetros a serem ajustados são:

$$\beta_1 = \max(Y_i)$$

$$\beta_2 = \min(Y_i)$$

$$\beta_3 = \text{med}(X)$$

$$\beta_4 = 1$$

e os demais parâmetros:

$$w_i = 1/(sY_i)$$

Y_i = i -ésimo valor DMOS

X_i = i -ésimo valor da métrica objetiva

σY_i = desvio padrão do i -ésimo valor DMOS

$$Y_i^w = w_i Y_i$$

$$\varepsilon_i^w = w_i \varepsilon_i$$

ε_i = i-ésimo valor residual

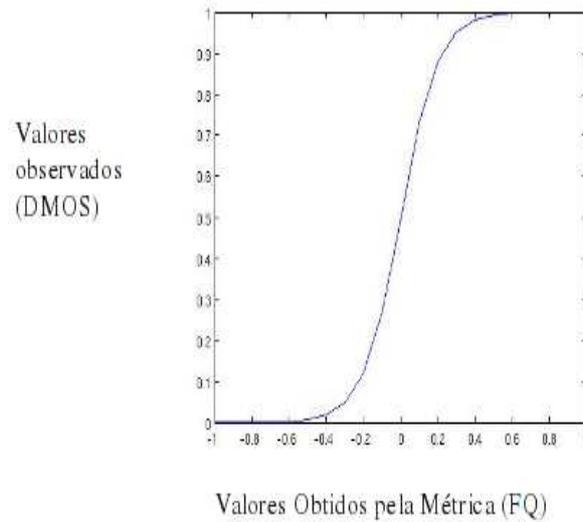


Figura 2.6: Gráfico característico da função logística

A função logística apresenta um comportamento conforme a Figura 2.6, onde os observadores conseguem distinguir uma melhora ou piora da qualidade observada de forma linear até determinados limites onde uma qualidade muito boa ou muito ruim levam aos limites de saturação na curva. Os parâmetros da função logística alteram a inclinação da curva e o deslocamento nos eixos.

Para que possamos verificar a qualidade dos vídeos no ambiente dos usuários, será necessário inserir marcas nos vídeos originais, logo após a produção destes conteúdos. Os vídeos serão marcados utilizando técnicas de marcas d'água digitais conforme descrito na próxima seção.

2.3 Marcas d'água Digitais

A ocultação de informações em uma mensagem é utilizada há bastante tempo. Vários trabalhos na área de esteganografia (do grego *steganós* = encobrir, *grapto* = escrita, ou ocultação de informações) e marcas d'água podem ser citados [94][95][96].

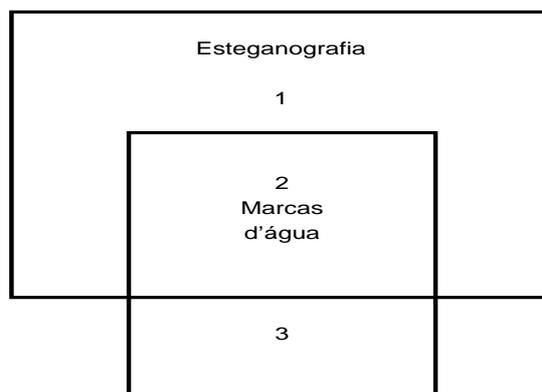


Figura 2.7: Esteganografia e marcas d'água [96].

As marcas d'água, porém, têm por objetivo assegurar algum tipo de validação, de autenticidade a determinado documento, podendo ser ou não ocultas, diferindo da esteganografia também por possuírem a característica de robustez contra ataques [97][98][99].

A Figura 2.7 mostra a relação entre a esteganografia e as marcas d'água. Algumas marcas não podem ser consideradas como esteganografia (tipo 3 na Figura 2.7). Certas técnicas de marca d'água não são totalmente ocultas, como as marcas das cédulas de dinheiro ou logotipos inseridos pelas emissoras de televisão. A Figura 2.8 mostra um exemplo de como as marcas d'água podem ser inseridas.

As melhores técnicas de inserção de marcas d'água do tipo invisível, em vídeos ou imagens, alteram a informação do vídeo ou imagem sem causar mudanças perceptíveis a olho nu.

Com a expansão dos meios digitais de armazenamento e transmissão de dados, o número de trabalhos na área de marcas d'água digitais aumentou consideravelmente. Atualmente, grande parte das pesquisas em marcas d'água são feitas na área de proteção de direito autoral [101][102][103].

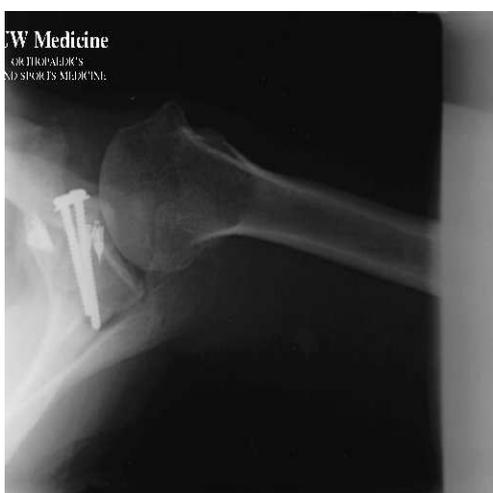
Em um sistema genérico de uso de marcas d'água, inicialmente é feita a inserção de uma marca em uma imagem, depois é feita a transmissão entre a origem e o destino, podendo ser tanto uma transmissão física quanto digital, com número indeterminado de intermediários. Após chegar ao destino, a imagem marcada passa por um detector ou extrator de marcas d'água. Se a resposta desse detector confirmar ou não a presença de uma marca d'água válida, o destinatário pode utilizar ou tomar as medidas necessárias para o tratamento da imagem.



(a) Imagem Original



(b) Marca d'água a ser inserida



(c) Imagem com marca d'água visível



(d) Imagem com marca d'água invisível

Figura 2.8: Tipos de Marcas d'água [100].

As principais propriedades [104][105] das marcas d'água para utilização em sistemas práticos, dependendo de suas finalidades, devem ser:

- invisibilidade: a marca d'água deve ser visualmente imperceptível e produzir a menor distorção possível na imagem original;
- detecção Confiável: um detector deve localizar uma marca d'água em uma imagem com o máximo de precisão possível;
- eficiência Computacional: os processos de inserção e detecção de marcas d'água devem ser computacionalmente viáveis, sendo necessária a detecção especialmente rápida se a aplicação envolver monitoração em tempo real;

- **robustez:** em aplicações que necessitam da marca d'água íntegra, como identificação de propriedade ou controle de cópias, a marca d'água deve resistir a ataques e modificações na imagem, ou então, em caso de tentativa de remoção, alterar a imagem, através da perda significativa de qualidade. Para aplicações de autenticação, a marca d'água deve ser alterada ou destruída quando ocorrer qualquer alteração na imagem;
- **capacidade de armazenamento:** propriedade que considera o total de informações que a marca d'água pode armazenar para posterior utilização.

Existem várias classificações que podem ser utilizadas para agrupar e diferenciar as técnicas existentes para criação de marcas d'água. Podemos classificar as marcas d'água em frágeis, semi-frágeis e robustas, de acordo com a capacidade de resistir às alterações em características da imagem [101][102].

Uma marca d'água robusta pode resistir a alterações de brilho, rotação, contraste, compressão com perdas em uma imagem. Uma marca classificada como robusta apenas poderia ser removida caso a qualidade da imagem sofresse grande alteração, comprometendo qualquer uso da imagem sem a marca. Este tipo de marca é utilizado normalmente para verificação de propriedade de imagens.

As marcas d'água frágeis são sensíveis às alterações na imagem, sendo facilmente removidas ou corrompidas e podem ser usadas para detectar edições nas imagens. Existem as marcas semi-frágeis, cujos métodos de detecção permitem alterações na imagem que não comprometam seu conteúdo, como compressão, melhoria de contraste dentre outras alterações.

As marcas d'água podem ser classificadas de acordo com modelos cujo detector da marca é baseado em algum tipo de correlação [106]:

- **correlação linear,** que utilizam funções lineares para detecção da marca e comparam seus resultados com determinados intervalos de valores para verificar a existência da marca;
- **correlação normalizada,** nos quais a marca de referência utilizada e a marca extraída da imagem são normalizadas para uma mesma unidade, antes de terem seus valores comparados com os intervalos definidos.

As técnicas de ocultação de informações podem ser divididas em categorias, baseadas no domínio utilizado pela aplicação [103]:

- baseadas no domínio dos dados, tanto espacial quanto temporal;
- baseadas no domínio de transformadas, normalmente Transformada Discreta do Cosseno (DCT), Transformada Discreta de Fourier (DFT) ou Transformada Discreta Wavelet (DWT);
- híbridas com características tanto dos domínios espacial/temporal quanto de transformadas;
- baseadas no domínio da compressão que, apesar de utilizarem técnicas dos domínios temporal, espacial ou de transformadas, possuem características específicas por manipularem diretamente sequências de bits comprimidas.

As marcas d'água, além dos domínios de transformadas e espacial, também podem ser divididas em outras técnicas:

- técnicas que utilizam Quantização Vetorial [107];
- técnicas com algoritmos genéticos, subdivididas nos domínios espacial e de transformadas [107][108][109];
- sistemas de marcas d'água utilizando redes neurais [107];
- sistemas baseados em Criptografia Visual [107].

Para a medição da qualidade de vídeo objetiva nas mesmas condições de visualização dos usuários, utilizam-se os parâmetros de resolução espacial do sistema óptico do ambiente do usuário. Esse ambiente é composto de diversos dispositivos que apresentam características ópticas distintas. Entre esses dispositivos, têm-se o monitor que apresenta o vídeo, as condições do ambiente de visualização, a resolução do monitor, a distância entre o monitor e o espectador e os parâmetros ópticos da câmera, que fará a coleta de informações. As informações das características ópticas do ambiente serão analisadas através da Função de Transferência de Modulação do sistema como um todo, conforme descrição na seção 2.4. No escopo desta pesquisa, a iluminação do ambiente não foi considerada para a geração da função de qualidade.

2.3.1 Marcas d'água Invariantes à Rotação, Escala e Translação

Um aspecto importante para escolha do tipo de marca a ser inserida nos vídeos está relacionado com o tipo de sistema utilizado para a verificação da qualidade de vídeo. O objetivo da metodologia é que a medida da qualidade de vídeo seja feita independentemente do tipo da rede, do tipo do vídeo e do tipo de dispositivo terminal do usuário, permitindo o monitoramento da qualidade em pontos próximos aos usuários, com dispositivos simples, de baixos custo e complexidade. Esses dispositivos devem interferir minimamente com os equipamentos já em utilização pelo distribuidor de vídeo. Para o atendimento destes objetivos, é utilizada uma câmera de vídeo desacoplada do sistema de recepção no ambiente do usuário, ou seja, a câmera não está fisicamente conectada ao receptor do usuário.

Pelo fato de estar desacoplada do sistema de recepção do usuário, não há sinais de sincronização chegando na câmera, o que é um desafio para a decodificação da marca d'água. A marca escolhida deve ser resistente às deformações geométricas e deve permitir a sua recuperação sem sincronismo com a marca inserida. Utilizamos uma Marca d'água do tipo “*RST Invariant*” (*rotation, scale, translation invariant*) nesta metodologia [110][111][112][113]. O diagrama da Figura 2.9 apresenta a sequência de atividades para a inserção da marca.

2.3.1.1 Normalização dos Quadros

A invariância à escala e à translação pode ser conseguida através da utilização da técnica de normalização de imagens [114]. Uma função de imagem $f(x, y)$ pode ser normalizada em relação à escala e à translação a transformando em uma nova imagem $h(x, y)$ da Equação 2.4:

$$h(x, y) = h\left(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y}\right) \quad (2.4)$$

onde (\bar{x}, \bar{y}) é o centróide de $f(x, y)$ e $a = \sqrt{\beta/m_{00}}$, β é um valor predeterminado e m_{00} é o momento de ordem zero.

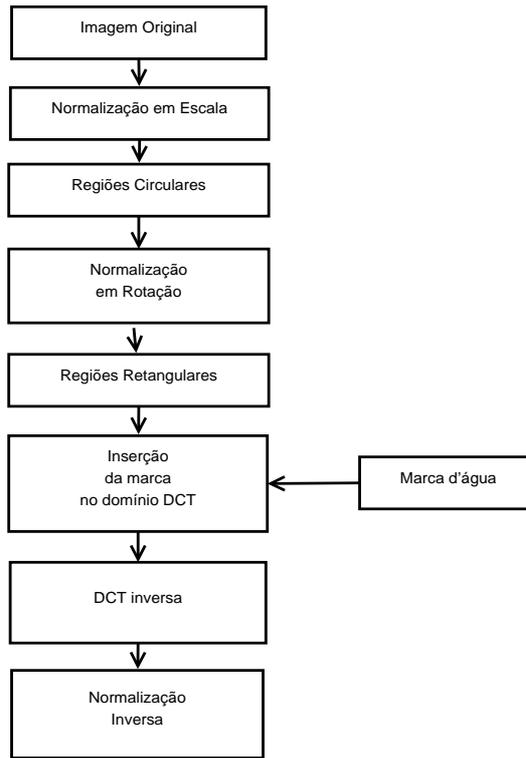


Figura 2.9: Algoritmo de inserção da marca d'água

Para calcular a nova imagem, primeiro movemos a origem da imagem para o centróide e transformamos a escala em um tamanho padrão.

A imagem normalizada é obtida a partir de um processo de transformação geométrica que é invariável às distorções geométricas da imagem.

Seja $f(x, y)$ a imagem digital de tamanho $M \times N$ ($1 \leq x \leq M, 1 \leq y \leq N$). Seu momento geométrico m_{pq} é definido na Equação 2.5 como:

$$m_{pq} = \sum_{x=1}^M \sum_{y=1}^N x^p y^q f(x, y) \quad (2.5)$$

Um processo de normalização invariante à escala e à translação é descrito nos seguintes passos para uma imagem $f(x, y)$. A invariância à translação é feita transformando a imagem original $f(x, y)$ em outra $f'(x, y) = f(x + \bar{x}, y + \bar{y})$, onde $\bar{x} = \frac{m_{10}}{m_{00}}$, $\bar{y} = \frac{m_{01}}{m_{00}}$. Então a origem é movida para o centróide.

A Invariância à escala é realizada aumentando ou reduzindo cada forma de tal modo que o seu momento de ordem 0 m_{00} seja definido por um valor pré-determinado β .

Dessa forma, a invariância à escala é conseguida através da transformação da função da imagem original $f'(x, y)$ em um nova função $h(x, y) = f'(\frac{x}{a}, \frac{y}{a})$, com $a = \sqrt{\beta/m_{00}}$. Então $h(x, y)$ é a imagem normalizada que é robusta às distorções de escala e translação.

No processamento das sequências de vídeo utilizadas, foi realizada a normalização espacial dos quadros do vídeo para um valor de referência. A normalização para uma resolução do quadro intermediária permite a detecção da marca quando são utilizadas resoluções arbitrárias de vídeo, de monitores ou de câmeras. O valor de referência para a normalização do vídeo foi escolhido como 512×512 *pixels*. A escolha do tamanho de normalização levou em conta parâmetros de desempenho, como baixo tempo de processamento e baixo consumo de recursos do processador, além da habilidade de detecção das marcas para resoluções diferentes. Valores de resolução de referência abaixo de 128×128 atendiam aos requisitos de consumo de recursos e afetava a velocidade do processamento mas limitavam o tratamento de resoluções altas como HDTV, enquanto valores de resolução acima de 512×512 afetavam a velocidade de processamento.

Independentemente da resolução do vídeo, que pode ser SDTV ou HDTV, da resolução da câmera utilizada para a detecção da marca ou da resolução do monitor de vídeo do usuário, todo o processamento de inserção e recuperação da marca é feito utilizando o vídeo normalizado para o valor de referência [115][116].

2.3.1.2 Detecção de Cantos e Seleção de Partes do Quadro

Após a normalização dos quadros, um processamento baseado em detecção de cantos espaciais utilizando um detector de Harris [117][118][119] é realizado e pontos de interesse com maior estabilidade local são selecionados (LMSP - *Locally Most Stable Points*) [114]. Cada um desses pontos selecionados define áreas circulares candidatas ao recebimento das marcas d'água.

Após a definição dessas áreas, são identificadas todas as que apresentam superposição espacial. Essa superposição pode afetar a inserção, e posterior detecção das marcas. Dessa forma, entre as áreas que apresentam superposição espacial, é escolhida aquela que apresentar um maior LMSP e descartadas as outras superpostas a esta. Esse processo de seleção é realizado para cada conjunto de áreas que apresentam superposição, resultando num conjunto de áreas circulares sem superposição. As marcas são então inseridas nesse conjunto de áreas circulares sem superposição.

O detector de cantos de Harris seleciona os pontos LMSP da matriz de segundo-momento (M) da imagem $f(x, y)$ mostrada na Equação 2.6[117].

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial f(x,y)}{\partial x}\right)^2 & \left(\frac{\partial f(x,y)}{\partial x}\right)\left(\frac{\partial f(x,y)}{\partial y}\right) \\ \left(\frac{\partial f(x,y)}{\partial x}\right)\left(\frac{\partial f(x,y)}{\partial y}\right) & \left(\frac{\partial f(x,y)}{\partial y}\right)^2 \end{bmatrix} \quad (2.6)$$

onde $\frac{\partial f(x,y)}{\partial x} = f(x, y) * \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$ e $\frac{\partial f(x,y)}{\partial y} = f(x, y) * \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T$.

A resposta (R) do detector de Harris é determinada pelo determinante e pelo traço de M ($\det(M) = AB - C^2$ e $\text{tr}(M) = (A + B)$), dada pela Equação 2.7:

$$R = AB - C^2 - k(A + B)^2 \quad (2.7)$$

onde k é uma constante. Os pontos LMSP são determinados comparando R com um valor limiar definido através dos testes subjetivos realizados.

2.3.1.3 Normalização em Rotação

As áreas circulares selecionadas pelo detector de Harris são normalizadas em rotação, de forma que mudanças de rotação da imagem não afetem a posterior detecção das marcas. Dessas áreas circulares, são retiradas novas áreas quadradas inscritas nas circunferências.

2.3.1.4 Marca de testes

A marca escolhida é um vetor binário de 64 bits com o seguinte formato: $w = \{w_1, w_2, \dots, w_n\}$.

onde w_i é gerado por uma sequência pseudoaleatória originada por uma chave K (expressão textual).

Como exemplo de chave, foi utilizada a expressão textual “**marca de agua**”, que após passar por uma transformação pseudoaleatória utilizando um *hash* com funções *XOR* (função lógica “ou exclusivo”). Esse processo gera a marca binária w de 64 bits.

2.3.1.5 Aplicação da marca em coeficientes DCT das áreas

O Vetor binário contendo a marca é inserido em n pares de coeficientes DCT (Transformada Discreta de Cossenos - *Discrete Cosine Transform*) de média frequência seguindo ordem de varredura Zig-Zag [120][121]. A escolha da DCT deve-se a utilização em algoritmos de compressão de imagens e pelo fato dos seus planos de frequência se aproximarem do comportamento do sistema visual humano. A marca é inserida nos coeficientes de frequência média, que apresentam robustez e mudanças visualmente imperceptíveis na qualidade da imagem.

Em cada par de coeficientes adjacentes ($c1$ e $c2$) da DCT, um bit da marca w é inserido.

Cada uma das áreas quadradas é então marcada com uma sequência de bits no domínio DCT. Esta sequência de bits é gerada a partir de uma cadeia de caracteres conhecidos (expressão textual) que também será utilizada na detecção.

A seguir é mostrado o processo de aplicação da marca escolhida $w = \{w_1, w_2, \dots, w_n\}$ nos coeficientes da DCT. O valor T é escolhido como o limiar de invisibilidade da marca.

- Se $w_i = 1$ e $D = c1 - c2 < T$, aumentamos $c1$ para o novo $c1'$ e diminuímos $c2$ para o novo $c2'$ a seguir conforme as Equações 2.8 e 2.9:

$$c1' = c1 + \frac{T - D}{2} \quad (2.8)$$

$$c2' = c2 - \frac{T - D}{2} \quad (2.9)$$

- Se $w_i = 1$ e $D > T$, $c1$ e $c2$ mantêm-se sem alteração

- Se $w_i = 0$ e $D = c1 - c2 < T$, diminuimos $c1$ para o novo $c1'$ e aumentamos $c2$ para o novo $c2'$ a seguir conforme as Equações 2.10 e 2.11:

$$c1' = c1 - \frac{T - D}{2} \quad (2.10)$$

$$c2' = c2 + \frac{T - D}{2} \quad (2.11)$$

- Se $w_i = 0$ e $D > T$, $c1$ e $c2$ mantêm-se sem alteração

2.3.1.6 DCT Inversa

Após a inserção da marca nos coeficientes DCT de cada área quadrada selecionada, é feita a transformada DCT inversa (IDCT) destas áreas. Essas novas áreas marcadas que passaram por este processo DCT-marcação-IDCT são inseridas na imagem no lugar das áreas originais não marcadas, resultando no novo quadro marcado.

2.4 Função de Transferência de Modulação

O principal parâmetro físico indicador da resolução espacial sugerido na literatura [122][123][124] é a função de transferência de modulação (MTF – *Modulation Transfer Function*), que expressa a correlação espacial na saída entre o sinal de entrada e sua vizinhança.

A teoria das funções de transferência aplicada à avaliação do comportamento de sistemas dinâmicos, como os sistemas de comunicações, para a análise das respostas em função das frequências, passou a ser utilizada também no estudo dos sistemas de imagens, para análise em termos de sua resposta em função da frequência espacial.

O desempenho de um sistema de imagens pode ser determinado verificando-se a relação entre o sinal de entrada (objeto) e o sinal de saída (imagem) que este sistema apresenta [125]. Dessa forma, com o conhecimento da característica de transferência do sistema, pode-se prever o comportamento de sua saída para qualquer entrada conhecida. Portanto, pode-se prever a deterioração que a imagem final apresentará em relação ao objeto tratando o sistema como uma "caixa preta" [122] (Figura 2.10).

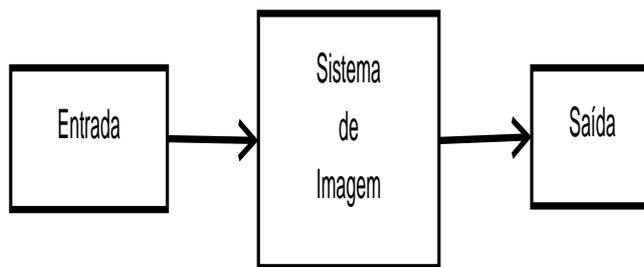


Figura 2.10: Característica de Transferência do Sistema de Imagens [126].

O sinal de entrada de um sistema de imagens pode ser representado por um padrão de fótons ou distribuição espacial de raios visíveis que, em cada ponto do receptor, representa a projeção da atenuação linear destes raios ao longo da trajetória entre a fonte e o detector, passando através do objeto. Já o sinal de saída é a imagem que representa tal propriedade do objeto. Em muitos casos, o termo intensidade é utilizado para se referir tanto ao sinal de entrada quanto ao de saída.

Para que a teoria das funções de transferência possa ser aplicada aos sistemas de imagens, duas propriedades devem ser observadas: linearidade e invariância espacial [127]. Um sistema é considerado linear se, para várias entradas distintas, a saída correspondente é igual ao somatório das saídas referentes a cada entrada isolada e se, multiplicando o sinal de entrada por uma constante, a saída também é multiplicada pela mesma.

Matematicamente, se dois objetos distintos $f_1(x,y)$ e $f_2(x,y)$ formam, respectivamente, as imagens $g_1(x,y)$ e $g_2(x,y)$, o sistema será linear se a imagem formada pelos dois objetos combinados na forma $[\alpha f_1(x,y) + \beta f_2(x,y)]$ resultar numa imagem que pode ser representada na forma $[\alpha g_1(x,y) + \beta g_2(x,y)]$. A propriedade de invariância espacial requer que a imagem de um objeto possua as mesmas características, independentemente da posição no plano do objeto em que ele é colocado [127].

Os processos físicos por trás da formação de imagens são tais que o sistema de imagem não registrará perfeitamente todas as flutuações espaciais da intensidade do sinal de entrada. Tais degradações podem aparecer de inúmeras formas, como por exemplo, na forma de “borramento” espacial, adição ou aumento dos níveis de ruído, redução do contraste e introdução de distorções geométricas.

A resolução espacial é a propriedade que descreve a capacidade de um sistema de imagens de exibir com exatidão detalhes de dois objetos de alto contraste, pequenos e

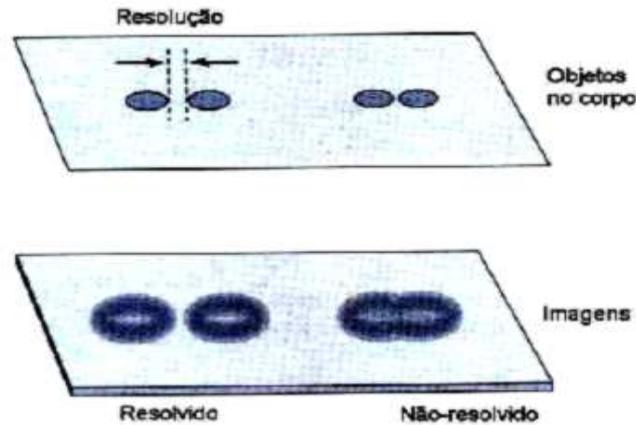


Figura 2.11: Resolução Espacial [126].

adjacentes, portanto, depende da resposta em frequência do sistema [125][128][129].

A resolução espacial é muitas vezes chamada simplesmente de resolução e pode ser interpretada como a separação mínima entre objetos pequenos (no corpo), ou seja, quando suas imagens deixam de ser distinguíveis umas das outras, conforme ilustrado na Figura 2.11. Em algum ponto, os dois objetos ficam tão próximos um do outro que eles parecem ser apenas um objeto e, neste ponto, se perde a resolução espacial [126][130]. Em sistemas de imagens, a resolução espacial depende das distorções introduzidas pelo sistema de imagem.

Uma maneira útil de expressar a resolução de um sistema de imagem é fazendo uso do domínio da frequência espacial [125][126][130]. O conceito de frequência espacial está relacionado à quantidade de ciclos que a função senoidal (sinal de entrada) apresenta por unidade de distância. Os sinais que apresentam uma quantidade pequena de ciclos por unidade de distância possuem baixa frequência espacial, enquanto que os que apresentam uma quantidade alta possuem alta frequência espacial.

A utilidade deste conceito em análise de desempenho de sistemas de imagem se justifica no fato de todos os objetos poderem ser considerados como compostos de uma soma de faixas de frequências espaciais com amplitudes e fases diferentes. Portanto, um objeto pode ser representado em dois sistemas de coordenadas distintos, tanto no domínio espacial quanto no domínio das frequências espaciais [125]. Uma simplificação conceitual muito comum, porém não estritamente correta, é pensar em uma função senoidal como uma função quadrada. A onda quadrada representa um padrão alternado de faixas com densidades diferentes na imagem. Com uma onda quadrada, cada ciclo se torna um

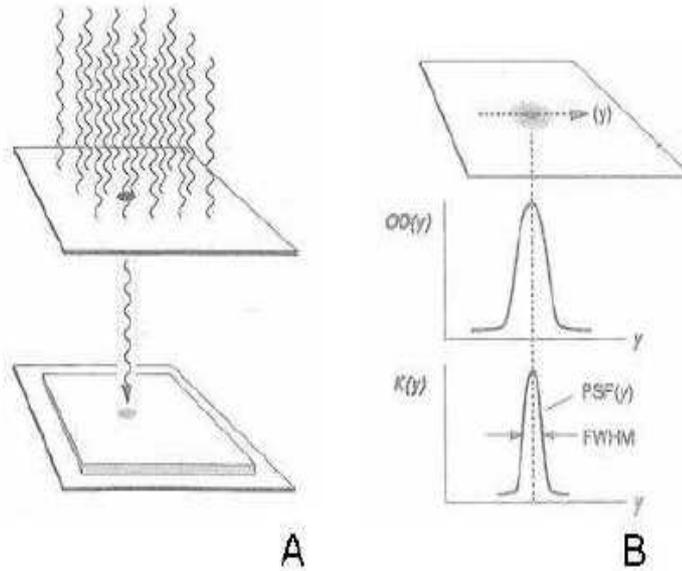


Figura 2.12: Função de Espalhamento de Ponto - PSF [126].

par de linhas – faixa clara e faixa escura. Desse modo, a unidade de frequência espacial é expressa como pares de linha por milímetro (lp/mm) [129][126][130]. A relação entre o tamanho de meio ciclo, Δ , e a frequência espacial é dada pela Equação 2.12, a saber:

$$F = \frac{1}{2}\Delta \quad (2.12)$$

Um objeto quadrado de largura Δ pode ser visto em termos da frequência espacial de acordo com a equação 2.12, ou seja, a frequência espacial é apenas uma outra forma de tratar os tamanhos de objetos. Baixos valores de frequência correspondem a objetos grandes, enquanto que altas frequências espaciais correspondem a objetos pequenos [126].

Uma outra forma conceitual de entender (e medir) a resolução espacial de um detector (sistema de detecção) no domínio do espaço é estimulando o detector com um único sinal de entrada pontual, e então observar como o detector responde. A imagem produzida a partir de um único estímulo pontual de entrada em um detector é chamada de função de espalhamento de ponto (PSF - *Point Spread Function*) [125][129][126][130].

A função de espalhamento de ponto (PSF) é definida como sendo a distribuição de intensidade dos raios luminosos dada pela imagem obtida a partir de uma fonte pontiforme de abertura infinitamente pequena *pinhole* e de intensidade unitária [125]. Dessa

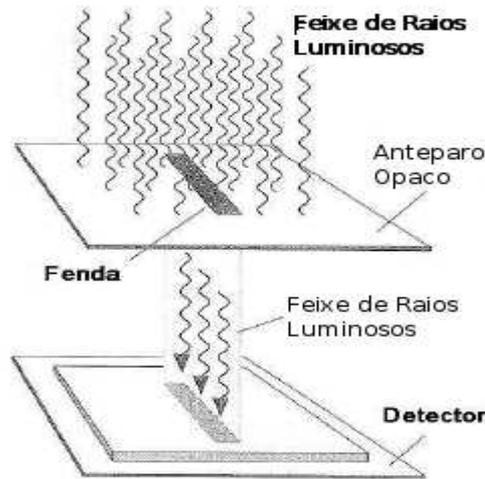


Figura 2.13: Função de Espalhamento de Linha - LSF [126].

forma, a PSF descreve a resposta de um sistema de imagem a uma fonte de imagem puntiforme (Figura 2.12).

A PSF descreve as propriedades de “borramento” de um sistema de imagem. Uma imagem é uma grande coleção de pontos individuais e, se um sistema de imagem é exposto a esta coleção de pontos, a PSF atua borrando cada um dos milhões de pontos de entrada que constituem a imagem. O processo de quebra da imagem de entrada em seus estímulos pontuais, que então são borrados individualmente utilizando a PSF do sistema de imagem, resulta em uma soma para se obter a imagem final através de uma operação matemática de convolução [126][130] Equação 2.13:

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x - \xi, y - \eta) PSF(\xi, \eta) d\xi d\eta \quad (2.13)$$

onde $f(x,y)$ é a distribuição de intensidade do objeto.

Na prática, a medição experimental da PSF é complicada. Isto se deve ao fato da abertura da câmera pontual ser muito pequena em relação ao tamanho do ponto focal, produzindo assim uma quantidade de raios luminosos muito baixa para o detector digital, e a dificuldade de alinhamento do ponto de abertura da câmera com o ponto focal. Estes problemas experimentais podem ser minimizados pela medição de outra característica de medição do sistema, a partir da qual se pode calcular a PSF, chamada de Função de Espalhamento de Linha (LSF – *Line Spread Function*).

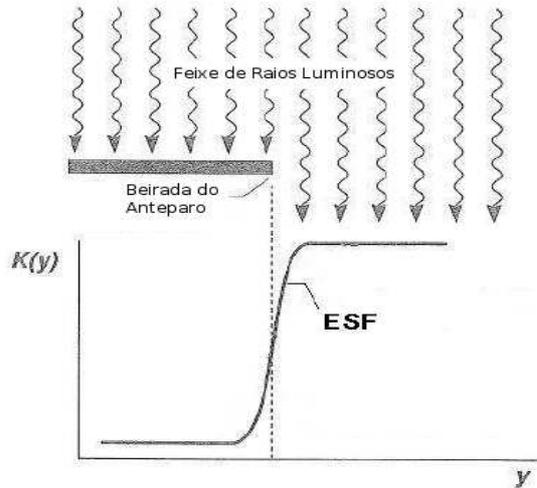


Figura 2.14: Função de Espalhamento de Borda - ESF [126].

A Função de Espalhamento de Linha (LSF), num sistema linear e isotrópico, representa a distribuição de intensidade de raios luminosos na imagem de uma fenda infinitamente longa iluminada com intensidade unitária. Neste caso, o anteparo opaco contém uma fenda estreita ao invés de um *pinhole* (Figura 2.13).

A LSF pode ser vista como uma coleção linear de um grande número de PSFs [126]. Na prática, a LSF é obtida fazendo-se uma varredura na direção perpendicular ao comprimento da fenda. Como os sistemas de imagem possuem características bidimensionais, para o cálculo da PSF a partir da LSF, torna-se necessária a medida das LSFs correspondentes a todas as orientações possíveis da fenda no plano-objeto, o que seria ainda mais complicado. No entanto, se o sistema de imagem for isotrópico, isto é, possuir uma PSF com simetria rotacional, o formato da LSF passa a ser independente de sua orientação no plano da imagem, sendo necessária apenas uma medida para que a LSF seja equivalente a PSF [125][129].

A LSF é mais fácil de ser medida experimentalmente porque a fenda linear que é usada precisa ser alinhada com o ponto focal apenas em uma dimensão (enquanto que o pequeno buraco usado para medir a PSF precisa ser alinhado em duas dimensões) [126][130].

Colimar um feixe de raios luminosos em uma linha fina o suficiente para a aproximação da LSF pode ser difícil experimentalmente [131]. Contudo, tanto a PSF quanto a LSF podem ser estimadas a partir da resposta de um sistema de imagem a um degrau de absorção uniformemente iluminado (Figura 2.14)

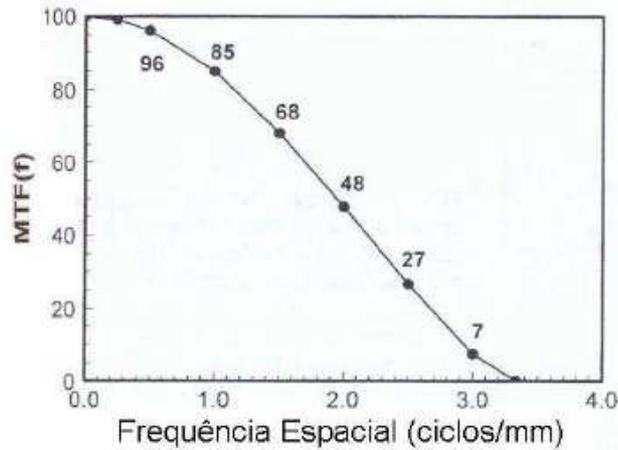


Figura 2.15: MTF em função da frequência espacial [126].

Tanto a PSF quanto a LSF podem ser determinadas por meio da diferenciação da ESF [126].

A Função de Transferência de Modulação (MTF – *Modulation Transfer Function*) descreve o desempenho de um sistema de imagem, em termos de sua habilidade de reproduzir sinais contendo uma ampla faixa de frequências espaciais e caracteriza ainda, a capacidade de resolução do sistema. A MTF é dada pela razão entre a amplitude (modulação) do sinal de saída em uma dada frequência e a amplitude (modulação) do sinal de entrada na mesma frequência espacial. Dessa forma, a MTF é uma função da frequência espacial, descrevendo a dependência da amplificação ou ganho do sistema com relação à frequência espacial [125].

Se todas as frequências espaciais do sinal forem passadas pelo sistema com o mesmo ganho (sistema perfeito), então a MTF do sistema será 1 para todas as frequências espaciais.

Praticamente todos os sistemas de imagem reproduzem sinais de alta frequência espacial pior do que os de baixa frequência, e conseqüentemente suas MTFs diminuem com o aumento da frequência espacial, conforme mostrado na Figura 2.15.

Um sistema de imagem que fornece uma boa reprodução das altas frequências espaciais produzirá imagens mais nítidas. Por outro lado, um sistema de imagem que possui uma MTF mais baixa para as altas frequências, produzirá uma imagem com menor definição espacial [125].

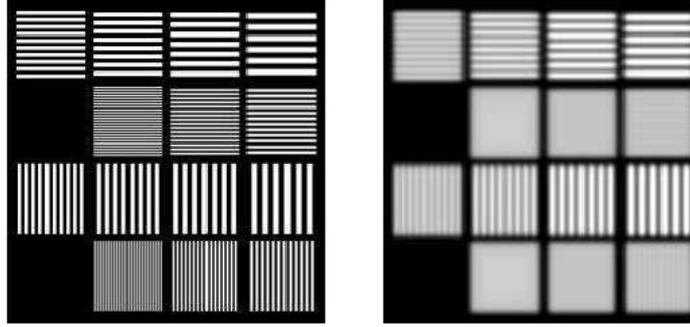


Figura 2.16: MTF e contraste

Na prática, a resolução é geralmente determinada pela utilização de padrões de teste que consistem em grupos formados por barras e espaços contendo a mesma largura, sendo que a largura de cada grupo diminui gradualmente, aumentando assim o número de pares de linha por unidade de distância (1 mm).

O "borramento" introduzido pelo sistema de imagem implica na diminuição da diferença entre a amplitude das barras e espaços conforme estes se tornam cada vez mais próximos, até que em um dado ponto não há mais diferença. Neste ponto os objetos deixam de ser "resolvíveis".

Dessa forma, a resolução pode ser representada também pela frequência espacial limite da MTF que tem sua magnitude abaixo de um valor crítico específico [125].

A função de transferência do sistema, nesse caso representada pela $MTF(f)$, pode ser determinada a partir da resposta de um sistema de imagem a um degrau de absorção uniformemente iluminado, conforme a equação 2.14:

$$MTF(f) = \int_{-\infty}^{\infty} PSF(x)e^{-2\pi i f x} dx \quad (2.14)$$

2.4.1 Método de estimativa da Resolução – MTF

A metodologia empregada para a medida da função de transferência de modulação (MTF) foi a análise de resposta ao degrau de absorção, ou seja, o método da borda.

Os resultados pelo método de borda se mostraram equivalentes aos resultados obtidos com o método da fenda [132].

Esse método consiste em obter a função de espalhamento de borda (ESF) do sistema a partir da imagem de um objeto opaco (disponível no vídeo de calibração) que apresente uma borda bem definida [132][133][124][123][134], conforme o esquema ilustrado na Figura 2.17.

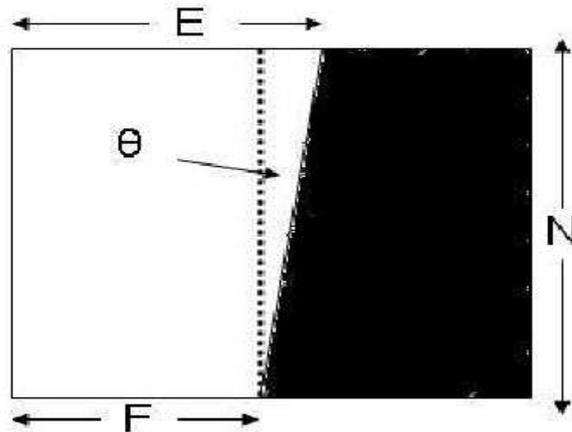


Figura 2.17: Imagem binária e método para o cálculo do ângulo (Equação (2.15))

Uma borda vertical pode ser utilizada para a estimativa da $MTF(f)$ até a frequência de Nyquist [123]. Entretanto, se a MTF possuir componentes de frequências maiores que este limite, estas causarão alteração na parte relativa às frequências mais altas da curva da MTF (*aliasing*). Por isso utiliza-se uma borda inclinada, a qual tem o efeito de aumentar a amostragem na determinação da ESF (projetada), que por sua vez melhora a exatidão da estimativa da MTF .

Portanto, a amostragem depende fortemente do ângulo de inclinação da borda porque quanto menor for o ângulo, maior será a amostragem e, portanto, melhor será a determinação da ESF . Por fim a MTF é obtida pela transformada de Fourier da PSF . A taxa de amostragem ideal da ESF projetada deve ser aproximadamente igual a 10% do tamanho do pixel, pois assim se garante um número razoável de pixels projetados e ao mesmo tempo promove uma boa resposta em frequência. Esta relação é suficiente para evitar erros associados à diferenciação de elementos finitos. Essa relação de 10% é alcançada fazendo o ângulo da borda ficar entre 2° e 10° [132].

Neste trabalho, dada a imagem digital do vídeo de calibração, o ângulo da borda é determinado precisamente através do esquema apresentado na Figura 2.17 e pela

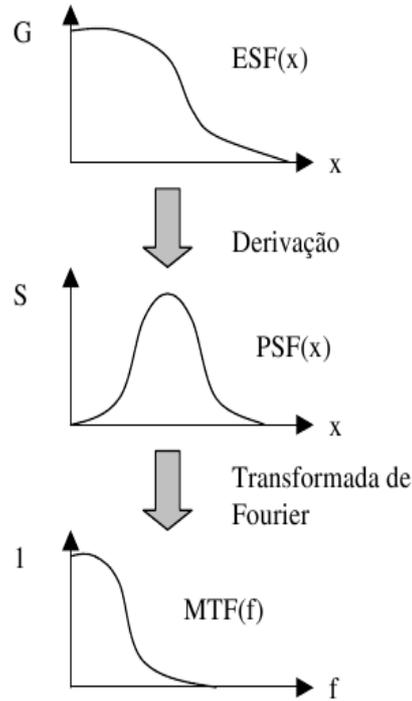


Figura 2.18: Representação dos passos para o cálculo da MTF

equação (2.15) [132][133].

$$\theta = \tan^{-1}\left(\frac{E - F}{N}\right) \quad (2.15)$$

onde E, F e N são as distâncias apresentadas na Figura 2.17.

Em seguida, a ESF é reconstituída a partir da região de interesse selecionada. Este método recebe o nome de reconstrução por projeção da borda, pois reconstitui a ESF combinando as projeções da borda bidimensional e realizando a média entre as regiões de sobreposição [132][133].

O procedimento completo para a determinação da MTF está ilustrado nas Figuras 2.18 e 2.19.

A função de espalhamento de linha (LSF) é obtida pela derivação numérica da ESF discreta. O processo de diferenciação é realizado pela convolução discreta da ESF com

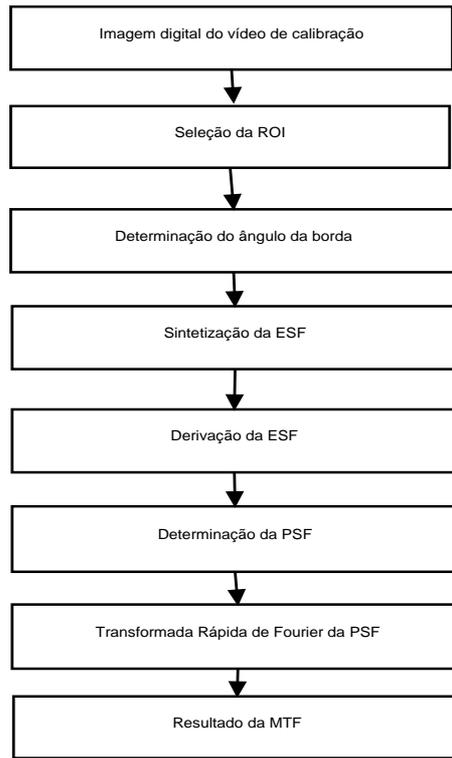


Figura 2.19: Diagrama de passos para o cálculo da MTF

a máscara de convolução ilustrada na Equação 2.16 [133]:

$$Md(r) = | - 1 | - 1 | - 1 | 0 | 1 | 1 | 1 | \quad (2.16)$$

$$LSF(r) = dESF(r)/dr = ESF(r) * Md(r) \quad (2.17)$$

A função de transferência de modulação é obtida pelo módulo da transformada de Fourier da LSF, normalizada com ganho unitário na frequência nula conforme Equação 2.18:

$$MTF(f) = |\{LSF(r)\}| / \int_{-\infty}^{\infty} LSF(r) dr \quad (2.18)$$

2.5 Acuidade visual humana

A capacidade do olho para resolver detalhes é conhecida como "acuidade visual". De forma simplificada, o olho humano normal pode distinguir padrões de linhas alternadas em preto e branco com características tão pequenas quanto um minuto de arco ($1/60$ graus ou $\pi/(60 * 180) = 0,000291$ radianos). Essa é a definição de visão normal 20-20 [135]. Para a maioria dos humanos, um padrão de frequência espacial maior que este aparecerá apenas como um tom de cinza. Padrões de baixo contraste com a frequência espacial máxima também aparecerão como cinza.

A afirmação de que o olho não consegue distinguir características menores do que um minuto de arco é muito simplificada. O olho tem uma resposta MTF assim como qualquer outro componente óptico, como ilustrado na Figura 2.20[135]. O eixo horizontal mostra a frequência angular em ciclos por grau (CPD). A MTF é mostrada para tamanhos de pupilas de 2 mm (iluminação brilhante, $f/8$), a 5,8 mm (pouca iluminação, $f/2.8$). Em 30 CPD, correspondendo a um minuto de arco de tamanho, a MTF cai de 0,4 para a pupila de 2 mm a 0,16 para a pupila de 5,8 mm. Esses são os valores "f-stop" do olho humano, de forma semelhante aos valores "f-stop" das máquinas fotográficas.

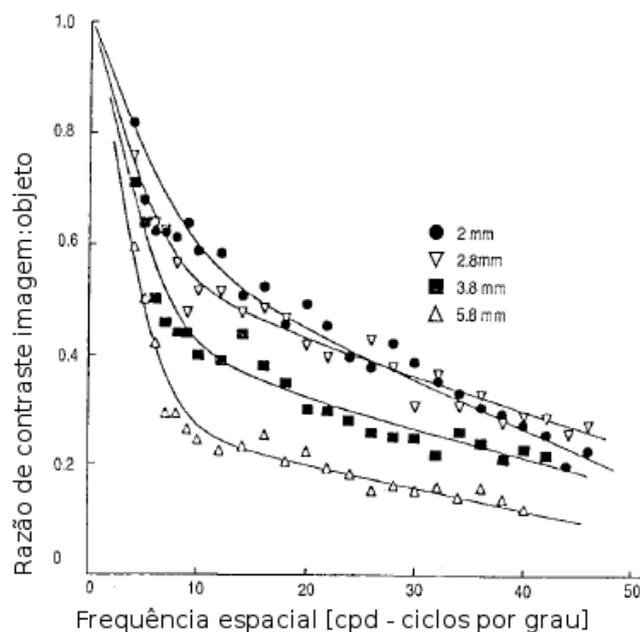


Figura 2.20: MTF do olho humano [135].

A MTF do olho humano, que é limitada em altas frequências angulares pelo sistema

óptico do olho e densidade de cones, não descreve totalmente a resposta do olho. Interações neuronais, tais como inibições laterais, limitam a resposta do olho em baixas frequências angulares, isto é, o olho é insensível às mudanças de densidade graduais. A resposta total do olho é conhecida como função de sensibilidade ao contraste (CSF - *Contrast Sensitivity Function*). A CSF de pico para os níveis de luz brilhantes (típico de condições de visualização de impressão) está na ordem de 6 e 8 ciclos por grau. A Figura 2.21 utiliza a aproximação da equação 2.21) cujo pico fica abaixo de 8 ciclos / grau.

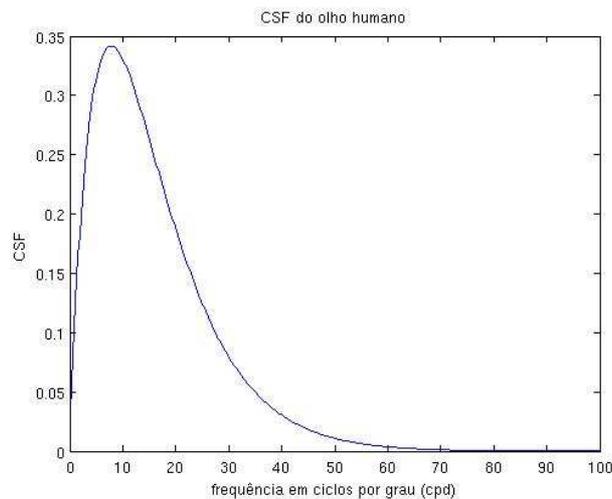


Figura 2.21: Função de sensibilidade ao contraste do olho humano [135].

A CSF é utilizada para a medida perceptual de nitidez de imagens conhecida como SQF descrita na seção 2.6.

2.6 Fator de Qualidade Subjetiva – SQF (*Subjective Quality Factor*)

O SQF é uma medida de nitidez percebida da imagem. O SQF é bastante utilizado na indústria fotográfica, mas permaneceu desconhecido para a maioria dos fotógrafos devido à dificuldade de sua medição e cálculo utilizando os recursos computacionais disponíveis em sua época, na década de 1970 [136]. O SQF inclui os efeitos de:

- MTF : Resposta de Frequência Espacial (SFR);
- CSF : Função de sensibilidade ao contraste do olho humano;
- altura da imagem; e

- distância de visualização.

O SQF foi introduzido no trabalho de Grainger [136] e os principais utilizadores do SQF eram fabricantes de equipamentos fotográficos, para o desenvolvimento de produtos, e a Revista Fotografia Popular [137], para testes de lentes. Nesse trabalho foi verificada a correlação do SQF com a preferência dos espectadores através de testes subjetivos [137][136].

A Equação 2.19 utilizada para o cálculo da SQF é:

$$SQF = K \int (CSF(f)MTF(f)/f)df \quad (2.19)$$

onde K é a constante de normalização e equivale a Equação 2.20:

$$K = 100\% / \int CSF(f)df \quad (2.20)$$

2.6.1 Relação entre o SQF e a MTF

A MTF é uma medida de nitidez do dispositivo ou sistema enquanto o SQF mede a nitidez percebida da imagem. O SQF é calculado a partir da MTF, da função de sensibilidade ao contraste (CSF) do sistema visual humano e de uma relação entre a altura da imagem e a distância de visualização.

O SQF pode ser comparado com a MTF pelas características a seguir:

A MTF mede o contraste da imagem em função da frequência espacial, descrevendo a nitidez do dispositivo ou sistema. A MTF não leva em consideração a distância de visualização.

O SQF mede a nitidez percebida como uma função da altura da imagem e distância de visualização do monitor, também descreve experiência do telespectador. O SQF utiliza os efeitos da distância de visualização e do sistema visual humano.

O SQF tem a mesma interpretação, independentemente do tamanho de imagem. Isso significa que uma imagem com $SQF = 92$ apresentaria a mesma "sensação de qualidade" tanto para uma imagem de tamanho em centímetros 4×6 como para uma de 24×36 . Se considerarmos as medidas da MTF na superfície da imagem, a interpretação é feita de forma diferente para diferentes tamanhos de imagens. O visualizador tem a tendência de aceitar MTFs mais baixas para imagens maiores, porque estas imagens são normalmente vistas a partir de distâncias maiores.

Os parâmetros ópticos do sistema são extraídos através da função de contraste [36] da Equação 2.21.

$$CSF(f) = 0.114 f e^{(-0.1254 f)} \quad (2.21)$$

onde CSF é a função de contraste do sistema e f é a frequência espacial em ciclos/grau. Para imagens digitais, em *pixels*, podemos utilizar a Equação 2.22 que pode ser utilizada para transformar a frequência espacial para ciclos por pixel (cpp).

$$f(\text{ciclos/grau}) = f(\text{ciclos/pixel}) \left(\frac{\pi}{180}\right) \left(d \frac{n_{PH}}{PH}\right) \quad (2.22)$$

onde n_{PH} é o número de pixels verticais (ao longo da altura da imagem), d é a distância de visualização e PH é a altura da imagem em unidades de distância.

As Equações 2.21 e 2.22 são utilizadas no cálculo do SQF da Equação 2.19 relacionando os parâmetros da MTF e da CSF.

3 Metodologia

3.1 Passos da Metodologia

O diagrama da Figura 3.1 mostra os passos da metodologia. No primeiro passo é apresentado o detalhamento do desenvolvimento da função de qualidade. O segundo passo detalha a calibração do ambiente do usuário. Este passo é importante, pois permite que a medida de qualidade possa ser feita junto ao usuário, de forma que a medida seja próxima ao espectador, refletindo sua percepção. O segundo passo, portanto, é executado no ambiente de cada usuário para que os parâmetros ópticos desses ambientes, como distâncias e tamanhos, sejam coletados para a parametrização da função de qualidade. Após a calibração e parametrização da função de qualidade com os dados do ambiente vem o terceiro passo que consiste da operação do sistema, na qual o usuário poderá então verificar a qualidade dos vídeos distribuídos por um sistema real em operação.

A Figura 3.2 apresenta o fluxograma que detalha a geração da função de qualidade apresentado no primeiro passo da Figura 3.1. Nas subseções a seguir são detalhadas as etapas apresentadas na Figura 3.2.

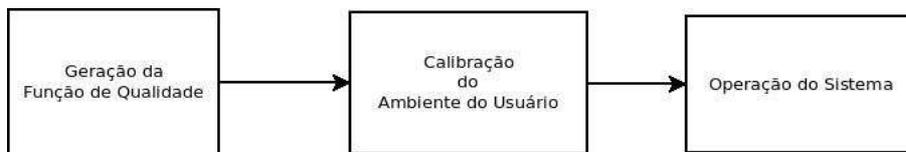


Figura 3.1: Etapas da Metodologia Proposta.

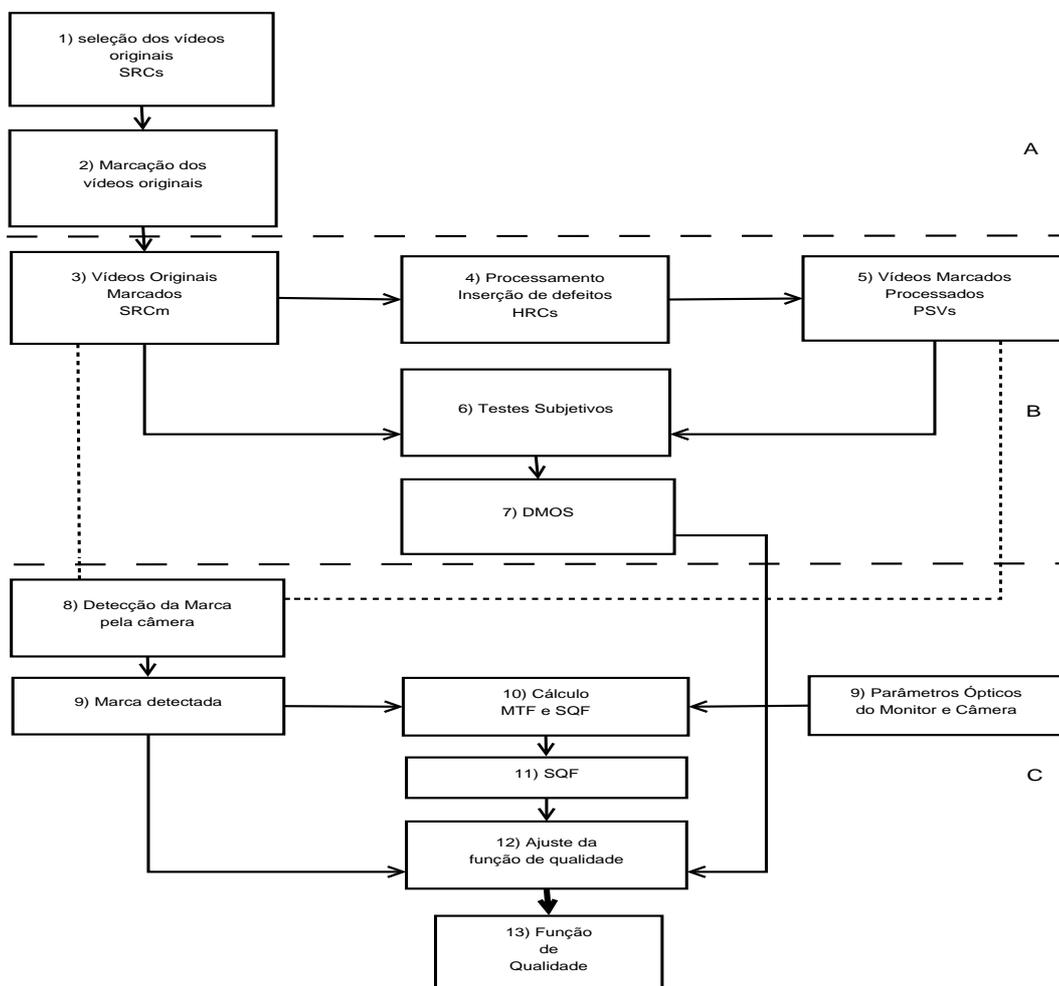


Figura 3.2: Método para geração da função de qualidade.

Nesta Metodologia a numeração de cada bloco do fluxograma está de acordo com a sequência cronológica das atividades realizadas, as quais serão descritas como etapas.

3.1.1 Etapa 1 - Seleção dos Vídeos Originais

A primeira etapa da metodologia, também comum à maioria dos projetos de métricas de qualidade de vídeo, é a seleção das sequências de vídeo a serem utilizadas. Foram selecionadas sequências utilizadas na fase I dos testes SDTV (*Standard Definition TV*) do VQEG e sequências em formato HDTV (*High Definition TV*).

As sequências de vídeo originais, também apresentadas na literatura como sequências de vídeo de referência ou SRCs (*Source Reference*) [91] são vídeos de alta qualidade sem a inserção de degradações. Para uma maior uniformidade e possibilidade de comparação de resultados de testes realizados em diversos laboratórios é desejável que estas sequências estejam disponíveis livremente para a comunidade científica. Essas sequências de vídeo são disponibilizadas em diversos repositórios disponíveis na Internet, em diversos formatos [28][88][89][90][91][138]. As sequências de referência foram escolhidas para abranger uma vasta gama de aplicações típicas de conteúdo para TV, como esportes, música e videoclipes. Além disso, foram selecionadas por conterem várias características, tais como superfícies planas, texturas complexas, movimento da câmera e de objetos, rostos e paisagens, abrangendo uma vasta gama de complexidade de codificação [91]. Cada sequência de vídeo tem uma duração de 10 segundos.

As sequências de vídeo utilizadas neste trabalho apresentam dois formatos que diferem apenas pela resolução em *pixels*: SDTV (720x480) e HDTV (1280x720). Nesses formatos, os vídeos compostos originalmente analógicos e com cores no formato NTSC foram convertidos em digitais com componentes dos seguintes sinais: luminância (Y), diferença de cor azul (CB), e diferença de cor vermelho (CR). A forma de amostragem da recomendação BT.601 do ITU-R [20] é conhecida como amostragem 4:2:2, onde o canal Y é amostrado em taxa completa enquanto os canais CB e CR são amostrados com metade da taxa do canal Y.

Esses vídeos seguem o padrão da Recomendação BT.601 do ITU-R e são armazenados no formato “*Big YUV*” ou também descrito como “*ABEKAS*” [23], onde o Y significa a informação de luminância, o U a diferença da cor azul (CB na BT.601), e V a diferença de cor vermelha (CR na BT.601). Nesse formato, todos os *pixels* são armazenados

sequencialmente em um arquivo binário contínuo. Os *pixels* das imagens em cada quadro são armazenados sequencialmente pelas linhas de vídeo como *bytes* na seguinte ordem: CB0, Y0, CR0, Y1, CB2, Y2, CR2, Y3, etc., onde o índice numérico representa o número do pixel [23].

A sequência de vídeo *Park Joy* foi escolhida para este estudo de verificação da qualidade. Um dos quadros desta sequência de vídeo é apresentado na Figura 3.3.



Figura 3.3: Quadro da Sequência de Vídeo *Park Joy*.

3.1.2 Etapa 2 - Marcação dos Vídeos Originais

Nessa etapa são inseridas as marcas de testes nos vídeos selecionados na etapa 1. Na metodologia proposta, os vídeos são divididos em quadros que receberão as marcas.

Um aspecto importante para escolha do tipo de marca a ser inserida nos vídeos está relacionado com o tipo de sistema utilizado para a verificação da qualidade de vídeo. O objetivo da metodologia é que a medida da qualidade de vídeo seja feita independentemente do tipo da rede, do tipo do vídeo e do tipo de dispositivo terminal do usuário, permitindo o monitoramento da qualidade em pontos próximos aos usuários, com dispositivos simples, de baixos custo e complexidade. Esses dispositivos devem interferir minimamente com os equipamentos já em utilização pelo distribuidor de vídeo. Para o

atendimento destes objetivos, é utilizada uma câmera de vídeo desacoplada do sistema de recepção no ambiente do usuário, ou seja, a câmera não está fisicamente conectada ao receptor do usuário.

Pelo fato de estar desacoplada do sistema de recepção do usuário, não há sinais de sincronização chegando na câmera, o que é um desafio para a decodificação da marca d'água. A marca escolhida deve ser resistente às deformações geométricas e deve permitir a sua recuperação sem sincronismo com a marca inserida. Utilizamos uma Marca d'água do tipo “*RST Invariant*” (*rotation, scale, translation invariant*) nesta metodologia [110][111][112][113].

As Figuras 3.4 e 3.5 mostram o processo de inserção da marca de testes. No corrente estudo de caso, a marca é inserida apenas na componente de luminância da sequência de vídeo. A Figura 3.4 apresenta a sequência normalizada para o tamanho de referência conforme a seção 2.3.1.1. A Figura 3.5 mostra as áreas que serão marcadas como na seção 2.3.1.2.



Figura 3.4: Sequência de Vídeo *Park Joy* Normalizada.

3.1.3 Etapa 3 - Vídeos Originais Marcados

Esta etapa resulta em um novo conjunto de sequências de vídeos (processados com inclusão de marcas resultantes da etapa 2), que servirão como base para a verificação de

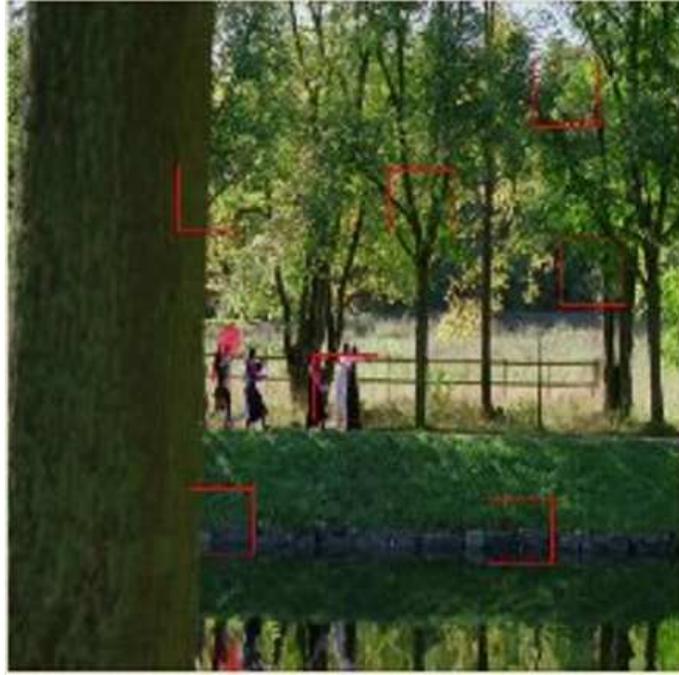


Figura 3.5: Áreas da Sequência de Vídeo *Park Joy*.

qualidade do vídeo distribuído, utilizando a recuperação das marcas de testes. As sequências de vídeo desta etapa são utilizadas como as novas sequências originais (SRC_m) ou de referência na métrica de qualidade objetiva [138].

A Figura 3.6 apresenta a sequência de vídeo após a inserção da marca.

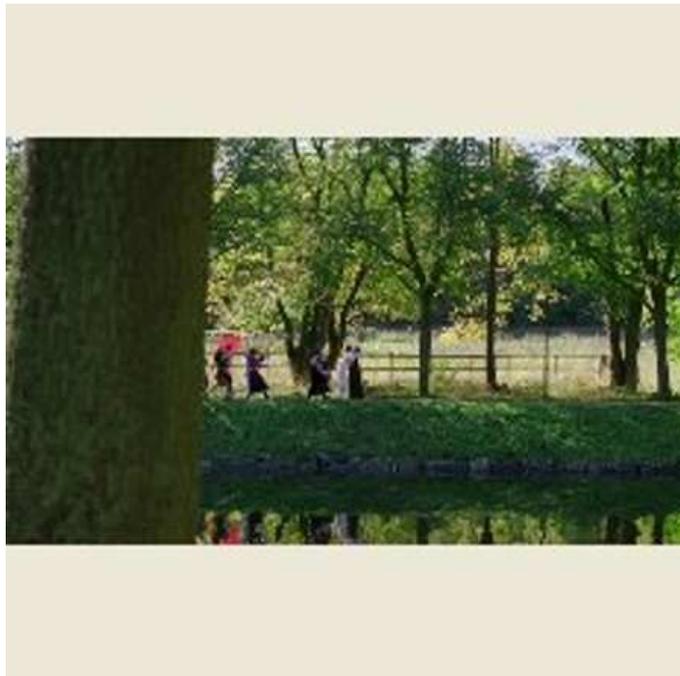


Figura 3.6: Quadro da Sequência de Vídeo *Park Joy* Marcada.

3.1.4 Etapa 4 - Processamento e Degradações dos Vídeos

No projeto de métricas de qualidade de vídeo objetivas [138], deve ser verificado o comportamento do resultado da métrica em relação aos resultados subjetivos indicados por avaliadores. Essas sequências têm como origem as sequências originais marcadas da Etapa 3 que passaram por diversos processamentos, que podem degradá-las, podendo se manifestar como problemas para os usuários.

Nos processamentos efetuados nesta etapa foram utilizados circuitos hipotéticos de referência (*HRC - Hypothetic Reference Circuits*) [91] utilizando os métodos apresentados na Tabela 3.1. O processamento das sequências originais marcadas (SRC_m) pelos diversos *HRC* resulta em novas sequências de vídeo processadas (*PVS - Processed Video Sequences*). Os *HRC* simulam o comportamento de uma cadeia de vídeo desde a produção até o consumo do vídeo.

Tabela 3.1: Características dos Circuitos de Referência - HRC

Número HRC	Degradação
1	vídeo codificado com MPEG-2 taxa de bits 1Mbps
2	vídeo codificado com MPEG-2 taxa de bits 10Mbps
3	vídeo codificado com H.264 taxa de bits 300kbps
4	vídeo codificado com H.264 taxa de bits 1Mbps
5	vídeo codificado com H.264 taxa de bits 10Mbps
6	vídeo codificado com H.264 taxa de bits 30Mbps
7	vídeo codificado com H.264 taxa de perda de pacote 1%
8	vídeo codificado com H.264 taxa de perda de pacote 3%
9	vídeo codificado com H.264 taxa de perda de pacote 10%

3.1.5 Etapa 5 - Vídeos Marcados Processados

O processamento da etapa anterior gera como resultados um conjunto de sequências de vídeo processadas (*PVS*) que é um número múltiplo das novas sequências originais ($PVS = SRC_m \times (1 + nrHRC)$). O valor $nrHRC$ indica a quantidade de degradações (coluna 1 da Tabela 3.1) que devem ser analisados pelos testes subjetivos e pela métrica objetiva de qualidade.

A Figura 3.7 apresenta a sequência de vídeo processada (PVS) com o codificador H.264 com taxa de bits de 30Mbps.

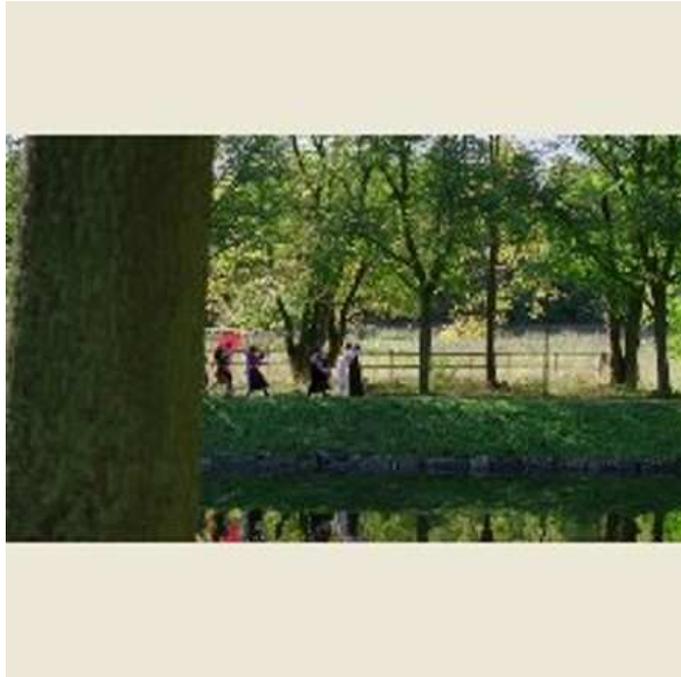


Figura 3.7: Quadro da Sequência de Vídeo *Park Joy* processada.

3.1.6 Etapa 6 - Testes Subjetivos

Os vídeos originais marcados (SRC_m) da etapa 3 e os vídeos processados (PVS) da etapa 5 são utilizados para a realização dos testes subjetivos. Os testes subjetivos contam com avaliadores humanos para a verificação da qualidade de vídeo de acordo com a metodologia descrita na seção 2.2.3.

Os avaliadores utilizaram o método DSCQS para atribuir a pontuação para cada vídeo. Nesse tipo de pontuação são avaliados um vídeo de referência contra um vídeo degradado gerando como resultado uma diferença entre as opiniões dos usuários DMOS.

Nesta seção, são selecionados dentre todos os testes subjetivos realizados neste trabalho, aqueles relacionados com a sequência de vídeo do caso de estudo.

3.1.7 Etapa 7 - Resultados DMOS

Os resultados de todos os avaliadores coletados nos testes subjetivos da etapa 6 serão utilizados para o ajuste da função de qualidade aos resultados psicofísicos.

A Figura 3.8 apresenta o resultado dos testes subjetivos da seção 3.1.6.

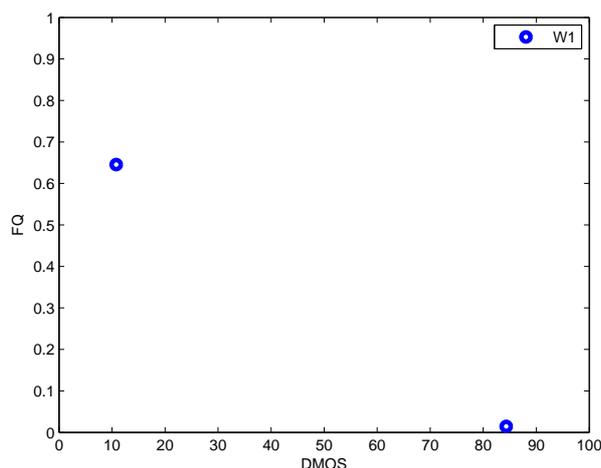


Figura 3.8: Valores DMOS para a Sequência de Vídeo *Park Joy*.

3.1.8 Etapa 8 - Detecção da Marca pela câmera de vídeo

Neste trabalho foi desenvolvido um método para a detecção utilizando a câmera de vídeo filmando o monitor de forma direta e sem sincronismo entre eles. Esse vídeo capturado é então processado para a extração da marca. O diagrama da Figura 3.9 apresenta a sequência de atividades para a detecção da marca, detalhada nos itens a seguir:

- No ambiente de recepção do usuário, uma câmera de vídeo filma a tela do monitor onde são assistidos os conteúdos audiovisuais. A câmera gera um "arquivo cru" (*RAW file*) sem compactação ou processamento, contendo apenas os bits indicando cada pixel;
- Cada quadro desse arquivo *RAW* é separado, gerando uma sequência de quadros ou imagens estáticas;

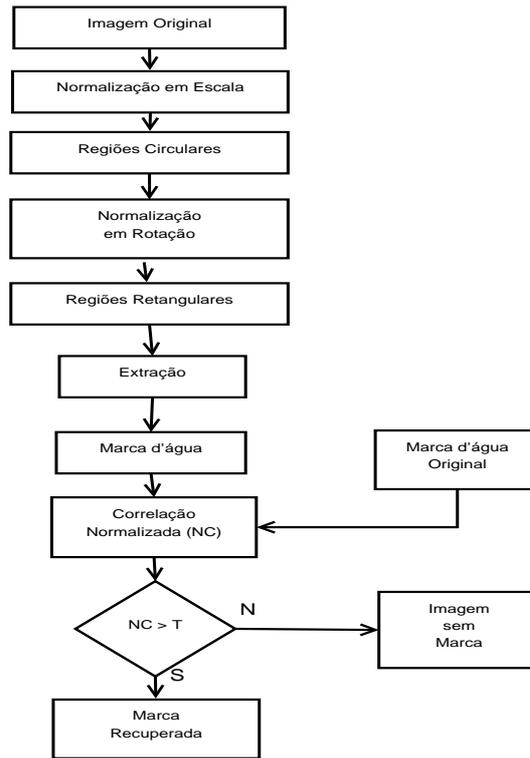


Figura 3.9: Detecção da Marca.

- Os quadros do vídeo passam por um processo de normalização equivalente ao executado na inserção, como na seção 2.3.1.1 [112];
- Os quadros normalizados passam por processamento de detecção de bordas de forma semelhante a seção 2.3.1.2. Esse processo seleciona áreas circulares onde a marca pode estar inserida [112];
- Cada área selecionada é testada pelo algoritmo de detecção da marca e pelo processo de seleção das áreas relevantes do quadro.

Essas áreas são transformadas para o domínio DCT, onde é feita a extração da marca em cada uma delas. A mesma chave utilizada na inserção é utilizada e é feita uma correlação entre a marca inserida e a marca recebida por cada área do quadro. Uma medida de qualidade da recuperação da marca é feita através da correlação normalizada da Equação 3.1 entre as marcas inserida e recuperada [112].

$$N_c = \frac{\sum_{i=1}^n w(i)\hat{w}(i)}{\sqrt{\sum_{i=1}^n w^2(i)}\sqrt{\sum_{i=1}^n \hat{w}^2(i)}} \quad (3.1)$$

onde N_c é a correlação normalizada entre as marcas, n é o comprimento da Marca, w é a marca original inserida durante a produção do conteúdo e \hat{w} é a

marca recuperada no ambiente do usuário.

Neste trabalho, o comprimento das marcas w e \hat{w} foi escolhido empiricamente como um vetor de 64 bits, após testes realizados com outros tamanhos de vetores.

O limiar de correlação normalizada para a detecção da marca foi escolhido como ($N_{cT} = 0.6$), de forma que a marca ficasse invisível para os observadores humanos e ao mesmo tempo identificável pela câmera de vídeo e pelo algoritmo de processamento. A escolha desse limiar foi baseada em testes subjetivos [21]. Um valor de correlação normalizada N_c acima de N_{cT} indica a presença da marca;

- O maior valor de N_c de cada área de um dado quadro se torna o valor N_{cframe} ;
- Um número fixo de quadros do vídeo (Num_{frames}) foi escolhido para o cálculo da correlação normalizada total do vídeo N_{cvideo} .

A escolha de um número de quadros maior que um, aumenta a eficiência da detecção da marca e ao mesmo tempo causa maior tempo de processamento. Neste trabalho, o valor escolhido para o número de quadros foi igual a 50 ($Num_{frames} = 50$).

Para melhorar ainda mais a eficiência da detecção da marca, eliminando alguns valores de N_{cframe} fora do comportamento médio, foi feito um ajuste dos valores de N_{cframe} para uma distribuição normal e utilizado o percentil 95% como o valor N_{cvideo} ;

- Esse N_{cvideo} foi então convertido para a mesma escala do SQF [137] para a utilização na função de qualidade. A Equação 3.2 mostra esta conversão de escala.

$$N_{cvideoQ} = 100 \times \frac{N_{cvideo} - 0.6}{(N_{cvideo} - 0.4) \times 0.6} \quad (3.2)$$

Na detecção da marca através da filmagem do monitor com a câmera de vídeo, de forma não sincronizada, há um grande desafio para que a marca seja decodificada de forma confiável. Para que a marca seja recuperada após passar por várias transformações geométricas, são necessários diversos processamentos que garantam a redundância de informações da marca.

Dessa forma, o vídeo é dividido em quadros, e o processamento é efetuado com a média dos resultados dos quadros. Cada quadro passa por normalizações em escala e rotação. Cada quadro é dividido em áreas circulares com maior significância (LMSP) e

as marcas são inseridas em quadrados inscritos nestas áreas. A redundância na inserção de múltiplas marcas em cada quadro, conforme a seção 3.1.2, permite que, na média, a recuperação seja obtida (seção 3.1.8).

No quadro do vídeo apresentado na Figura 3.10 podem ser observadas as áreas com possibilidade de haver marcas. É possível notar que as áreas detectadas nesse quadro não são exatamente as mesmas apresentadas no quadro da Figura 3.5, porém algumas das áreas podem encontrar coincidências. Na Figura 3.5, apenas uma das áreas apresenta coincidência com a área marcada na inserção, sendo possível recuperar esta marca. No conjunto de quadros, mais marcas são recuperadas e retirada a média dos resultados, aumentando a robustez do processo.

O valor de $N_{cvideoQ}$ obtido para este vídeo é $N_{cvideoQ} = 79$.



Figura 3.10: Detecção de Marcas na Sequência de Vídeo *Park Joy*

3.1.9 Etapa 9 - Parâmetros Ópticos

Os parâmetros ópticos dos componentes do sistema são utilizados para o cálculo da MTF do sistema óptico. Neste trabalho são utilizados os parâmetros da câmera de vídeo, do monitor de vídeo e da distância de visualização. A Tabela 3.2 apresenta os dados utilizados.

Tabela 3.2: Parâmetros da Câmera e do Monitor

Equipamento	Parâmetro	Valor
Câmera Canon D10	tamanho do sensor	6.17x4.55mm
	quantidade de pixels	4000x3000
	tamanho do pixel	1.54x1.52μm
Monitor Samsung P2270HN	tamanho da diagonal	22" (55.88cm)
	quantidade de pixels	1920x1080
	tamanho do pixel	0.248x0.248mm
Ambiente	distância de visualização	100cm

3.1.10 Etapa 10 - Cálculo da MTF e do SQF

A MTF total do sistema óptico combina as MTF de cada componente. No sistema óptico estudado neste trabalho estão incluídos os efeitos da câmera de vídeo, que incluem o sensor e a lente, os efeitos do monitor de vídeo e os efeitos do ambiente onde estes equipamentos estão sendo utilizados [139]. A Figura 3.11 mostra um diagrama típico para a MTF total do sistema óptico. A Equação 3.3 mostra como pode ser calculada a MTF no domínio da frequência.

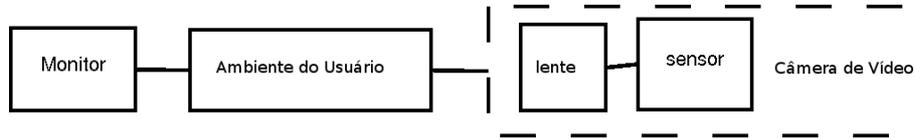


Figura 3.11: Diagrama da MTF do sistema óptico

$$MTF_T = MTF_{mon} \times MTF_{env} \times MTF_{len} \times MTF_{sen} \quad (3.3)$$

onde MTF_T é a MTF total do sistema óptico, MTF_{mon} , MTF_{env} , MTF_{len} e MTF_{sen} são as MTFs referentes ao *monitor*, *ambiente*, *lente da câmera* e *sensor da câmera* respectivamente.

O domínio da frequência é utilizado, pois a teoria da função de transferência do sistema óptico torna o cálculo da MTF mais simples e rápido [139][140]. No domínio da frequência (Transformada de Fourier), o produto das MTFs de cada componente do sistema resulta na MTF total do sistema óptico.

A Equação 3.3 pode ser simplificada mantendo apenas os termos devidos ao ambiente MTF_{env} e aos componentes eletrônicos utilizados MTF_{elect} , resultando na Equação 3.4.

$$MTF_T = MTF_{elect} \times MTF_{env} \quad (3.4)$$

onde $MTF_{elect} = MTF_{mon} \times MTF_{len} \times MTF_{sen}$, da Equação 3.3.

Utilizando a MTF_T obtida com o método *slanted-edge* [141], é possível obter os valores do SQF através da Equação 3.5 e da Função de Qualidade (FQ) da Equação 3.6. A MTF_T possibilita o cálculo dos valores da MTF_{elect} ou dos valores da MTF_{env} . Se um destes valores for fixado, como por exemplo, mantendo os mesmos equipamentos eletrônicos ($MTF_{elect} = MTF_{elect.1}$) fixos durante os ensaios, pode ser calculada a outra MTF (MTF_{env}).

O SQF (fator de qualidade subjetivo) da Equação 3.5 [137] é calculado utilizando a MTF_T das Equações 3.3 e 3.4. O SQF representa um valor objetivo relacionado com a CSF do olho humano. O SQF tem grande importância pois foi verificado empiricamente em um conjunto de testes de nitidez de fotos conduzidos com observadores humanos [137][136].

$$SQF = 100 \times \int_{v^-}^{v^+} (CSF(v_d) MTF(v_d) / v_d) dv_d \quad (3.5)$$

onde v_d é a frequência espacial (posição no espaço) em ciclos por grau na retina e os limites de integração são $v^- = 3$ e $v^+ = 12$ ciclos por grau para o olho humano [137].

As Figuras 3.12 e 3.13 apresentam os gráficos da MTF e do SQF respectivamente. Os gráficos são obtidos conforme os passos da seção 3.1.10.

3.1.11 Etapa 11 - Fator de Qualidade Subjetivo

A MTF e os parâmetros ópticos dos componentes do sistema permitem o cálculo do

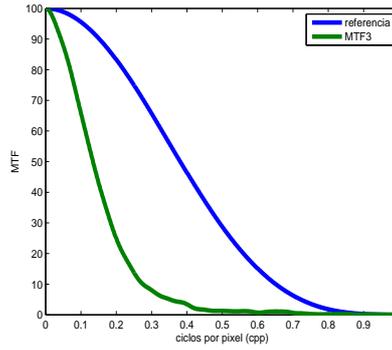


Figura 3.12: MTF da Sequência de Vídeo *Park Joy*

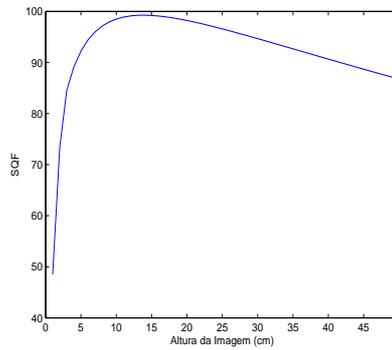


Figura 3.13: SQF da Sequência de Vídeo *Park Joy*

Fator de Qualidade Subjetivo (SQF). Com o SQF e a marca detectada pela câmera é possível calcular a métrica de qualidade objetiva.

O valor obtido de SQF do gráfico da Figura 3.13 ($SQF = 93$) será utilizado para o cálculo de FQ .

3.1.12 Etapa 12 - Ajuste da Função de Qualidade

Utilizando o valor de $N_{cvideoQ}$ calculado na recuperação da marca conforme a Equação 3.2 e o valor de SQF calculado utilizando a MTF do sistema óptico da Equação 3.5, ambos ajustados para a mesma escala, a função de qualidade é definida como FQ na Equação 3.6.

Na formulação de FQ foi utilizada uma combinação linear dos componentes da recu-

peração da marca N_c e das características ópticas e visuais SQF .

$$FQ = a \times N_{cvideoQ} + b \times SQF \quad (3.6)$$

O ajuste dos pesos da Equação 3.6 utiliza além dos valores $N_{cvideoQ}$ e SQF , os resultados DMOS. Esses valores são adaptados aos resultados dos testes subjetivos conduzidos [21] durante os passos de definição da métrica de qualidade da seção 3.1.6 [138].

3.1.13 Etapa 13 - Função de Qualidade

A métrica de qualidade é obtida por uma função que depende da marca inserida e dos parâmetros ópticos do sistema e possui a seguinte formulação da Equação 3.7:

$$FQ = 0.408 \times N_{cvideoQ} + 0.349 \times SQF \quad (3.7)$$

Do ajuste feito com os dados dos testes subjetivos em SQF e da combinação linear entre os parâmetros de recuperação das marcas e do SQF são obtidos os coeficientes de ponderação mostrados na Equação 3.7.

Com o ajuste dos dados obtem-se a função de qualidade como na Equação 3.8, com $a = 0.408$ e $b = 0.349$.

$$FQ = 0.408 \times N_{cvideoQ} + 0.349 \times SQF \quad (3.8)$$

Como resultado, para esta sequência de vídeo, $FQ = 0.408 \times 79 + 0.349 \times 93$, o valor de qualidade para este vídeo é $FQ = 65$.

3.2 Calibração e Registro do sistema

O passo de calibração do ambiente do usuário deverá ser feito uma vez, antes das medidas de qualidade de vídeo em regime de operação. A calibração do ambiente será

responsável pela parametrização da função de qualidade com os dados do ambiente do usuário, tais como resoluções do monitor e distância de visualização.

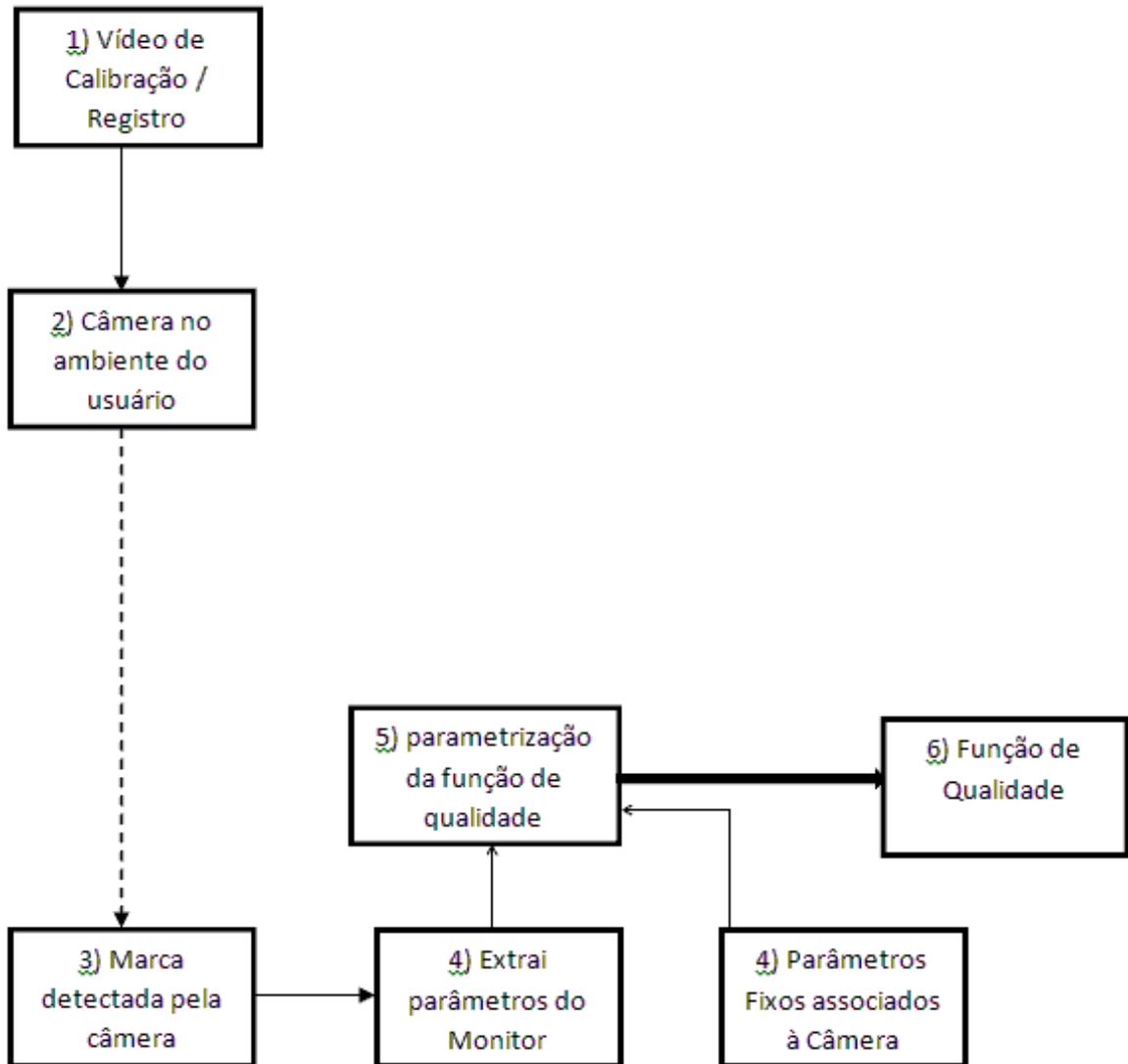


Figura 3.14: Calibração do sistema no ambiente do usuário

A Figura 3.14 ilustra os passos para a adaptação do sistema ao ambiente do usuário. O usuário será orientado a realizar a calibração a cada vez que alterar o seu ambiente de visualização de vídeo. Esse ambiente é composto pelo monitor, que apresenta os vídeos, a resolução e o tamanho deste monitor, as distâncias de visualização entre o monitor e a posição de visualização e as características da câmera. O processo de calibração utiliza um vídeo fornecido juntamente com o sistema. Esse vídeo de calibração apresenta uma sequência com degradações conhecidas para que a câmera do usuário possa calcular os parâmetros de resolução e tamanho vertical do monitor e distância de visualização.

Uma vez obtidos esses parâmetros, eles serão gravados no sistema. Esses parâmetros

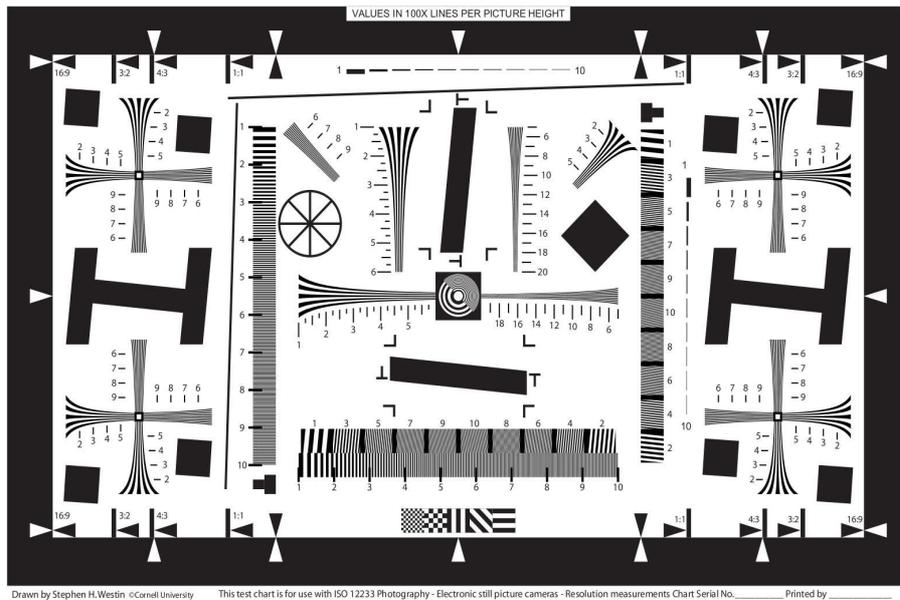


Figura 3.15: Diagrama ISO12233 utilizado como referência para o vídeo de calibração [141].

são inseridos no cálculo da métrica e permitem que o sistema seja parametrizável de acordo com o ambiente sob teste. Os parâmetros da câmera de vídeo são fixos e dependem das especificações da câmera que poderá ser fornecida com o sistema.

Para fazer a calibração do ambiente do usuário, foi gerado um vídeo de calibração, que deve ser enviado em conjunto com a câmera de vídeo para a etapa de calibração em cada ambiente. Esse vídeo é baseado no padrão ISO12233 da Figura 3.15 e será utilizado para a determinação da MTF_{env} de cada ambiente de usuário, calibrando parâmetros que personalizarão a FQ da Equação 3.6.

O padrão ISO12233 foi escolhido como base para o vídeo de calibração por apresentar regiões com alto contraste, diversas barras com resoluções distintas e regiões apresentando bordas inclinadas [141] adequadas para as medidas de MTF_T .

A geração do vídeo foi feita a partir da concatenação de várias sequências de vídeo com as características a seguir:

- Cada sequência de vídeo concatenada apresenta características comuns como: duração de 10 segundos com 25 quadros por segundo de taxa e resolução em *pixels* de 1280x720;
- A primeira sequência de vídeo (Tabela 3.3) foi criada pela simples repetição do

padrão de testes ISO12233 em cada quadro do vídeo sem nenhum tratamento, sem compressão ou codificação;

- As demais sequências de vídeo utilizadas foram geradas a partir desta primeira sequência, variando o codificador, a taxa de bits ou taxa de perda de pacotes como mostrado na coluna 2 da Tabela 3.3.

Tabela 3.3: Sequências utilizadas no vídeo de calibração

Sequência de Vídeo	Descrição	Fator de Qualidade
1	Sequência ISO12233 sem tratamento	100
2	vídeo codificado com H.264 e taxa de bits de 30Mbps	95
3	vídeo codificado com H.264 e taxa de bits de 3Mbps	80
4	vídeo codificado com H.264 e taxa de bits de 300kbps	10
5	vídeo codificado com MPEG-2 e taxa de bits de 10Mbps	90
6	vídeo codificado com MPEG-2 e taxa de bits de 1Mbps	20
7	vídeo codificado com H.264 e taxa de perda de pacotes de 1%	50
8	vídeo codificado com H.264 e taxa de perda de pacotes de 3%	40
9	vídeo codificado com H.264 e taxa de perda de pacotes de 10%	30

O fator de qualidade da coluna 3 da Tabela 3.3 foi estimado através de testes subjetivos realizados durante a execução deste trabalho [21]. Esses testes subjetivos foram feitos usando sequências de vídeo com as mesmas degradações mostradas na Tabela 3.3.

Utilizando os valores conhecidos de qualidade da Tabela 3.3 na função de qualidade FQ 3.6, o valor de $MTF_T = MTF_{T_{FQ}}$ pode ser calculado. Esta $MTF_{T_{FQ}}$ e a $MTF_{elect.1}$ calculada anteriormente podem ser substituídas na Equação 3.4 para a obtenção da MTF_{env} do ambiente do usuário.

No processo de calibração do ambiente, diversas medidas são efetuadas em intervalos de tempo constantes, menores que 10 segundos. A cada medida, o vídeo de calibração estará apresentando uma das sequências da Tabela 3.3. Nestas medidas, valores discretos que podem ser configurados no sistema de verificação de qualidade servem como parâmetros de ambiente para o cálculo da MTF e do SQF para aquele ambiente. Uma vez que o resultado da função de qualidade do vídeo de calibração é conhecido, conforme a coluna 3 da Tabela 3.3, os valores dos parâmetros utilizados que minimizarem o erro entre a Função de Qualidade e a Sequência de Vídeo são escolhidos para a cali-

bração do ambiente (MTF_{env}). Os valores discretos são testados nas Equações 2.21 e 2.22, do capítulo anterior, para o cálculo da CSF e do SQF.

3.3 Verificação da Qualidade em Operação

A Figura 3.16 apresenta o procedimento de operação do sistema após a calibração e que será utilizado no dia a dia do usuário. Em uma operação de distribuição de conteúdos audiovisuais é considerado que o distribuidor faça a marcação dos conteúdos antes de qualquer outro processamento que seja necessário para a distribuição (passo 2). Os vídeos marcados serão então enviados para os usuários, conforme a tecnologia utilizada por cada distribuidor.

Em um cenário em que um processo de medição de qualidade utilizando esta metodologia seja regulamentado e tornado obrigatório para os distribuidores de conteúdo audiovisuais, alguns aspectos devem ser considerados, como o processo de marcação do vídeo, que necessitará um novo processo de auditoria e certificação pelos órgãos competentes, da mesma forma que os processos atuais de telecomunicações [142].

O usuário poderá a qualquer momento, em tempo real, solicitar uma verificação da qualidade do conteúdo que estiver assistindo. Outra forma de operação do sistema de verificação de qualidade, este pode solicitar automaticamente, sem intervenção do usuário e de forma transparente, uma verificação da qualidade.

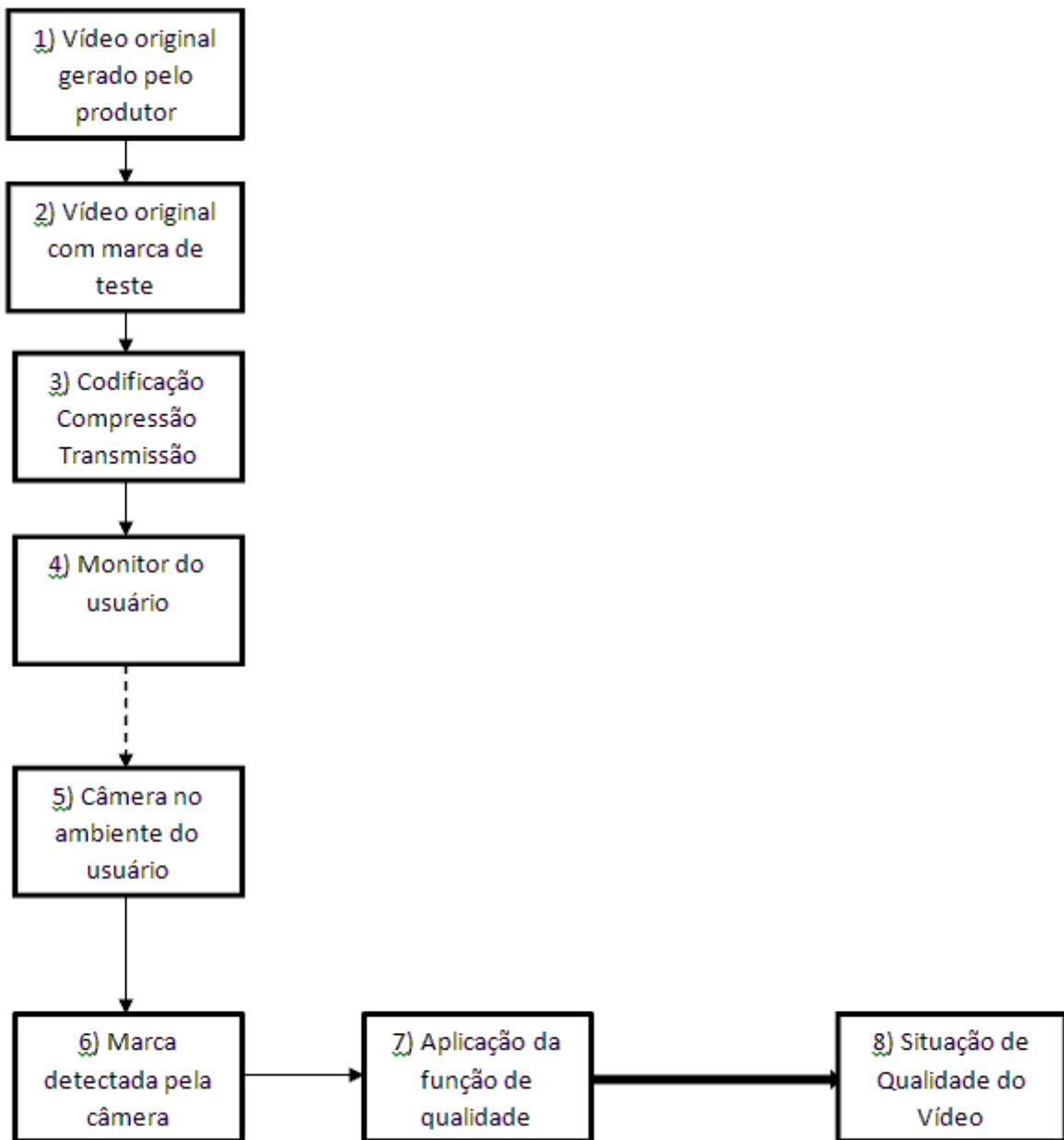


Figura 3.16: Procedimento de operacional para verificação da qualidade de vídeo

4 Resultados e Discussões

4.1 Introdução

Neste capítulo são apresentados os resultados obtidos durante a realização dos testes subjetivos, apresentados na seção 3.1.6 utilizando a metodologia da seção 2.2.3 do capítulo 2. Os resultados da geração da métrica de qualidade com a função de qualidade FQ são apresentados na seção 4.3.

Para o estudo da qualidade de vídeo, um conjunto de experiências foi conduzido no Departamento de Engenharia Elétrica da Universidade de Brasília e descritos na seção 4.2. A metodologia utilizada nos experimentos psicofísicos realizados neste trabalho é apresentada e são descritos o aparato, as instruções, o processo de geração das sequências de teste e a análise de dados.

4.2 Testes Subjetivos

As pesquisas em desenvolvimento [87] sobre qualidade de vídeo se baseiam em experimentos psicofísicos, uma vez que o resultado mais preciso de qualidade vem das estimativas dos receptores humanos ao material de vídeo. Para a concepção de métricas de qualidade objetiva de vídeo, é necessário realizar uma série de experimentos psicofísicos.

4.2.1 Aparato Utilizado

O aparato, representado de forma simplificada na Figura 4.1, é composto dos seguintes itens: uma câmera de vídeo, um tripé, um computador com disco de estado sólido SSD, um monitor de computador, um teclado e um mouse instalados em um laboratório com isolamento acústico. As características ópticas do ambiente de testes e dos componentes utilizados, como distâncias entre a câmera e o monitor (D), a altura da tela de

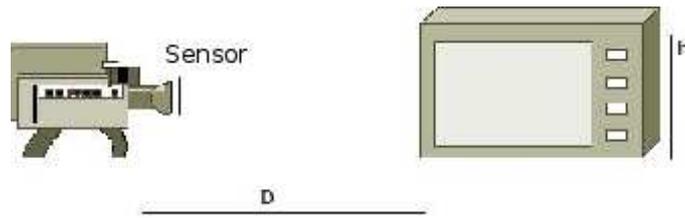


Figura 4.1: Diagrama do aparato experimental.

visualização (h) e o tamanho do sensor de imagens (I), são utilizados na composição das funções ópticas apresentadas nas Equações 2.19, 2.21 e 2.22.

A Tabela 4.1 mostra as especificações dos equipamentos utilizados. Os vídeos são mostrados usando um subconjunto de vídeos, conforme a Tabela 4.2. O comprimento das sequências de vídeo de teste foram limitadas a dez segundos e são armazenadas no disco rígido SSD e exibidas no monitor de vídeo do computador PC. Cada sequência de vídeo requer cerca de 440 Megabytes e todo o conjunto de teste para um experimento utiliza cerca de 5 Gigabytes de espaço em disco. Um programa com funcionalidade dedicadas a testes psicofísicos (Presentation [143]) é usado para controlar e executar o experimento e registrar os dados de cada avaliador. Depois da apresentação de cada sequência de teste, o programa exibe perguntas e registra as respostas dos avaliadores em um arquivo de dados. Os experimentos são executados com um avaliador de cada vez. O avaliador permanece sentado de forma alinhada em frente ao monitor, que é centrado ligeiramente abaixo da altura dos olhos para a maioria dos indivíduos, com o teclado e mouse com fácil acesso. A distância entre os olhos da pessoa e do monitor de vídeo é de três vezes a altura do monitor [21].

Tabela 4.1: Lista de Equipamentos utilizados

Monitor 1	Philips 52PFL7404D 52" Resolução 1920x1080 h = 650mm Tamanho Pixel 0,602x0,602mm
Monitor 2	Samsung P2270HN 22" Resolução 1920x1080 h = 268,596mm Tamanho Pixel 0,248x0,248mm
PC Dell	Modelo Inspiron 1525 Core2Duo 1,8MHz, 2GB ram Resolução 1440x900 h = 207mm
Câmera	Canon + Tripé Modelo D10 sensor CCD 12M 1/2.3" (6.17 x 4.55 mm) Resolução 4000x3000 pixels Tamanho Pixel 1,54ux1,52u Focal Length 35-105mm Ap F2.8-F4.9

Um total de 9 vídeos de alta qualidade foi utilizado neste trabalho, conforme a Tabela 4.2. Esses vídeos são limitados a dez segundos de duração e estão no formato HD (1280x720) com 25 quadros por segundo.

Um subconjunto de 6 das sequências originais disponíveis é escolhido. A Tabela 4.2 mostra os originais utilizados para a experiência.

Parâmetros utilizados para a criação das sequências de vídeos a serem utilizadas nos testes.

- Número de vídeos originais (9);
- Número de vídeos marcados (9);
- Codificadores MPEG2, H264 (2);
- Taxas de Bits 1 e 10 Mbps (MPEG2) (2), 300kbps, 1 e 30 Mbps (H264) (5);
- Erros de Transmissão, taxa de perda de pacotes 1 e 10% (2);
- Número Total de sequências de vídeo = $9 \times (1 + 1 + 2 + 5 + 2) = 99$.

Tabela 4.2: Sequências de Vídeo utilizadas

sequência	nome	utilização
W1	Park Joy	testes
W2	Ducks Take Off	treinamento
W3	Mobcal	testes
W4	Park Run	treinamento
W5	Bee	treinamento
W6	Stockholm	testes
W8	Crowd Run	testes
W10	Cactus	testes
W12	BQ Terrace	testes

4.2.2 Procedimento de testes

Os avaliadores foram escolhidos de um grupo heterogêneo de estudantes do ensino fundamental, ensino médio, ensino superior e servidores públicos, com idade variando

entre 11 e 55 anos de ambos os sexos. Eles são considerados leigos em relação a maioria dos tipos das degradações de vídeo digital e da terminologia associada. Os avaliadores foram orientados a usar óculos ou lentes de contato, caso eles precisassem deles para assistir TV. Neste experimento 18 indivíduos participaram dos testes para garantir resultados robustos [21].

Cada seção de testes foi dividida nas seguintes etapas:

- instruções orais;
- treinamento;
- ensaio preliminar;
- ensaios experimentais; e
- avaliação por entrevista.

Nesta seção cada uma das etapas é descrita.

- Instruções orais

Antes de cada experiência, um roteiro é elaborado para ajudar a realizar a experiência. O *script* contém detalhes do que o experimentador deve fazer em cada etapa da experiência. O roteiro contém instruções orais que devem ser dadas ao avaliador para ter certeza de que ele entende bem a tarefa a ser realizada. Antes das instruções orais, o experimentador precisa verificar se o avaliador está posicionado corretamente na distância adequada. Os avaliadores são apresentados às tarefas que serão realizadas nos ensaios experimentais. Eles são orientados a desconsiderar o conteúdo dos vídeos e a desempenho dos artistas e julgar apenas as deficiências que eles conseguem visualizar.

- Treinamento

Em cada sessão experimental o avaliador é solicitado a realizar uma tarefa que consiste em um julgamento sobre a degradação ou artefato vista no vídeo. A fim de completar essas tarefas os avaliadores precisam ter uma ideia de como vídeos sem degradações (originais) comparam com os vídeos com deficiência fortes. Portanto, uma sessão de treinamento está incluída no âmbito do processo

que consiste em apresentar os vídeos originais, seguido por exemplos de vídeos com as deficiências mais fortes encontradas no experimento. Os avaliadores são orientados a assistir a esses vídeos com cuidado e atribuir um valor de 0 para as piores qualidades de vídeo neste subconjunto.

- Ensaio Preliminar

Os julgamentos iniciais de um avaliador de teste são geralmente irregulares. É preciso tempo para um avaliador se acostumar com a tarefa de julgar ou detectar deficiências. A Recomendação ITU sugere que os primeiros cinco a dez ensaios sejam desprezados [21]. No nosso experimento, em vez de descartar os primeiros ensaios, nós incluímos um ensaio preliminar para testes práticos. Antes de iniciar esta fase, os avaliadores são informados de que esta é uma fase para testes práticos e que nenhum dado está sendo gravado.

Além de eliminar respostas erráticas, as provas práticas têm outros benefícios. Eles expõem os avaliadores a sequências de vídeo com uma boa variedade de deficiências e propiciam a chance de experimentar o processo de entrada de dados. Eles também permitem que os voluntários possam ganhar confiança em seus julgamentos.

- Ensaio experimental

Os dados subjetivos para cada experimento são coletados durante os ensaios experimentais. Esta fase é realizada com um conjunto completo de sequências de teste apresentados numa ordem aleatória. Após cada sequência de teste ser apresentada para o avaliador, perguntas sobre a qualidade do vídeo são realizadas. Nessas experiências, todos os avaliadores de teste visualizam todas as sequências de testes. O número de sequências de testes foi limitado a aproximadamente 100 sequências conforme apresentado na Seção 2.2.3.1.

As notas são registradas no próprio sistema de apresentação dos testes. O usuário atribui uma nota contínua (variando de ruim a excelente) a cada uma das duas sequências, o que é feito inserindo uma marca em cada uma das escalas. O sistema aguardará que as notas sejam dadas pelo usuário antes de passar à sequência seguinte.

As Figuras 4.2 e 4.3 apresentam fotos do estúdio do laboratório NMI (Núcleo de Multimídia e Internet), da Universidade de Brasília - UnB, utilizado para a realização dos testes subjetivos.



Figura 4.2: Fotos do espaço físico do Laboratório - Foto 1



Figura 4.3: Fotos do espaço físico do Laboratório - Foto 2

4.2.3 Resultados DMOS dos Testes Subjetivos

Os dados subjetivos coletados no experimento proposto na seção 3.1.6 e executado na seção 4.2 para as sequências de vídeo da Tabela 4.3 são comparados com os resultados da métrica objetiva de qualidade proposta neste trabalho e apresentados nos gráficos da Figura 4.5.

Na tabela 4.3 são apresentados os resultados dos testes subjetivos comparados com aqueles obtidos pela métrica objetiva de qualidade desenvolvida neste trabalho. De forma semelhante, pode-se observar nos gráficos das Figuras 4.4 e 4.5, a correlação entre os valores DMOS obtidos nos testes subjetivos e os resultados da métrica FQ desenvolvida.

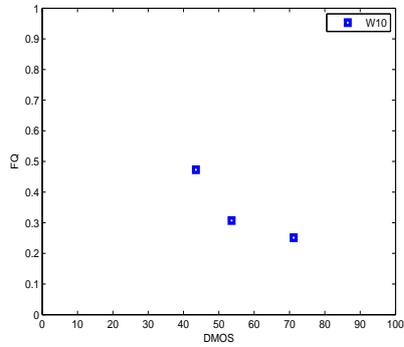
Tabela 4.3: Resultados dos Testes Subjetivos e Função de Qualidade

Código da Sequência	Nome da Sequência	Codificador e taxa de bits (Mbps) ou de perda de pacotes (%)	DMOS	FQ
W12	BQ Terrace	H.264 @ 30M	9.6667	0.9977
W12	BQ Terrace	MPEG2 @ 10M	14.0000	0.6756
W12	BQ Terrace	MPEG2 @ 1M	46.3333	0.3734
W12	BQ Terrace	H.264 @ 0.3M	53.2222	0.3010
W10	Cactus	H.264 @ 10.00%	71.1667	0.2510
W10	Cactus	MPEG2 @ 1M	53.6111	0.3071
W10	Cactus	H.264 @ 1.00%	43.5000	0.4729
W8	Crowd run	H.264 @ 0.3M	73.6667	0.0650
W8	Crowd run	H.264 @ 10.00%	61.6667	0.0918
W8	Crowd run	H.264 @ 1.00%	47.1111	0.4378
W8	Crowd run	H.264 @ 30M	8.0000	0.8584
W1	Park joy	H.264 @ 30M	10.7778	0.6455
W1	Park joy	H.264 @ 0.3M	84.3333	0.0145
W6	Stockholm	H.264 @ 1.00%	51.5556	0.4086
W6	Stockholm	H.264 @ 30M	11.2778	0.9806
W3	Mobcal	H.264 @ 10.00%	44.0000	0.3294
W3	Mobcal	H.264 @ 1.00%	34.5556	0.5536

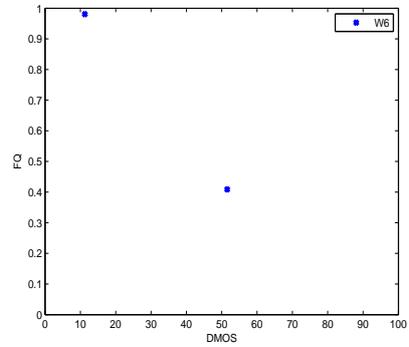
Os gráficos na figura 4.4 apresentam os resultados dos testes subjetivos efetuados nesta pesquisa. Os cinco primeiros gráficos representam os resultados obtidos nos testes subjetivos para cada sequência de vídeo da Tabela 4.3. O último gráfico apresenta os resultados dos testes subjetivos considerando todas as sequências de vídeo. Todos os gráficos mostram a relação entre os resultados DMOS dos testes subjetivos no eixo y contra os resultados da função de qualidade no eixo x. Os valores DMOS variam de 0 a 100, onde 0 representa a melhor qualidade possível e 100 a pior. Os valores de FQ estão no intervalo de 0 a 1, onde 0 indica baixa qualidade e 1 alta qualidade. Cada ponto nos gráficos representa um par de valores de DMOS e FQ para a mesma sequência de vídeo e um mesmo tipo de degradação.

Nos gráficos representando as sequências individuais, é importante ressaltar que devido ao método utilizado durante os testes subjetivos, conforme apresentado na seção 2.2.3, onde cada sequência de vídeo com ou sem degradações é apresentada para o avaliador de forma aleatória, de modo que na média dos resultados dos testes, algumas degradações específicas de determinadas sequências não foram apresentadas aos avaliadores. Este comportamento pode ser percebido nos gráficos com menor ou maior quantidade de pontos plotados.

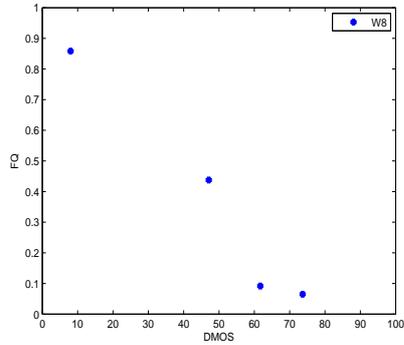
A tabela 4.3 apresenta os valores utilizados para cada gráfico. Nessa tabela podemos observar os tipos de degradação utilizados durante os testes. As maiores degradações levam o ponto do gráfico para o canto superior esquerdo, com o valor alto de DMOS (acima de 50) e baixo valor de FQ (abaixo de 0,5).



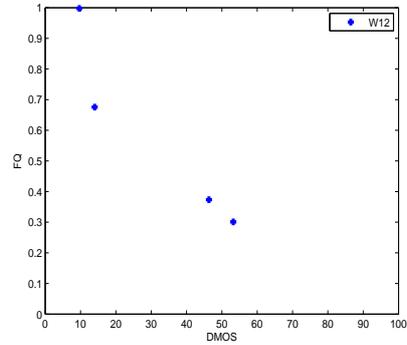
(a) sequência *Cactus*



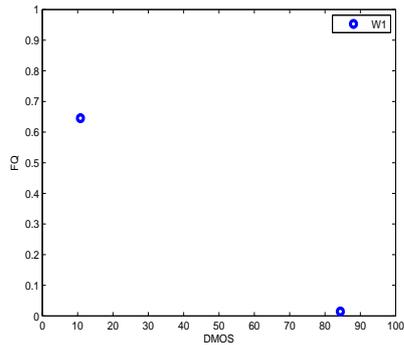
(b) sequência *Stockholm*



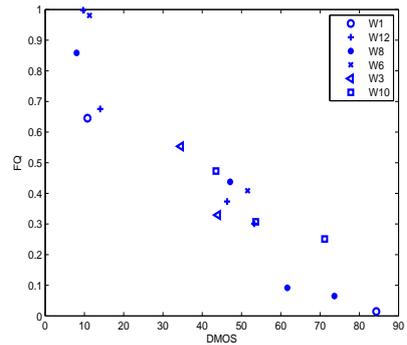
(c) sequência *Crowd run*



(d) sequência *BQ Terrace*



(e) sequência *Park Joy*



(f) Todas as sequências de vídeo

Figura 4.4: Resultados dos Testes Subjetivos

A Figura 4.5 apresenta a correlação linear entre os resultados das avaliações DMOS dos testes subjetivos com os resultados da métrica de qualidade FQ . O valor de $r = -0.94355$, apresentado em destaque na legenda, mostra como a métrica de qualidade acompanha as avaliações dos observadores humanos. A correlação linear é maior que 94% indicando uma excelente convergência de resultados.

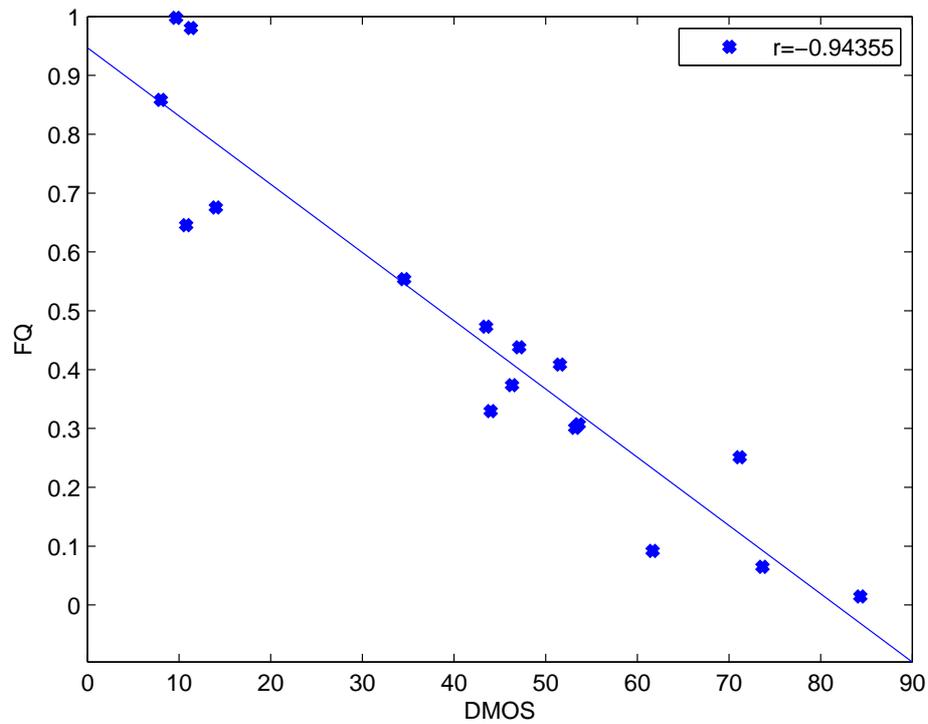


Figura 4.5: Correlação entre os testes subjetivos e a métrica proposta

4.3 Função de Qualidade - Métrica de Qualidade

A métrica objetiva de qualidade de vídeo proposta neste trabalho é baseada na premissa de que a marca d'água adicionada durante a geração do vídeo, sofrerá as mesmas degradações do vídeo que está carregando esta marca (vídeo hospedeiro), ou seja, o vídeo que se deseja medir a qualidade. Desta forma, a degradação da marca medida pela FQ pode ser utilizada como uma medida de qualidade do sistema.

Outro aspecto de interesse para a medida da qualidade é o ambiente de visualização do usuário, onde a marca d'água será decodificada através da câmera de vídeo. Dessa forma, uma combinação de parâmetros de recuperação da marca e do ambiente de visualização deve compor a métrica de qualidade.

Nas próximas seções são apresentadas as medidas da MTF e do SQF para a composição da métrica de qualidade da Equação 4.1.

$$FQ = a \times N_{cvideoQ} + b \times SQF \quad (4.1)$$

4.4 Cálculo da MTF

Os parâmetros ópticos do sistema são avaliados através da Função de Transferência de Modulação (MTF) que é calculada conforme o diagrama da Figura 2.19. A utilização de imagens de teste contendo bordas definidas pode ser calculada a MTF pelo método *slant-edge*.

A MTF total do sistema no domínio da frequência pode ser calculada a partir das MTFs de cada subsistema óptico que compõem estes sistema. Dessa forma, após o cálculo da MTF total do sistema óptico, são calculados componentes de MTFs de equipamentos utilizados nesse sistema conforme Equações 3.3 e 3.4.

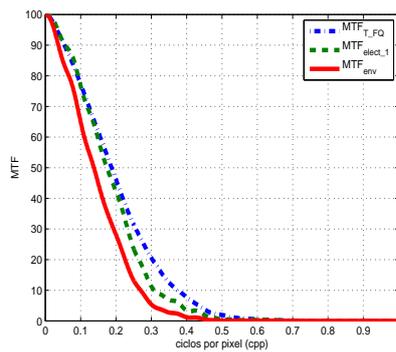
Neste trabalho as MTFs de interesse são aquelas referentes ao ambiente do usuário e as referentes aos equipamentos eletrônicos utilizados. Os equipamentos utilizados durante os testes são fixos conforme a Tabela 3.2 e dessa forma a MTF devida a eles pode ser isolada como uma constante MTF_{elect} .

A MTF_T do sistema é então utilizada para o cálculo do SQF e da métrica de qualidade objetiva.

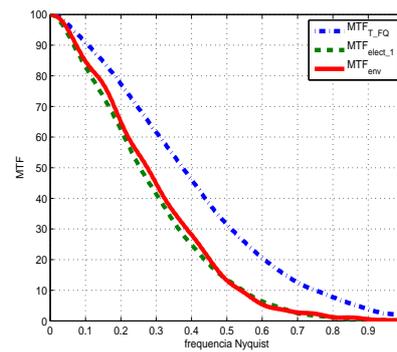
A Figura 4.6 mostra MTFs calculadas utilizando a metodologia desenvolvida. A MTF_T é utilizada como referência para a avaliação da MTF das imagens no monitor de vídeo, sem considerar as perdas do sistema óptico da câmera de vídeo e das distorções geométricas adicionadas no percurso entre a câmera e o monitor. Com essa referência medida com imagens ou vídeos com qualidade máxima, é possível estimar as perdas no sistema óptico.

A avaliação da qualidade de vídeos é feita considerando a perda (MTF_{env}) devido ao ambiente de visualização.

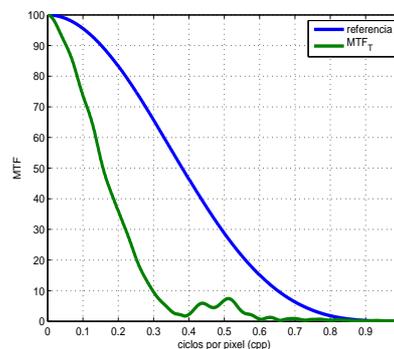
O resultado das MTFs do sistema óptico é utilizado no cálculo do SQF e da métrica de qualidade.



(a) MTF dos dispositivos eletrônicos



(b) MTF dos dispositivos eletrônicos pela frequência de Nyquist



(c) MTF total

Figura 4.6: MTF Total e MTF dos dispositivos Eletrônicos

Os gráficos das figuras 4.7 apresentam os resultados da Função de Transferência de Modulação (MTF) calculados utilizando o método apresentado nas seções 2.4 e 3.1.10 seguindo o diagrama da Figura 2.19. Os gráficos representam os resultados obtidos do cálculo da MTF para cada sequência de vídeo da Tabela 4.3.

Os gráficos mostram a MTF de referência calculada através de uma região de interesse (ROI) utilizando o vídeo de calibração sem degradações. A primeira curva apresentada em cada gráfico como MTF de referência (MTF_{REF}) é para um sistema desconsiderando as perdas do caminho óptico, como as perdas devido ao sistema óptico da câmera, de suas lentes e do caminho entre a câmera e o monitor. Essa MTF_{REF} leva em conta a MTF_{elec} , do monitor e do sistema de apresentação de imagens, que no caso dos testes apresentados nesta pesquisa é o computador Dell da Tabela 4.1.

A segunda curva apresentada em cada gráfico leva em conta a perda óptica do ambiente, conforme calculado pelas Equações 3.3 e 3.4 para o equipamento e distância da Tabela 3.2 utilizados nesta pesquisa. Podemos observar que há uma perda de resolução óptica devido aos componentes ópticos e a distância fixa de visualização, o que é representado pelo deslocamento dessa curva para a esquerda. A perda de resolução óptica se comporta como um filtro passa-baixos para a frequência espacial.

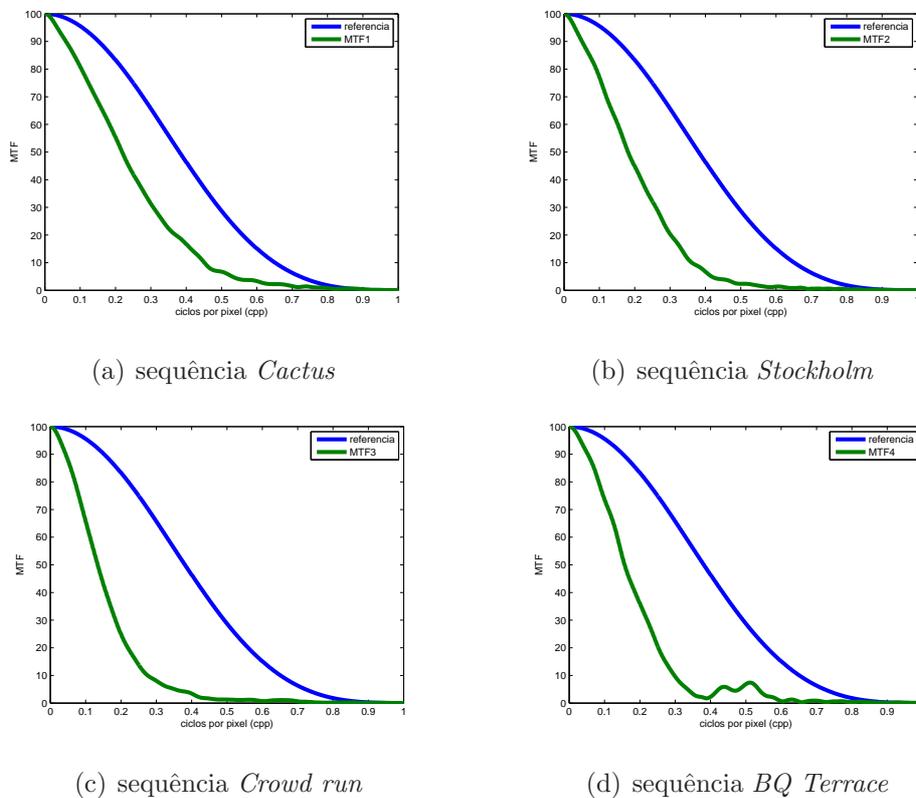


Figura 4.7: Função de Transferência de Modulação - MTF

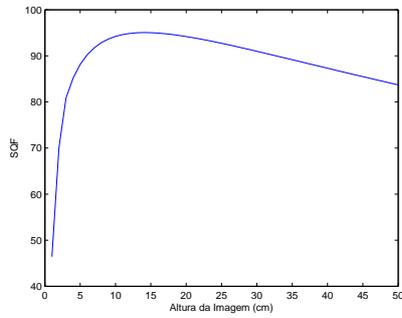
4.5 Cálculo do SQF

O valor da MTF_T é utilizado para calcular o Fator de Qualidade Subjetivo (SQF) através da Equação 2.19 repetida na Equação 4.2 para facilitar a visualização.

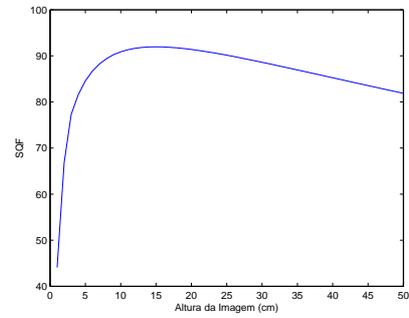
$$SQF = K \int (CSF(f)MTF(f)/f)df \quad (4.2)$$

A utilização do SQF, derivado da CSF, aproxima o comportamento da métrica de qualidade de vídeo com o sistema visual humano (HVS).

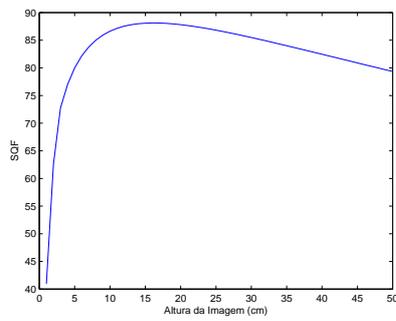
Os gráficos da figura 4.8 apresentam os resultados do Fator de Qualidade Subjetivo (SQF) calculados utilizando o método apresentado nas seções 2.6 e 3.1.10, substituindo a MTF calculada na seção 4.4 anterior no cálculo do SQF da Equação 4.2. Os gráficos representam os resultados obtidos do cálculo do SQF para cada sequência de vídeo da Tabela 4.4. Todos os gráficos apresentam o mesmo comportamento, que se assemelha com o comportamento da CSF do olho humano apresentado na Figura 2.21. Esses gráficos apresentam um pico no ponto de visualização ótima. No caso dos gráficos do SQF, o pico se encontra na melhor distância de visualização apresentado no eixo x do gráfico.



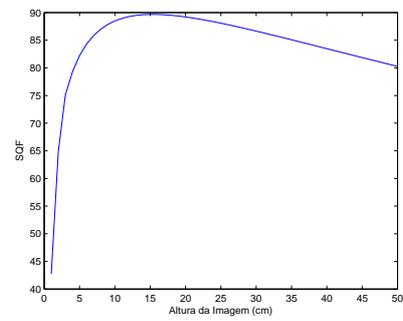
(a) sequência *Cactus*



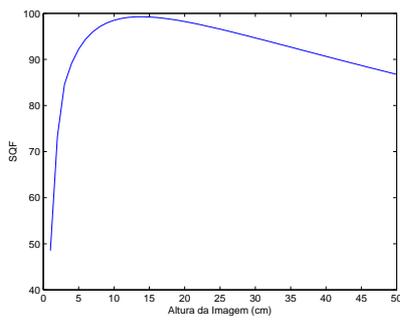
(b) sequência *Stockholm*



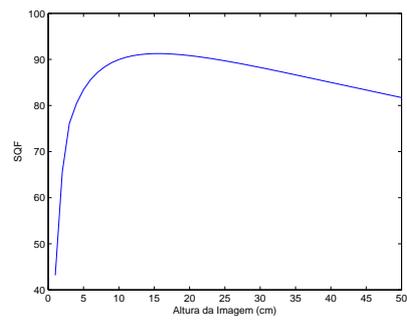
(c) sequência *Crowd run*



(d) sequência *BQ Terrace*



(e) sequência *Park Joy*



(f) sequência *MobCal*

Figura 4.8: Fator de Qualidade Subjetivo - SQF

Tabela 4.4: Resultados da Função de Qualidade

Código da Sequência	Nome da Sequência	Codificador e taxa de bits (Mbps) ou de perda de pacotes (%)	SSIM	VQM	PSNR	FQ
W12	BQ Terrace	H.264 @ 0.3Mbps	0.9946	0.9870	48.9470	0.9977
W12	BQ Terrace	MPEG2 @ 10M	0.9717	0.9447	40.6079	0.6756
W12	BQ Terrace	MPEG2 @ 1M	0.8754	0.6336	32.4961	0.3734
W12	BQ Terrace	H.264 @ 0.3M	0.8711	0.5927	31.6951	0.3010
W8	Crowd run	H.264 @ 0.3M	0.9025	0.3004	30.5056	0.2510
W8	Crowd run	H.264 @ 10.00%	0.8749	0.4307	33.6848	0.3071
W8	Crowd run	H.264 @ 1.00%	0.9477	0.6281	37.1515	0.4729
W8	Crowd run	H.264 @ 30M	0.8267	0.2345	30.8936	0.2718
W10	Cactus	H.264 @ 10.00%	0.9806	0.9408	43.3981	0.7022
W10	Cactus	MPEG2 @ 1M	0.9943	0.9858	49.4828	0.9142
W10	Cactus	H.264 @ 1.00%	0.6918	0.2186	25.7758	0.0650
W10	Cactus	H.264 @ 0.3M	0.8485	0.2876	27.0725	0.0918
W10	Cactus	MPEG2 @ 10M	0.9321	0.6813	34.1506	0.4378
W10	Cactus	H.264 @ 30M	0.9915	0.9797	45.1649	0.8584
W1	Park joy	MPEG2 @ 1M	0.7927	0.5198	28.8922	0.3474
W1	Park joy	MPEG2 @ 10M	0.8924	0.7957	32.7158	0.4943
W1	Park joy	H.264 @ 30M	0.9661	0.9774	42.2440	0.6455
W1	Park joy	H.264 @ 0.3M	0.5959	0.1655	24.0654	0.0145
W6	Stockholm	H.264 @ 1.00%	0.9429	0.6826	36.1772	0.4086
W6	Stockholm	H.264 @ 30M	0.9940	0.9850	49.0531	0.9806
W6	Stockholm	H.264 @ 10.00%	0.8976	0.3393	31.7971	0.3785
W6	Stockholm	H.264 @ 0.3M	0.8902	0.5811	33.4170	0.2676
W6	Stockholm	MPEG2 @ 1M	0.8793	0.5817	33.4864	0.3648
W6	Stockholm	MPEG2 @ 10M	0.9774	0.9524	42.6703	0.7537
W3	Mobcal	H.264 @ 10.00%	0.8958	0.4532	31.8511	0.3294
W3	Mobcal	H.264 @ 1.00%	0.9349	0.7369	35.2829	0.5536
W3	Mobcal	H.264 @ 0.3M	0.8561	0.5762	31.1572	0.3244
W3	Mobcal	MPEG2 @ 1M	0.8457	0.5752	31.5596	0.4291
W3	Mobcal	MPEG2 @ 10M	0.9577	0.9031	38.1551	0.7033
W3	Mobcal	H.264 @ 30M	0.9909	0.9809	45.7954	0.8447

4.6 Testes subjetivos e Métricas de referência

Os gráficos apresentados nas Figuras 4.9, 4.10 e 4.11 desta seção mostram a correlação linear entre os testes subjetivos realizados nesta pesquisa e os valores dos resultados de sistemas de referência escolhidos para a comparação. O resultado do coeficiente de correlação está apresentado na parte superior de cada gráfico e a reta sobreposta aos valores representa a função linear com este coeficiente de correlação.

É importante observar que os sistemas de referência utilizados para comparação utilizam métricas de qualidade utilizando referência completa (FR). Na bibliografia descrita na seção 2.2.2 e Tabela 2.2, as métricas FR pelo fato de utilizarem a sequência de vídeo sem degradações para a comparação com a sequência degradada apresentam coeficientes de correlação entre a métrica e os testes subjetivos com valores maiores que 90%, estes valores indicam uma ótima correlação. As métricas sem referência (NR), como a desenvolvida nesta pesquisa, apresentam coeficientes de correlação da ordem de 80%, que neste caso indicam correlação ótima.

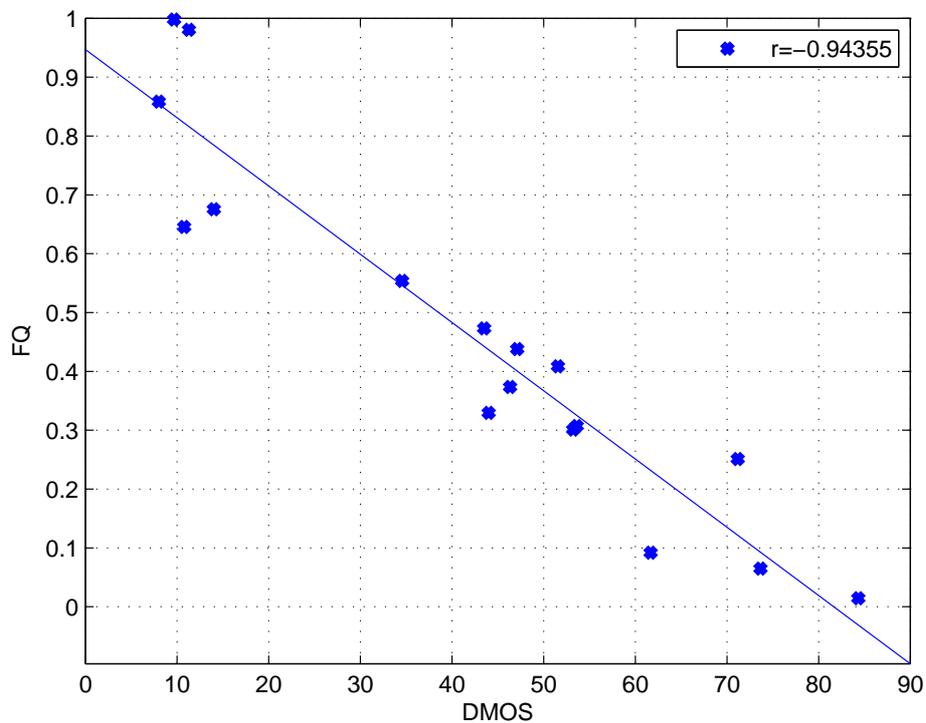


Figura 4.9: Correlação entre o teste subjetivo e a métrica proposta

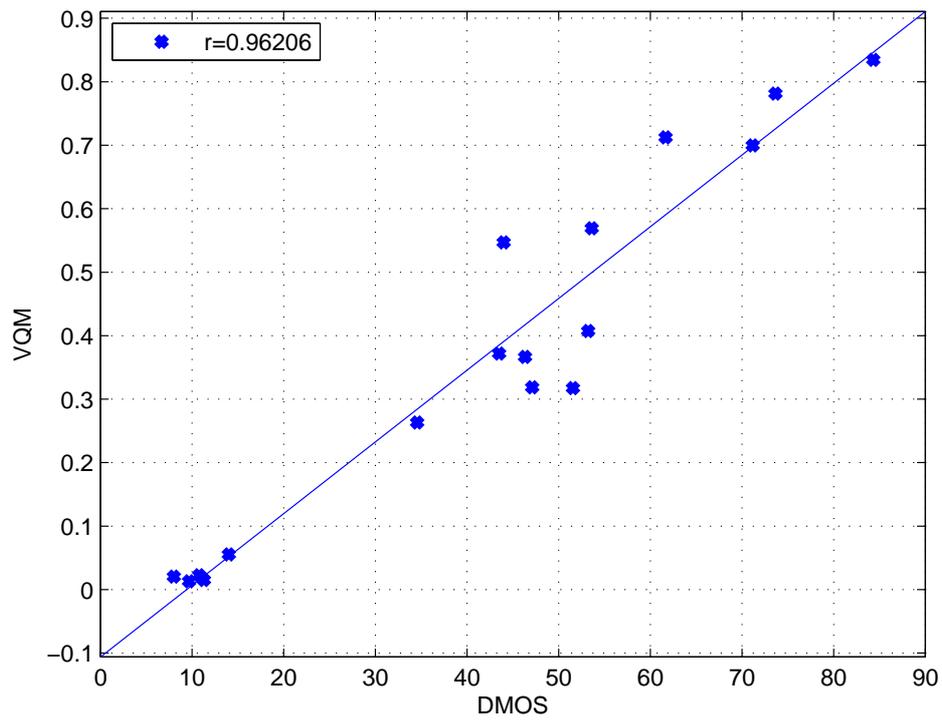


Figura 4.10: Correlação entre o teste subjetivo e a métrica de referência VQM

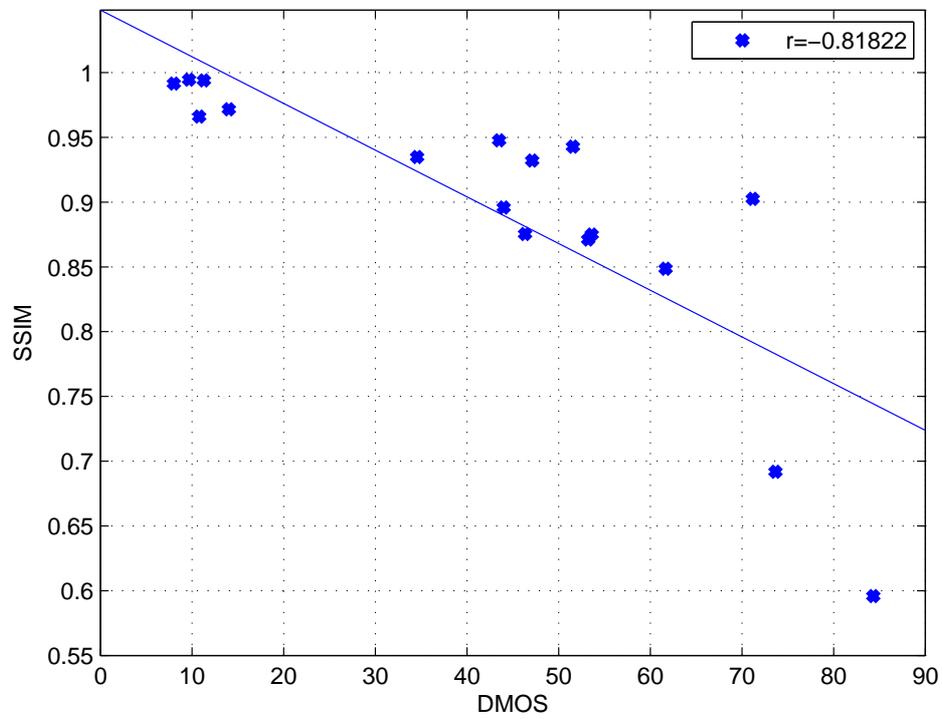


Figura 4.11: Correlação entre o teste subjetivo e a métrica de referência SSIM

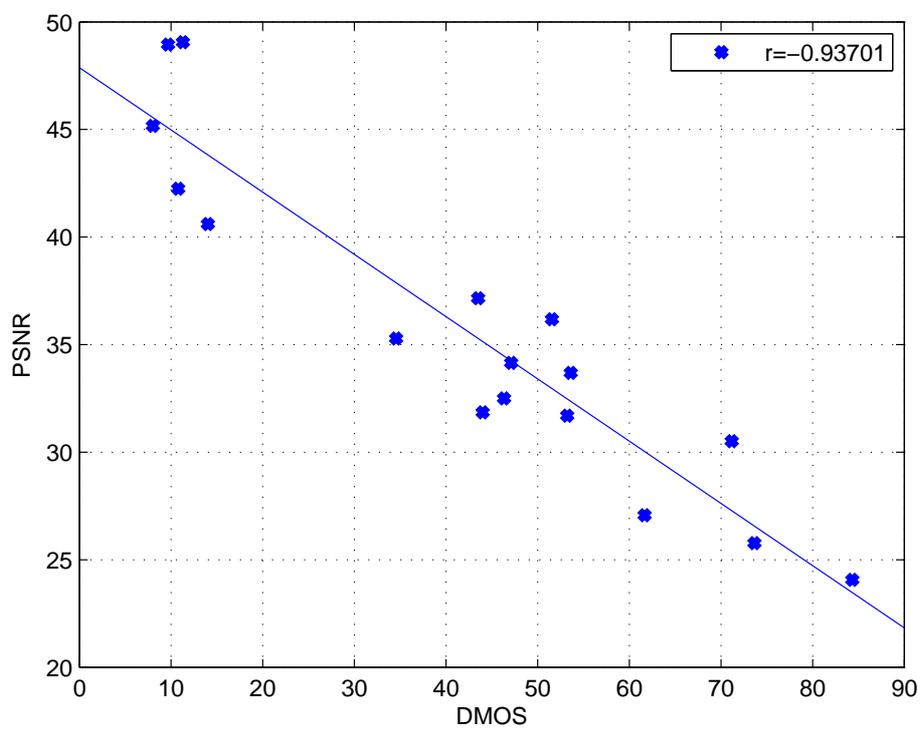


Figura 4.12: Correlação entre o teste subjetivo e a métrica de referência PSNR

4.7 Comparação das Métricas de Qualidade

A Tabela 4.4 apresentada na seção 4.5, mostra os valores para as métricas de qualidade para as sequências de vídeo utilizadas durante os testes com esta metodologia.

Os gráficos das Figuras 4.13, 4.14 e 4.15 apresentam a correlação linear entre os resultados da métrica de qualidade de vídeo, desenvolvida com esta metodologia, e as métricas padronizadas VQM, SSIM e PSNR respectivamente. Na legenda em destaque em cada gráfico é apresentado o valor r da correlação linear entre as métricas. Cada ponto desses gráficos representa um par relacionando a qualidade medida pelas métricas VQM x FQ, SSIM x FQ e PSNR x FQ respectivamente, onde um ponto representa a qualidade para o mesmo vídeo, conforme as linhas da Tabela 4.4. A linha sólida cortando cada gráfico mostra a função de correlação que minimiza o erro entre os pontos.

A métrica FQ foi normalizada para apresentar valores entre 0 e 1, da mesma forma que a métrica VQM e SSIM, onde valores próximos ao valor 0 indicam baixa qualidade de vídeo, e próximos ao valor 1 indicam excelente qualidade.

A distribuição dos pontos no gráfico é compatível com as degradações utilizadas (Tabela 4.4). Os vídeos codificados com taxas de bit maiores, da ordem de 10 Mbps ou 30 Mbps, apresentam melhor qualidade, enquanto aqueles codificados com taxas de bit menores ou com perda de pacotes apresentam qualidade inferior.

A comparação da métrica FQ com a SSIM apresentou o menor valor de correlação, mesmo assim indicando convergência entre as métricas, com correlação maior que 82%.

Na figura 4.13 é apresentado o resultado da correlação entre a métrica proposta e os resultados das métricas de referência, VQM [23][144] (video quality metric). Os pontos representam os resultados para os vídeos utilizados 4.4, sendo que para cada vídeo, no eixo horizontal temos os valores obtidos pela FQ proposta e no eixo vertical os valores obtidos pelo VQM.

Podemos observar pelo gráfico uma alta correlação entre as duas formas de medida de qualidade. A correlação linear entre estas medidas é de 0,91.

Esta correlação indica o desempenho da metodologia proposta se considerarmos que o sistema VQM também tenha um bom desempenho para a verificação da qualidade

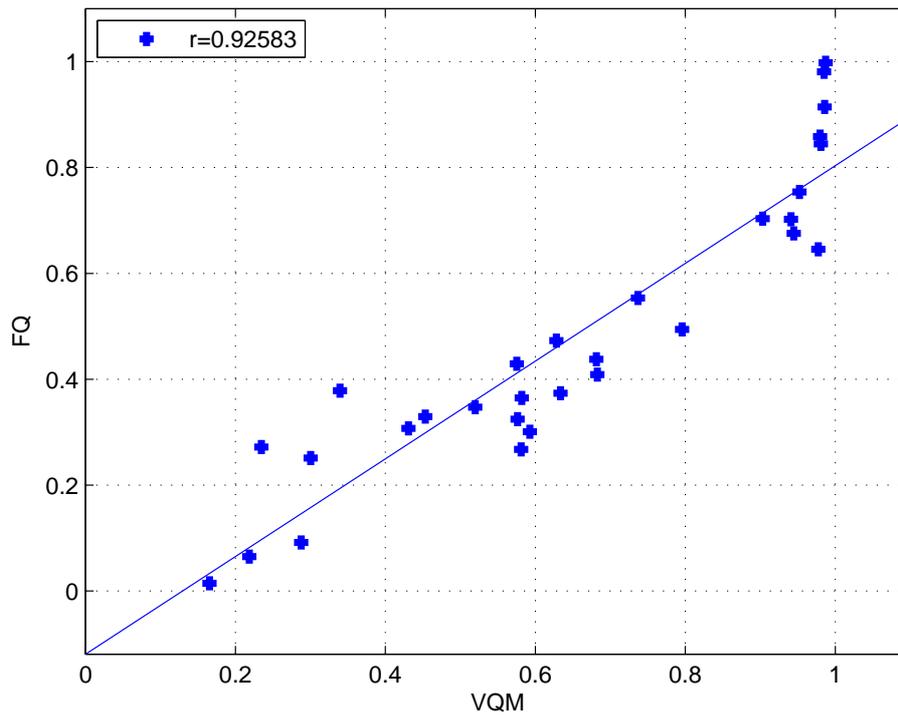


Figura 4.13: Correlação entre a métrica proposta e o sistema de Referência VQM

dos vídeos. A métrica VQM utiliza vídeos com referência completa para calcular a qualidade, enquanto a FQ não utiliza referência.

Considerando os valores de correlação apresentados nos gráficos das Figuras 4.13, 4.14 e 4.15 podemos observar que a métrica FQ apresentada nesta pesquisa possui uma ótima correlação com os testes subjetivos realizados além de ser possível comparar com a ordem de grandeza dos coeficientes de correlação das métricas FR.

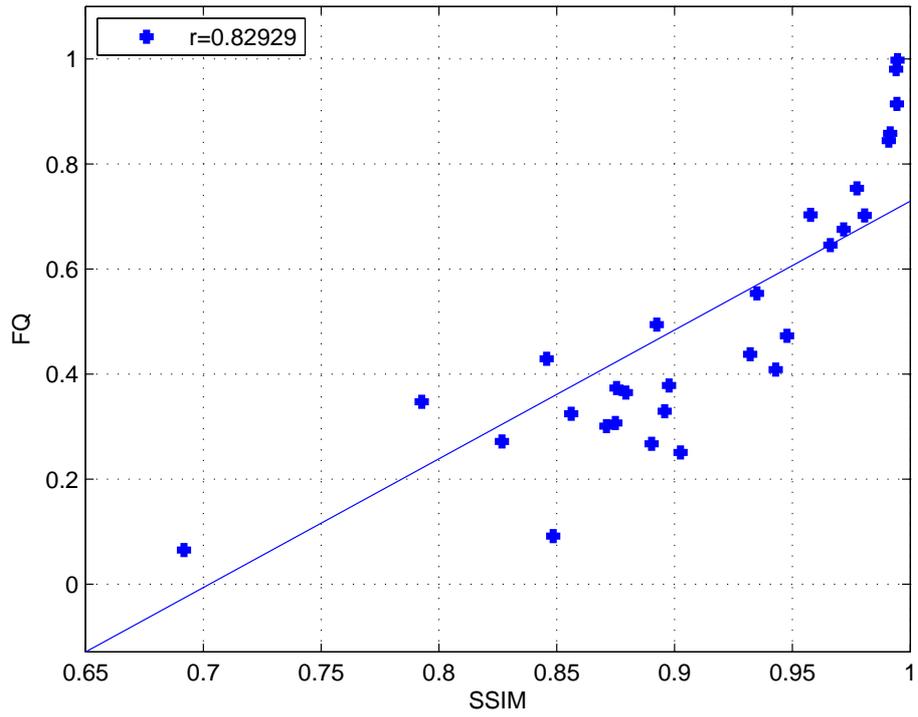


Figura 4.14: Correlação entre a métrica proposta e a Referência SSIM

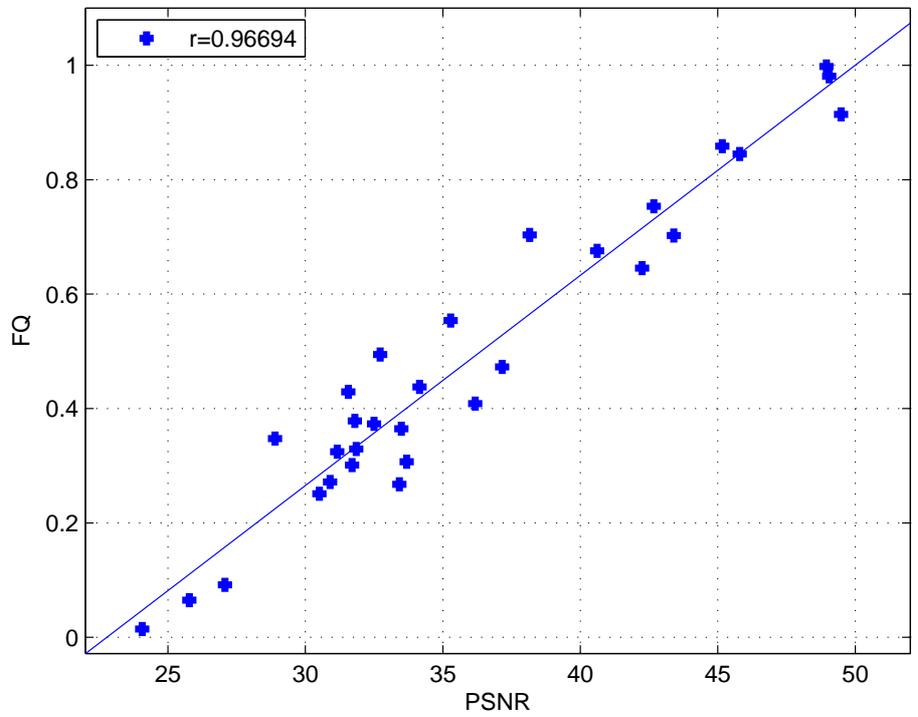


Figura 4.15: Correlação entre a métrica proposta e a Referência PSNR

5 Conclusões

O principal objetivo deste trabalho é o desenvolvimento de uma metodologia para medição de qualidade de vídeo, de forma objetiva em sistemas de distribuição em massa de conteúdos audiovisuais, automaticamente, no ambiente do usuário, de forma natural e com o mínimo de interferência do usuário. A metodologia desenvolvida permite medir a qualidade de vídeo independente do tipo da rede, do tipo do vídeo e do tipo de dispositivo terminal do usuário podendo ser aplicada nos mais diferentes cenários encontrados em países em desenvolvimento como o Brasil. A proposta é de que os pontos de medição estejam próximos aos usuários, com um comportamento semelhante ao dos usuários, ou seja, visualizando o monitor. Os dispositivos para a medição da qualidade devem ser simples e encontrados facilmente no mercado e ao mesmo tempo não devem interferir na rede do distribuidor de vídeos, evitando aumento de custos.

Com esta metodologia e dispositivos será possível aos consumidores realizar a fiscalização da qualidade dos vídeos em tempo real, alertando aos interessados e melhorando a qualidade do sistema como um todo.

O projeto da métrica leva em conta a correlação entre dados subjetivos coletados por usuários, que avaliam a qualidade de sequências de vídeo e servem para avaliar o resultado do processamento da métrica.

A métrica objetiva avalia a qualidade de forma independente de observadores humanos. A construção da métrica prevê que seja independente do tipo de codificação e da rede do distribuidor de vídeos, possua tempo de resposta rápido que seja parametrizável, de forma a se adaptar ao ambiente dos usuários.

A métrica apresentada utiliza características de marcas d'água digitais e dados escondidos (*data hiding*) com propriedades de invariância no tempo, tornando a detecção mais robusta aos problemas de alinhamento e falta de sincronismo.

A métrica, além das características da marca, utiliza também as características ópticas

do ambiente do usuário, que pode apresentar distorções geométricas entre o dispositivo terminal do usuário (monitor de vídeo) e o dispositivo de medidas de qualidade (câmera de vídeo).

Outro aspecto importante é a calibração do sistema no ambiente do usuário, o que permite que o sistema se adapte à grande variedade de situações encontradas na prática.

A metodologia desenvolvida permite que tenhamos uma métrica de qualidade de vídeo objetiva sem referência para avaliações em tempo real em sistemas de distribuição de vídeo.

A utilização dos parâmetros ópticos do ambiente do usuário permite a utilização em situações semelhantes aquelas que os usuários tem em seu dia a dia. Estes parâmetros são independentes das tecnologias de distribuição de vídeo, que podem ser digitais ou analógicas, são independentes dos equipamentos terminais da rede ou do usuário.

Os resultados apresentados no capítulo 4 demonstram o atendimento dos objetivos deste trabalho. Os resultados fornecidos pela métrica de qualidade foram comparados com os resultados dos testes subjetivos e com duas importantes métricas de qualidade de vídeo com referência completa. As correlações lineares entre estas métricas foi maior que 80% indicando a convergência dos resultados.

Pode-se observar dos resultados iniciais que a metodologia utilizando esta abordagem sem referência para testes e utilizando equipamentos de baixo custo como medidores de qualidade, pode apresentar qualidade comparável a de sistemas de referência completa.

5.1 Contribuições

Como principais contribuições deste trabalho, tem-se:

- uma metodologia para verificação de qualidade de vídeos em sistemas de distribuição em massa, sem utilização de referência;
- uma forma de disseminar os medidores de qualidade para uma situação próxima a do usuário, fazendo medidas com o foco da percepção pelo usuário;

- uma proposta de sistema de verificação de qualidade de baixo custo utilizando câmeras de vídeo;
- uma forma de medida de qualidade de vídeos desacoplada e não sincronizada com a rede de distribuição, podendo ser utilizada nas mais diversas tecnologias de distribuição de vídeo;
- uma forma de decodificação de marcas d'água digitais utilizando câmeras de vídeo;
- uma forma de verificar a qualidade baseada em parâmetros de contraste do sistema óptico de forma semelhante ao sistema visual humano.

5.2 Próximos Passos

As técnicas de processamento de marcas d'água digitais apresentaram grande evolução nos últimos anos. A inserção e recuperação das marcas podem ser feitas utilizando múltiplas transformadas no domínio da frequência e os coeficientes destas transformadas podem passar por vários processamentos distintos. O acompanhamento dessa evolução tecnológica apresenta várias soluções promissoras para o aprimoramento da metodologia desenvolvida neste trabalho. Algumas inovações recentes não foram testadas na implementação deste trabalho e merecem atenção. Novas técnicas para a recuperação de marcas d'água digitais invariantes às distorções geométricas apresentam promissores resultados quanto a eficiência e robustez na detecção da marca, sem necessitar de sincronismo entre as marcas inserida e recuperada, da mesma forma que a utilizada neste trabalho. A possibilidade de efetuar uma quantidade menor de leituras da marca em uma quantidade menor de quadros, sem a necessidade de obter diversas medidas e cálculo de suas médias, pode significar um aumento do desempenho da métrica. Dessa forma, como próximos passos para este trabalho, podem ser investigados:

- Utilização do domínio da Transformada Quaternion Fourier [145][146] e a utilização de polinômios de Krawtchouk para a definição dos coeficientes das transformadas no domínio da frequência [147];
- Utilização de novas técnicas para a detecção de áreas de importância, com detecção de bordas, para a inserção e posterior recuperação das marcas [119][148];

- Utilização de técnicas para a detecção das marcas d'água digitais envolvendo a inserção de múltiplas marcas em diferentes domínios de transformadas. Uma marca mais resistente às distorções geométricas pode ser utilizada para que após sua recuperação, uma pequena quantidade de dados recuperados possa ser utilizada para sincronizar outras marcas que conterão os dados úteis a serem trabalhados [149][150];
- Utilização de uma maior quantidade de dados parametrizáveis no ambiente de visualização do usuário, de forma que a métrica de qualidade consiga atingir com maior precisão um maior número possível de casos. Como exemplo de parâmetros podemos citar o nível de iluminação e detecção automática de tamanhos dos monitores de vídeo.

Além disso, melhorias podem ser implementadas para tornar o sistema de medidas de qualidade mais compacto e com menor custo para atingir um número maior de usuários, como:

- Projeto de aplicação protótipo rodando de forma autônoma em um dispositivo móvel do tipo *smartphone* nos novos sistemas operacionais destes dispositivos;
- Projeto de sistema utilizando o SoC em desenvolvimento no projeto INCT-Namitec, utilizando o sensor APS e processador reconfigurável embarcado com a aplicação implementando a metodologia e a função de qualidade objetiva.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ANSI,. ANSI T1.801.03-2003(R2008), Digital Transport of One-Way Video Signals - Parameters for Objective Performance Assessment. *American National Standard Institute*, 2008.
- [2] Pereira, F., Ferreira, A., Salema, C., Trancoso, I., Correia, P., Assunção, P., e Faria, S. Comunicações audiovisuais: Tecnologias, normas e aplicações, 2009.
- [3] Wu, Y., Hirakawa, S., Reimers, U. H., e Whitaker, J. Overview of digital television development worldwide. *Proceedings of the IEEE*, 94(1):8–21, 2006.
- [4] Farias, M. C., Carvalho, M. M., e Alencar, M. S. Digital television broadcasting in brazil. *Multimedia, IEEE*, 15(2):64–70, 2008.
- [5] Hart, J. A. Television, technology, and competition: Hdtv and digital tv in the united states, western europe, and japan. *New York Cambridge University Press.*, 2004.
- [6] Zuffo, M. K., Carvalho, E., Barros, G., COSTA, L., FARIA, R., NUNES, R., e LOPES, R. The brazilian digital television system access device architecture. *Journal of the Brazilian Computing Society*, 13:19, 2007.
- [7] Cordero, C. e Rowello, R. Optimizing Customer Experience through Operational Excellence. *Cisco Internet Business Solutions Group (IBSG)*, 2009.
- [8] TCU,. Relatório de Auditoria de Natureza Operacional; Auditoria de natureza operacional, Avaliação da atuação da Anatel no acompanhamento da qualidade da prestação dos serviços de telefonia, determinações e recomendações, monitoramento. acórdão 2109/2006 – plenário – tcu. ac-2109-46/06-p. 2006. *Tribunal de Contas da União*, 2006.
- [9] BRASIL,. Lei nº 9472 de 16 de julho de 1997. *Diário Oficial da União*, 1997.
- [10] BRASIL,. Lei nº 4117 de 27 de agosto de 1962. *Diário Oficial da União*, 1962.

- [11] BRASIL,. Constituição de 1988. *Diário Oficial da União*, 1988.
- [12] BRASIL,. Regulamentação de televisão por assinatura. *Ministério das Comunicações*, 1997.
- [13] BRASIL,. Norma nº 13 de 1996. *Ministério das Comunicações*, 1996.
- [14] BRASIL,. Lei nº 12485 de 12 de setembro de 2011. *Diário Oficial da União*, 2011.
- [15] Anatel,. Resolução nº 581 de 26 de março de 2012, Regulamento do serviço de acesso condicionado (SeAC). *Agência Nacional de Telecomunicações*, 2011.
- [16] Farias, M. C. Visual-quality estimation using objective metrics. *Journal of the Society for Information Display*, 19(11):764–770, 2011.
- [17] Klein, S. A. Image quality and image compression: A psychophysicist’s viewpoint. In *Digital images and human vision*, pages 73–88. MIT Press, 1993.
- [18] Sohn, H., Yoo, H., De Neve, W., Kim, C. S., e Ro, Y. M. Full-reference video quality metric for fully scalable and mobile svc content. *Broadcasting, IEEE Transactions on*, 56(3):269–280, 2010.
- [19] ITU-T,. Recommendation P.910, Subjective video quality assessment methods for multimedia applications. *International Telecommunication Union, Geneva, Switzerland*, 1996.
- [20] ITU-R,. Recommendation BT.601-7, Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. *International Telecommunication Union, Geneva, Switzerland*, 2011.
- [21] ITU-R,. Recommendation BT.500-12. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union, Geneva, Switzerland*, 2009.
- [22] Lambrecht, C. J. e Kunt, M. Characterization of human visual sensitivity for video imaging applications. *Signal Processing*, 67(3):255–269, 1998.
- [23] Wolf, S. e Pinson, M. Video quality measurement techniques. *NTIA Report 02-392*, 2002.
- [24] Daly, S. J. Visible differences predictor: an algorithm for the assessment of image fidelity. In *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pages 2–15. International Society for Optics and Photonics, 1992.

- [25] Lubin, J. A human vision system model for objective picture quality measurements. *IET*, 1997.
- [26] Yeh, E. M., Kokaram, A. C., e Kingsbury, N. G. Psychovisual measurement and distortion metrics for image sequences. In *Proceedings of European Signal Processing Conference*, volume 2, pages 1061–1064, 1998.
- [27] Watson, A. B. Towards a visual quality metric for digital video. In *Proc. European Signal Processing Conference*, volume 2, 1998.
- [28] VQEG,. Vqeg subjective test plan. *Technical Report, Video Quality Experts Group*, 1999.
- [29] Winkler, S. Issues in vision modeling for perceptual video quality assessment. *Signal Processing*, 78(2):231–252, 1999.
- [30] ANSI,. ANSI T1.801.02-1996(R2006), Digital Transport of Video Teleconferencing/Video Telephony Signals - Performance Terms, Definitions and Examples. *American National Standard Institute*, 2006.
- [31] Takahashi, A., Hands, D., e Barriac, V. Standardization activities in the itu for a qoe assessment of iptv. *Communications Magazine, IEEE*, 46(2):78–84, 2008.
- [32] ITU-T,. Recommendation J.144, Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference. *International Telecommunication Union, Geneva, Switzerland*, 2001.
- [33] ITU-T,. Recommendation J.143, User requirements for objective perceptual video quality measurements in digital cable television. *International Telecommunication Union, Geneva, Switzerland*, 2000.
- [34] Winkler, S. *Digital video quality: vision models and metrics*. Wiley, 2005.
- [35] SCHADE, S. e others,. Optical and photoelectric analog of the eye. *JOsA*, 46(9):721–738, 1956.
- [36] Mannos, J. e Sakrison, D. The effects of a visual fidelity criterion of the encoding of images. *Information Theory, IEEE Transactions on*, 20(4):525–536, 1974.
- [37] Lukas, F. e Budrikis, Z. Picture quality prediction based on a visual model. *Communications, IEEE Transactions on*, 30(7):1679–1692, 1982.

- [38] Tong, X., Heeger, D. J., e Branden Lambrecht, C. J. Van den . Video quality evaluation using st-cielab. In *Electronic Imaging'99*, pages 185–196. International Society for Optics and Photonics, 1999.
- [39] Zhang, X. e Wandell, B. A spatial extension of cielab to predict the discriminability of colored patterns. In *SID Symposium Digest*, volume 27, pages 731–735, 1996.
- [40] Watson, A. B. The cortex transform: rapid computation of simulated neural images. *Computer vision, graphics, and image processing*, 39(3):311–327, 1987.
- [41] Watson, A. B. e Ahumada Jr, A. J. A hexagonal orthogonal-oriented pyramid as a model of image representation in visual cortex. *Biomedical Engineering, IEEE Transactions on*, 36(1):97–106, 1989.
- [42] Daly, S. J. Visible differences predictor: an algorithm for the assessment of image fidelity. In *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pages 2–15. International Society for Optics and Photonics, 1992.
- [43] Lubin, J. A visual discrimination model for imaging system design and evaluation. *Vision models for target detection and recognition*, 2:245–357, 1995.
- [44] Lubin, J. e Fibush, D. Sarnoff jnd vision model, 1997.
- [45] Teo, P. C. e Heeger, D. J. Perceptual image distortion. In *Proceedings of SPIE*, volume 2179, page 127, 1994.
- [46] Teo, P. C. e Heeger, D. J. Perceptual image distortion. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, volume 2, pages 982–986. IEEE, 1994.
- [47] Lambrecht, C. Perceptual models and architectures for video coding applications. *Swiss Federal Institute of Technology/Lausanne*, 1996.
- [48] Masry, M. A. e Hemami, S. S. A metric for continuous quality evaluation of compressed video with severe distortions. *Signal processing: Image communication*, 19(2):133–146, 2004.
- [49] Watson, A. B. Image data compression having minimum perceptual error, June 20 de 1995. US Patent 5,426,512.
- [50] Watson, A. B. Image data compression having minimum perceptual error, May 13 de 1997. US Patent 5,629,780.

- [51] Watson, A. B. Toward a perceptual video-quality metric. In *Photonics West'98 Electronic Imaging*, pages 139–147. International Society for Optics and Photonics, 1998.
- [52] Pinson, M. e Wolf, S. A new standardized method for objectively measuring video quality. *IEEE Trans. Broadcast.* vol. 50. no, 50(3):312–322, September de 2004.
- [53] Tan, K., Ghanbari, M., e Pearson, D. E. An objective measurement tool for mpeg video quality. *Signal Processing*, 70(3):279–294, 1998.
- [54] Wang, Z., Lu, L., e Bovik, A. Video quality assessment based on structural distortion measurement. *Signal Process. Image Commun.*, 19(2):121–132, February de 2004.
- [55] Wang, Z., Bovik, A., Sheikh, H., e Simoncelli, E. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* vol. 13. no, 13(4):600–612, april de 2004.
- [56] Wang, Z., Simoncelli, E. P., e Bovik, A. C. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. IEEE, 2003.
- [57] Wang, Z. e Li, Q. Video quality assessment using a statistical model of human visual speed perception. *J. Opti. Soc. America A (Optics. Image Sci. Vision)*, 24(12):B61–B69, December de 2007.
- [58] Stocker, A. A. e Simoncelli, E. P. Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, 9(4):578–585, 2006.
- [59] Sheikh, H. e Bovik, A. Image information and visual quality. *IEEE Trans. Image Process.*, 15(2):430–444, February de 2006.
- [60] Lu, L., Wang, Z., Bovik, A. C., e Kouloheris, J. Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of mpeg video. In *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, volume 1, pages 61–64. IEEE, 2002.
- [61] Shnayderman, A., Gusev, A., e Eskicioglu, A. Multidimensional image quality measure using singular value decomposition. 5294(1):82–92, 2003.

- [62] Tao, P. e Eskicioglu, A. M. Video quality assesment using m-svd. In *Electronic Imaging 2007*, pages 649408–649408. International Society for Optics and Photonics, 2007.
- [63] Pessoa, A., Falcao, A., Nishihara, R., Silva, A., e Lotufo, R. Video quality assessment using objective parameters based on image segmentation. *Soc. Motion Pictures Television Eng. (SMPTE) J.*, 108(12):865–872, December de 1999.
- [64] Okamoto, J., Hayashi, T., Takahashi, A., e Kurita, T. Proposal for an objective video quality assessment method that takes temporal and spatial information into consideration. *Electronics and Communications in Japan (Part I: Communications)*, 89(12):97–108, 2006.
- [65] Lee, S.-O. e Sim, D.-G. New full-reference visual quality assessment based on human visual perception. In *Consumer Electronics, 2008. ICCE 2008. Digest of Technical Papers. International Conference on*, pages 1–2. IEEE, 2008.
- [66] Bhat, A., Richardson, I., e Kannangara, S. A new perceptual quality metric for compressed video. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 933–936. IEEE, 2009.
- [67] Branden Lambrecht, C. van den e Verscheure, O. Perceptual quality measure using a spatio-temporal model of the human visual system. *Proc. Int. Soc. Opt. Eng. (SPIE).*, 2668:450–461, 1996.
- [68] Xiao, F. e others,. Dct-based video quality evaluation. *Final Project for EE392J*, page 769, 2000.
- [69] Watson, A., Hu, J., e McGowan, J. Digital video quality metric based on human vision. *J. Electron. Imaging. vol. 10. no*, 10(1):20–29, January de 2001.
- [70] Lee, C. e Kwon, O. Objective measurements of video quality using the wavelet transform. *Optical Eng. vol. 42. no*, 42(1.):265–272, January de 2003.
- [71] Seshadrinathan, K. e Bovik, A. C. Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Trans. Image Process. vol. 19. no*, 19(2.):335–350, February de 2010.
- [72] Born, R. T. e Bradley, D. C. Structure and function of visual area mt. *Annu. Rev. Neurosci.*, 28:157–189, 2005.

- [73] Hekstra, A., Beerends, J., Ledermann, D., Caluwe, F.de , Kohler, S., Koenen, R., Rihs, S., Ehram, M., e Schlauss, D. Pvqm—a perceptual video quality measure. *Signal Process. Image Commun. vol. 17. no*, 17(10.):781–798, November de 2002.
- [74] Lu, Z., Lin, W., Ong, E., Yang, X., e Yao, S. Psqm-based rr and nr video quality metrics. In *Visual Communications and Image Processing 2003*, pages 633–640. International Society for Optics and Photonics, 2003.
- [75] Ong, E., Yang, X., Lin, W., Lu, Z., e Yao, S. Video quality metric for low bi-trate compressed videos. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 5, pages 3531–3534. IEEE, 2004.
- [76] Ong, E., Lin, W., Lu, Z., e Yao, S. Colour perceptual video quality metric. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages III–1172. IEEE, 2005.
- [77] Ndjiki-Nya, P., Barrado, M., e Wiegand, T. Efficient full-reference assessment of image and video quality. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 2, pages II–125. IEEE, 2007.
- [78] Chandler, D. e Hemami, S. Vsnr: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Trans. Image Process. vol. 16. no*, 16(9.):2284–2298, September de 2007.
- [79] ITU-T,. Recommendation J.247, Objective perceptual multimedia video quality measurement in the presence of a full reference. *International Telecommunication Union, Geneva, Switzerland*, 2008.
- [80] VQEG,. Report on the validation of video quality models for high definition video content, 2010.
- [81] Wolf, S. e Pinson, M. Application of the ntia general video quality metric (vqm) to hdtv quality monitoring. In *Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-07), Scottsdale, Arizona*, volume 56, 2007.
- [82] Sugimoto, O., Naito, S., Sakazawa, S., e Koike, A. Objective perceptual picture quality measurement method for high-definition video based on full reference framework. In *IS&T/SPIE Electronic Imaging*, pages 72421A–72421A. International Society for Optics and Photonics, 2009.

- [83] Okamoto, J., Watanabe, K., Honda, A., Uchida, M., e Hangai, S. Hdtv objective video quality assessment method applying fuzzy measure. In *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pages 168–173. IEEE, 2009.
- [84] ITU-T,. Recommendation J.341, Objective perceptual multimedia video quality measurement of hdtv for digital cable television in the presence of a full reference. *International Telecommunication Union, Geneva, Switzerland*, 2011.
- [85] Ferguson, K. An adaptable human vision model for subjective video quality rating prediction among cif, sd, hd and e-cinema. In *Proceedings of the ird International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM*, 2007.
- [86] Farias, M. C., Carli, M., Neri, A., e Mitra, S. K. Video quality assessment based on data hiding driven by optical flow information. In *Electronic Imaging 2004*, pages 190–200. International Society for Optics and Photonics, 2003.
- [87] Farias, M. C. e Mitra, S. K. Perceptual contributions of blocky, blurry, noisy, and ringing synthetic artifacts to overall annoyance. *Journal of Electronic Imaging*, 21(4):043013–043013, 2012.
- [88] De Simone, F., Goldmann, L., Lee, J.-S., e Ebrahimi, T. Towards high efficiency video coding: Subjective evaluation of potential coding technologies. *Journal of Visual Communication and Image Representation*, 22(8):734–748, 2011.
- [89] CDVL,. The consumer digital video library. 2013.
- [90] HDFORUM,. The official hd video clip list. 2013.
- [91] VQEG,. Final report from the video quality experts group on the validation of objective models of video quality assessment. *Video Quality Experts Group*, 2000.
- [92] 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*, 2008.
- [93] Snedecor, G. W. Statistica methods. *The Iowa State University Press*, 1989.
- [94] Anderson, R. J. e Petitcolas, F. A. On the limits of steganography. *Selected Areas in Communications, IEEE Journal on*, 16(4):474–481, 1998.
- [95] Hartung, F. e Kutter, M. Multimedia watermarking techniques. *Proceedings of the IEEE*, 87(7):1079–1107, 1999.

- [96] Kahn, D. The history of steganography. In *Information Hiding*, pages 1–5. Springer, 1996.
- [97] Sencar, H. T., Ramkumar, M., e Akansu, A. N. *Data hiding fundamentals and applications: content security in digital multimedia*. Academic Press, 2004.
- [98] Bender, W., Gruhl, D., Morimoto, N., e Lu, A. Techniques for data hiding. *IBM systems journal*, 35(3.4):313–336, 1996.
- [99] Petitcolas, F. A., Anderson, R. J., e Kuhn, M. G. Information hiding-a survey. *Proceedings of the IEEE*, 87(7):1062–1078, 1999.
- [100] UWMedicine,. Treatment of recurrent instability. *University of Washington*, February de 2013.
- [101] Rey, C. e Dugelay, J.-L. A survey of watermarking algorithms for image authentication. *EURASIP Journal on Applied Signal Processing*, 2002(1):613–621, 2002.
- [102] Yang, Z. Dual domain semi-fragile watermarking for image authentication. *University of Toronto, Canada*, 2003.
- [103] Barni, M. e Bartolini, F. *Watermarking systems engineering: enabling digital assets security and other applications*, volume 21. CRC Press, 2004.
- [104] Del Colle, F. A. e Gómez, J. C. Dwt based digital watermarking fidelity and robustness evaluation. *Journal of Computer Science & Technology*, 8, 2008.
- [105] N.Corina,. Filigranage dans le domaine des ondelettes Watermarking in the wavelet domain. *Luniversité Politehnica Timisoara Faculté dElectronique et Télécommunications*, 2004.
- [106] Cox, I., Miller, M., Bloom, J., e Miller, M. *Digital watermarking*. Morgan Kaufmann, 2001.
- [107] Pan, J.-S., Huang, H.-C., e Jain, L. C. *Intelligent watermarking techniques*, volume 7. World Scientific Publishing Company, 2004.
- [108] Khan, A. Intelligent perceptual shaping of a digital watermark. *Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan*, 2006.
- [109] Cong, J. e Peng, J. Affine invariant watermark scheme using genetic algorithm. *Journal of Computer Science*, 4:20–26, 2005.

- [110] Bi, H., Zhang, Y., e Li, X. Video watermarking robust against spatio-temporal attacks. *Journal of Networks*, 6(6):932–936, 2011.
- [111] Al-Temeemy, A. A. e Spencer, J. Invariant spatial chromatic processors for region image description. In *Imaging Systems and Techniques (IST), 2010 IEEE International Conference on*, pages 421–425. IEEE, 2010.
- [112] Li, L. e Guo, B. Image adaptive rst invariant watermark using pseudo-zernike moments. *Frontiers of Electrical and Electronic Engineering in China*, 3(1):20–24, 2008.
- [113] Lee, H.-Y., Kim, H., e Lee, H.-K. Robust image watermarking using local invariant features. *Optical Engineering*, 45(3):037002–037002, 2006.
- [114] Li, L., Guo, B., e Pan, J.-S. Robust image watermarking using feature based local invariant regions. *International Journal of Innovative Computing, Information and Control*, 4(8):1977–1986, 2008.
- [115] Hsin, C. e Van Hau, B. Affine invariant local features based on novel keypoint detection and grouping. In *Communications and Electronics (ICCE), 2012 Fourth International Conference on*, pages 296–301. IEEE, 2012.
- [116] Pei, S.-C. e Lin, C.-N. Image normalization for pattern recognition. *Image and Vision Computing*, 13(10):711–723, 1995.
- [117] Harris, C. e Stephens, M. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.
- [118] ZHANG, Y., BI, H., e ZHANG, H. Robust watermarking scheme by harris interest regions. *Journal of Computational Information Systems*, 8(20):8421–8429, 2012.
- [119] Chen, D.-Y. e Chiu, Y.-M. Robust video copy detection based on constrained feature points matching. In *Advances in Intelligent Systems and Applications-Volume 2*, pages 359–368. Springer, 2013.
- [120] Ahmed, N., Natarajan, T., e Rao, K. R. Discrete cosine transform. *Computers, IEEE Transactions on*, 100(1):90–93, 1974.
- [121] Burger, W. e Burge, M. J. *Digital image processing*. Springer, 2008.
- [122] Rossmann, K. Point spread-function, line spread-function, and modulation transfer function tools for the study of imaging systems. *Radiology*, 93(2):257–272, 1969.

- [123] Padgett, R. e Kotre, C. Development and application of programs to measure modulation transfer function, noise power spectrum and detective quantum efficiency. *Radiation protection dosimetry*, 117(1-3):283–287, 2005.
- [124] Buhr, E., Günther-Kohfahl, S., e Neitzel, U. Accuracy of a simple method for deriving the presampled modulation transfer function of a digital radiographic system from an edge image. *Medical Physics*, 30:2323, 2003.
- [125] Workman, A. e Brettle, D. Physical performance measures of radiographic imaging systems. *Dentomaxillofacial Radiology*, 26(3):139–146, 1997.
- [126] Bushberg, J. T. *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.
- [127] Epstein, C. L. *Introduction to the mathematics of medical imaging*. Society for Industrial and Applied Mathematics, 2007.
- [128] Huda, W. e Slone, R. M. *Review of radiologic physics*. Lippincott Williams & Wilkins, 2003.
- [129] Wolbarst, A. B. e Cook, G. Physics of radiology. *Appleton & Lange Norwalk Connecticut Connecticut*, 1993.
- [130] Johns, H. E. e others,. *The physics of radiology*. Charles C. Thomas, 1983.
- [131] Barrett, H. H. e Swindell, W. *Radiological imaging: the theory of image formation, detection, and processing*, volume 2. Academic Press, 1996.
- [132] Samei, E., Flynn, M. J., e Reimann, D. A. A method for measuring the pre-sampled mtf of digital radiographic systems using an edge test device. *Medical Physics*, 25:102, 1998.
- [133] Cao, X., Huang, H., Lou, S. L., e others,. A novel algorithm for measuring the mtf of a digital radiographic system with a ccd array detector. In *Proc. SPIE*, volume 3977, pages 580–589, 2000.
- [134] Samei, E. e Flynn, M. J. An experimental comparison of detector performance for computed radiography systems. *Medical physics*, 29:447, 2002.
- [135] Bernd, G. Handout #9: Human visual perception. *course EE368B - Image and Video Compression*, 2001.

- [136] Granger, E. e Cupery, K. An optical merit function (sqf), which correlates with subjective image judgments. *Photographic science and engineering*, 16(3):221–230, 1972.
- [137] Keelan, B. W. Objective and subjective measurement and modeling of image quality: a case study. In *Proc. of SPIE Vol*, volume 7798, pages 779815–1, 2010.
- [138] Farias, M. C. e Mitra, S. K. A methodology for designing no-reference video quality metrics, 2009.
- [139] DeVelis, J. B., PARRENT, J., GEORGE, B., e others,. Transfer function for cascaded optical systems. *JOSA*, 57(12):1486–1487, 1967.
- [140] Goodman, J. *Introduction to Fourier optics*. McGraw-hill, 2008.
- [141] Estribeau, M. e Magnan, P. Fast mtf measurement of cmos imagers using iso 12333 slanted-edge methodology. In *Proceedings of SPIE*, volume 5251, pages 243–252, 2003.
- [142] Anatel,. Resolução nº 242 de 30 de novembro de 2000, Regulamento para certificação e homologação de produtos para telecomunicações. *Agência Nacional de Telecomunicações*, 2011.
- [143] NeuroBS,. Presentation. <http://www.neurobs.com/>, 2013.
- [144] Pinson, M. e Wolf, S. Video quality metric software, Version 2. *NTIA Software. Data Product SD-03-396. Volumes, 1-5*, October de 2002.
- [145] Wang, X., Wang, C., Yang, H., e Niu, P. A robust blind color image watermarking in quaternion fourier transform domain. *Journal of Systems and Software*, 86(2):255–277, 2013.
- [146] Wang, X.-y., Wang, C.-p., Yang, H.-y., e Niu, P.-p. A robust blind color image watermarking in quaternion fourier transform domain. *Journal of Systems and Software*, 2012.
- [147] Asli, B. H. S. e Paramesran, R. Digital filter implementation of orthogonal moments. 2013.
- [148] Yu, F. J., Luan, X., Song, D. L., Li, X. F., e Zhou, H. H. A novel high accuracy sub-pixel corner detection algorithm for camera calibration. *Applied Mechanics and Materials*, 239:713–716, 2013.

- [149] Liu, J., She, K., e Wu, H. Blind image watermarking using dual embedding scheme in the wavelet transform domain. *J. Comput. Inf. Syst*, 6(6):1887–1896, 2010.
- [150] Pramila, A., Keskinarkaus, A., Rahtu, E., e Seppänen, T. Watermark recovery from a dual layer hologram with a digital camera. In *Image Analysis*, pages 146–155. Springer, 2011.

APÊNDICES

A Avaliação de Marcas de Testes para Inserção nos Vídeos

Nesta pesquisa foram avaliados dispositivos como câmeras de vídeo utilizando sensores de imagens e softwares de tratamento de imagens que podem ser embutidos nos decodificadores ou em dispositivos portáteis junto a recepção do usuário.

Uma imagem de testes foi embutida com o vídeo de interesse de forma que os observadores humanos não consigam enxergá-lo de forma natural, porém o dispositivo de testes consegue recuperar estes padrões de testes e verificar a qualidade do vídeo.

O padrão de testes é inserido na geração do vídeo de forma que possa ser testada toda a cadeia de distribuição, desde a codificação, transmissão e posterior decodificação no equipamento do usuário. Com base na análise do padrão de testes são estimados os possíveis fatores de impacto na qualidade deste vídeo.

Na preparação desta metodologia, foram avaliadas algumas abordagens para a inclusão do padrão de testes. A primeira foi o aumento da taxa de amostragem de quadros de vídeo, de forma que o sistema de visão humano não perceba estes quadros extras, onde os padrões de teste foram incluídos. A segunda foi a inclusão de imagens em uma faixa espectral imperceptível ao olho humano. A terceira foi a inclusão de marcas embutidas no vídeo que podem ser decodificadas pelo sistema de testes, porém sem afetar a visualização do vídeo.

Nesta pesquisa é utilizada a terceira abordagem, pois esta se demonstrou mais robusta e ao mesmo tempo apresenta uma forma de implementação mais barata utilizando dispositivos convencionais comumente utilizados pelos usuários de forma a atender aos objetivos propostos.

Metodologia para transmissão de sinais de controle utilizando a área visível do monitor de vídeo, para dispositivos de verificação da qualidade de vídeo que tratarão estes sinais. Os dados de controle são codificados e inseridos subliminarmente no sinal de vídeo, preferencialmente logo após a geração do vídeo, antes de qualquer processamento, de forma que o vídeo original mais o sinal subliminar sejam enviados em uma sequência de campos selecionados do vídeo para os usuários finais que poderão fazer a análise da

qualidade do vídeo decodificando estes sinais utilizando um dispositivo com sensor de imagens.

O processo de inclusão dos sinais de teste subliminares terá o comportamento da modulação de campos de vídeo dentro da área de visualização de uma televisão, tendo cada um, alternadamente, sua luminância levantados e abaixados proporcionalmente em suas linhas de varredura horizontais, monitorados por um dispositivo sensível à luz posicionado em frente ao monitor de televisão do usuário.

Estes dispositivos poderão decodificar o sinal, dentro do alcance da câmera e detectar energia nesta faixa de frequência, decodificando os comandos incorporados nos dados de controle e executar um algoritmo para a medida da qualidade de vídeo baseado em parâmetros dos sensores de imagem em conjunto com a sequência de dados que foi adicionada na geração do vídeo.

O método de codificação de dados envolve um modulador de sinal de vídeo em frequências que estão relacionados com múltiplos e submúltiplos da taxa de linha horizontal, que se comportará como uma subportadora de vídeo de forma imperceptível para os usuários. Embora a modulação esteja dentro da área de visualização da tela da televisão, ela é praticamente invisível para o espectador por causa de sua intensidade relativamente baixa, de sua relação de tempo com a taxa de linha horizontal e da integração e das características inerentes a resolução do olho humano.

Utilizando este sistema de modulação de subportadora, podem ser incluídos vários sinais em uma transmissão de TV, estes sinais podem ser adicionados para aumentar a capacidade de informações a serem transmitidas, facilitando a decodificação pelo equipamento sensor de imagens do lado do receptor do usuário.

O sistema é composto de um codificador para incluir as informações de testes no vídeo de interesse e um decodificador utilizando sensor de imagens que captará a imagem da tela do receptor do usuário.

O codificador é instalado no ambiente de geração dos vídeos a serem monitorados e é responsável por adicionar a estes vídeos os sinais que possibilitam que o decodificador possa medir a qualidade do vídeo no final da cadeia de vídeo.

Os sinais de teste são inseridos no vídeo bruto, para possibilitar que codificações e compressões posteriores possam ser avaliadas.

Nesta pesquisa os sinais de testes são definidos como um padrão de bits que é conhecido pelo decodificador, que manipula estas informações recebidas e define qual a qualidade do vídeo. Pode-se ainda estimar qual o tipo de problema na cadeia de distribuição do vídeo, de forma a auxiliar a resolução dos problemas de qualidade.

O padrão de bits inserido no vídeo original é feito de forma a alterar minimamente a qualidade de vídeo para o usuário.

Na primeira fase da pesquisa foram avaliadas duas formas de inclusão de informações de teste no vídeo de interesse, a primeira utilizando sobre amostragem, inserindo quadros de testes intercalados com os quadros originais do vídeo, estes quadros eram apresentados com uma taxa tal que o observador humano não consegue observar a transição entre os quadros do vídeo e os quadros de testes.

A taxa de amostra do vídeo para que não seja percebida pelo olho humano e como esta taxa afetará o tamanho do arquivo e a qualidade do vídeo. É importante ressaltar que o padrão de teste deverá ser nesse caso uma transição suave em relação ao vídeo original para evitar que quadros de teste entrelaçados com os quadros úteis (*payload*) retirem toda a vantagem das codificações MPEG.

O sistema visual humano tem a capacidade de perceber “*flickering*” de aproximadamente 45 quadros por segundo. Aumentando a taxa de amostragem dos quadros para valores acima deste limite, o olho humano não terá capacidade de perceber os quadros adicionais e esta característica pode ser utilizada para inserir padrões de teste nos quadros amostrados a taxas maiores intercalados aos quadros do vídeo de interesse.

Uma das formas mais simples de verificar a qualidade de vídeo, independente do dispositivo de recepção de vídeo, é através da tela, e podemos utilizar várias técnicas para através da tela do receptor, medirmos a qualidade do vídeo transmitido originalmente. Para a recepção do vídeo diretamente através da tela de visualização, utilizaremos um sensor de imagens.

Técnicas de inclusão de imagens para a certificação de que a imagem original não foi alterada ou não está sendo utilizada para os fins a que ela foi fornecida são muito comuns na literatura, principalmente para proteção contra cópias.

A inclusão de imagens pode ser feita de diversas formas e a análise da que melhor se adapta ao objetivo do trabalho é avaliada.

Para não alterar a percepção dos expectadores técnicas de esteganografia são utilizadas, embutindo imagens de testes não visíveis para os observadores humanos, mas que poderão ser detectadas pelo sistema de testes.

Nesta pesquisa são avaliados essas formas e a escolha da que melhor atende ao objetivo de mínima interferência nos sistemas de distribuição de vídeo, custos menores e possibilidade de espalhar sensores por todos os usuários independentemente do dispositivo monitor de imagens pode ser aplicada de forma que atinja o maior número possível de situações que ocorrem nos mercados de distribuição de vídeo.

A.1 Utilização de Super amostragem

Para este tipo de teste são inseridas imagens de testes com taxa maior que o vídeo de interesse, as taxas utilizadas para o teste são de 100 quadros por segundo para um vídeo original com taxa de 30 quadros por segundo.

Na maior taxa são incluídos quadros em duas situações de testes, com baixa e com alta correlação com o vídeo original, 10% e 90% respectivamente. Os quadros inseridos com alta correlação tiram melhor proveito dos codificadores de vídeo mais modernos.

A.2 Utilização de Modulação de linhas horizontais do vídeo

Os testes utilizaram diversas câmeras comerciais com sensores APS e monitores com diversas resoluções de vídeo. A resolução dos monitores e dos sensores permite definir o algoritmo para a medida da qualidade do vídeo. Outras características dos sensores APS foram utilizadas para o cálculo da qualidade de vídeo.

Os vídeos de origem são os vídeos utilizados pelos testes da fase I do VQEG disponíveis na internet, estes vídeos estão em formato bruto sem compactação apresentados na Tabela A.1.

Estes vídeos são modulados com diversos limiares que tornarão a modulação perceptível ou não para o usuário dependendo da amplitude do fator de modulação.

Os vídeos após a modulação passarão por processos de inclusão de erros que deteriorarão a qualidade. Entre estes processos está a compressão com taxas muito baixas, perda de pacotes na rede distribuição e outros. Cada processo de deterioração gera artefatos de vídeo que poderão ser detectados na aplicação pelo sensor de imagens.

O vídeo com problemas será apresentado no monitor de televisão e o dispositivo sensor receberá estas informações e calculará a qualidade do vídeo.

A.3 Utilização de Marcas d'água digitais no vídeo

Descrito nos capítulos 1 e 2.

A.4 Resultados Obtidos

Os resultados apresentados envolvem formas de verificação de qualidade com sensores de imagem testados.

A verificação com super amostragem apresentada e também com a inclusão de marcas de testes no vídeo modulando horizontalmente os valores de luminância do vídeo.

No testes o vídeo “Foreman” foi utilizado com a inclusão de uma imagem de teste com correlação menor que 10% em relação aos quadros do vídeo, implicando em uma codificação com baixa eficiência devido a grande diferença entre os quadros originais do vídeo e as imagens de vídeo adicionadas. Com uma imagem de teste com correlação maior que 90% com os quadros do vídeo original, temos uma melhor compactação do vídeo.

Entre algumas das desvantagens da sobre amostragem está o aumento do tamanho dos arquivos, pois cada quadro adicional carregaria uma maior quantidade de bits. Outra desvantagem, que também se relaciona com o tamanho do arquivo, é a escolha do padrão de testes que será embutido ao vídeo com uma taxa de amostragem maior. Este padrão de testes deve ter alta correlação com o vídeo de interesse para que os codificadores que se aproveitam de correlação temporal possam compactar o vídeo de forma otimizada.

Tabela A.1: Novos Tamanhos dos Vídeos com Super Amostragem

Vídeo	Amostragem Original	Nova Amostragem	Novo Tamanho Correlação 10%	Novo Tamanho Correlação 90%	PSNR
Foreman	30	100	+10%	+20%	42
Vid2	30	100	+5%	+15%	38

Outra desvantagem seria a incapacidade de muitos monitores de apresentar vídeos em taxas mais elevadas o que inviabilizaria a recuperação do padrão de testes.

Utilizando a técnica de super amostragem com os equipamentos que temos disponíveis se mostrou ineficiente, pois a câmera deveria ter a capacidade de capturar muitos quadros por segundo, o que não é fácil de encontrar no mercado e ao mesmo tempo aumentaria os custos da aplicação e do sensor de imagens para o projeto.

Com as câmeras de mercado utilizadas, não foi possível fazer a detecção apropriada das marcas adicionadas.

O resultado apresentado no gráfico da Figura A.1 é o resultado da comparação dos resultados da aplicação proposta utilizando sensores de imagem com o sistema VQM padronizado pelo VQEG e ANSI. Para a obtenção dos resultados, utilizamos diversos vídeos, onde cada tipo de vídeo apresenta determinado tipo de cena, alguns com movimentos rápidos, outros com tons de pele, de forma a atingir a grande maioria das situações presentes em vídeos reais.

Na Figura A.1 pode-se verificar uma alta correlação de 0,79 entre os dois sistemas, sendo que o sistema VQM calcula sua métrica com referência completa, utilizando o vídeo original e o vídeo modificado.

O gráfico da Figura A.1 apresenta resultados obtidos com a aplicação com sensores de imagem.

Pode-se verificar dos resultados comparando o sistema proposto com o sistema VQM utilizado como referência, que existe uma grande correlação entre os dois, conforme o resultado de 0,79 apresentado no gráfico da Figura A.1.

Das simulações realizadas, pode-se notar que a alteração de parâmetros na cadeia de

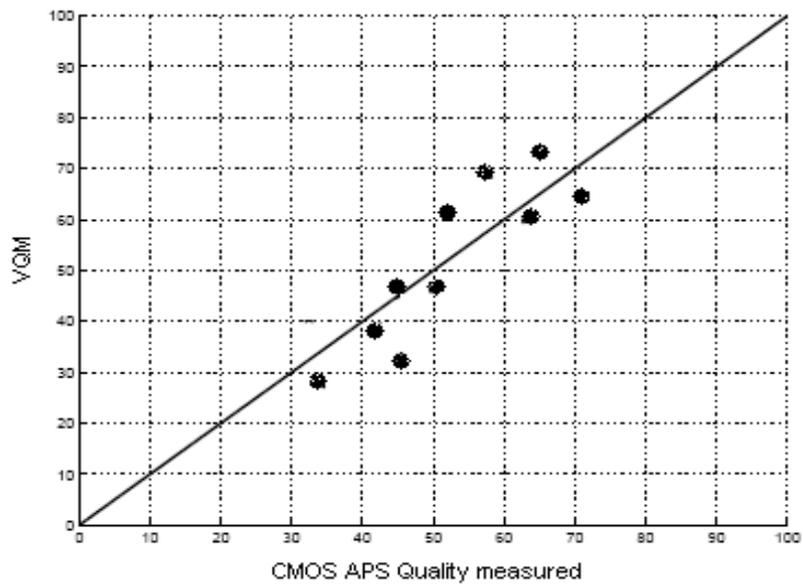


Figura A.1: Resultados da aplicação com relação ao VQM

transmissão do vídeo afeta a qualidade do vídeo de forma negativa, e esta alteração é captada pelo novo sistema proposto e identificado através de seus parâmetros.

B Testes APS-UFRJ NAMITEC

A seguir é reproduzido um documento de trabalho desenvolvido para o estudo sobre sensores de imagem APS (Active pixel Sensor). Nesse documento foram executados diversos testes para a caracterização de sensores de imagem desenvolvidos no Projeto NAMITEC.

Os resultados desses testes podem subsidiar projetos de novos dispositivos para a implementação da metodologia de testes de qualidade de vídeo.

Introdução

O objetivo do teste do chip APS-UFRJ é verificar o comportamento óptico na faixa do infravermelho próximo (NIR) acima de 700nm. A Responsividade deste chip para os comprimentos de onda na faixa de espectro visível é apresentada na figura abaixo.

A caracterização foi feita em duas etapas, sendo a primeira, verificando a resposta dos fotodiodos de teste, medindo a corrente para cada comprimento de onda. A segunda etapa foi o teste dos pixels que compõem o sensor conectado ao buffer interno (amplificador operacional conectado como amplificador isolador) do próprio chip, onde obtivemos a tensão na saída para cada comprimento de onda.

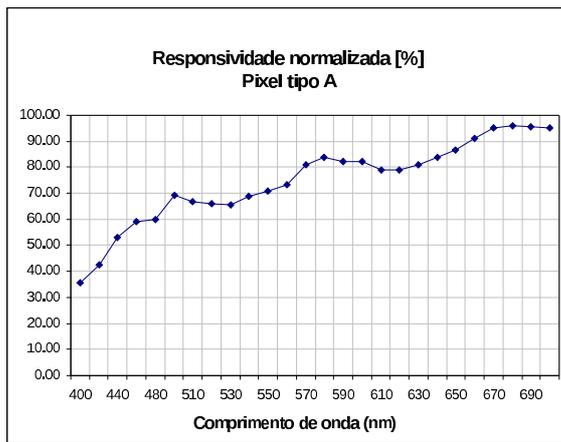


Figura 1 – Responsividade x Comprimento de onda (visível) Pixel APS-UFRJ.

Medidas Fotodiodos isolados de teste

A primeira experiência realizada para a caracterização dos fotodiodos foi a medida das curvas IxV para vários comprimentos de onda, na figura 2 são mostradas curvas típicas.

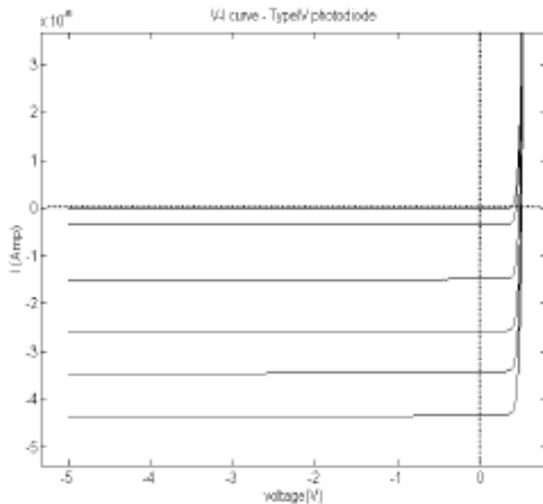


Figura 2 – Curvas Típicas IxV para fotodiodos com diferentes iluminações.

Na figura 3 a seguir são mostrados os resultados obtidos com os fotodiodos de testes do chip APS-UFRJ. A figura 4 apresenta estes resultados ajustados em relação a eficiência próprio fotodiodo para os comprimentos de onda medidos.

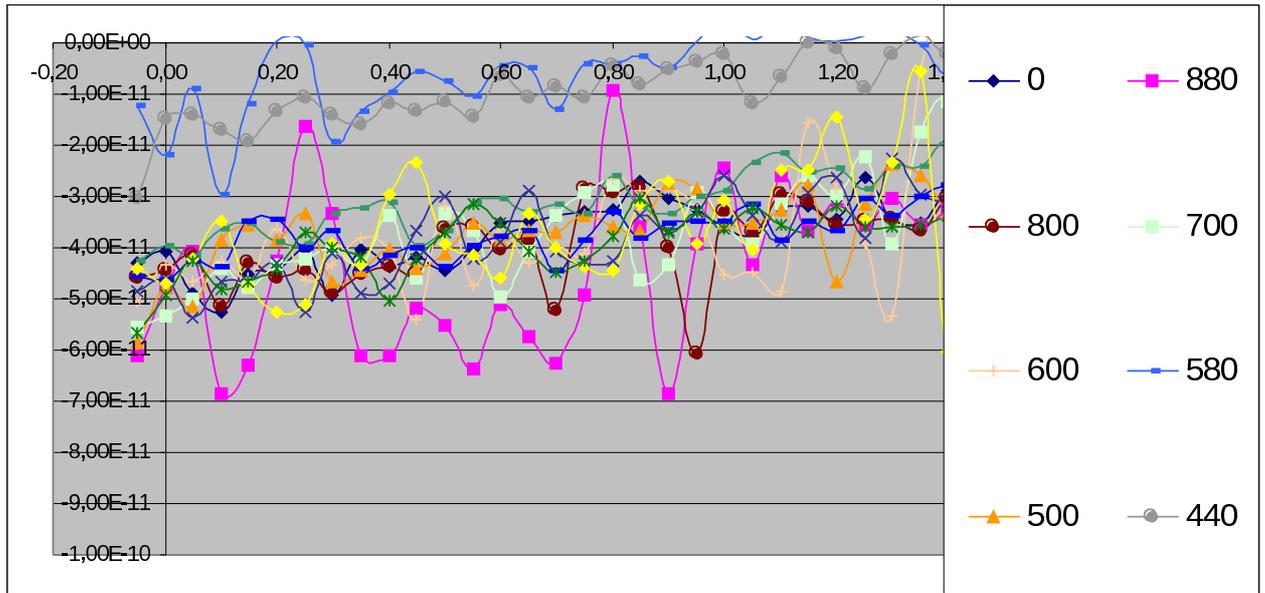


Figura 3 – Medidas de corrente para o fotodiodo de testes do APS-UFRJ para os comprimentos de onda da legenda, onde 0 = corrente de escuro. [Gráfico: $-V \times I$]

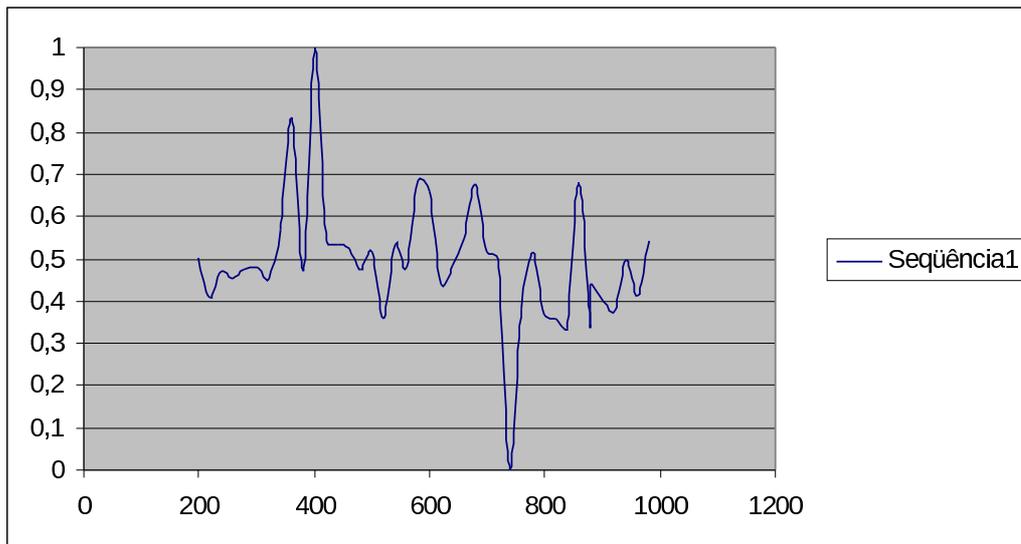


Figura 4 – Eficiência quântica relativa ao próprio fotodiodo.

Medidas APS-UFRJ

Os resultados apresentados na figura 6 são relativos à eficiência do próprio sensor. Para gerar o gráfico, foram feitas medições de tensão analógica na saída do amplificador isolador conectado a cada pixel. A forma de onda na saída do pixel é apresentada na figura abaixo, mostrando o reset do e a integração no pixel.

O varredura dos pixels do APS foi testada para frequências de 150KHz até 1MHz não apresentando problemas. Para os resultados da figura 6 foi utilizada frequência de clock de 500KHz.

Sensores APS CMOS típicos são lidos usando amostras em intervalos de tempo regulares. Após o tempo de integração, os valores dos pixels são lidos um a um, e então resetados para um determinado valor.

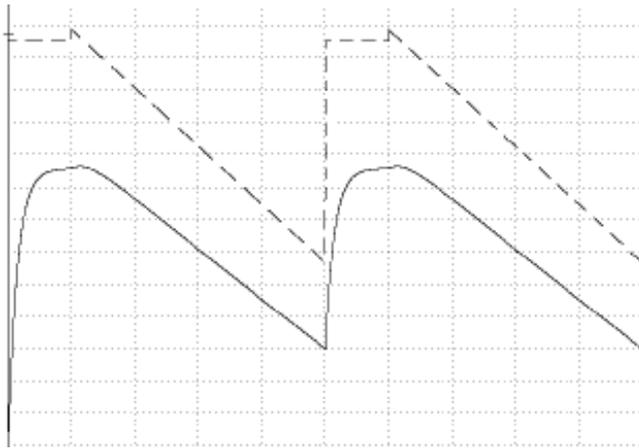


Figura 5 – Tensão de saída do Pixel Ativo.

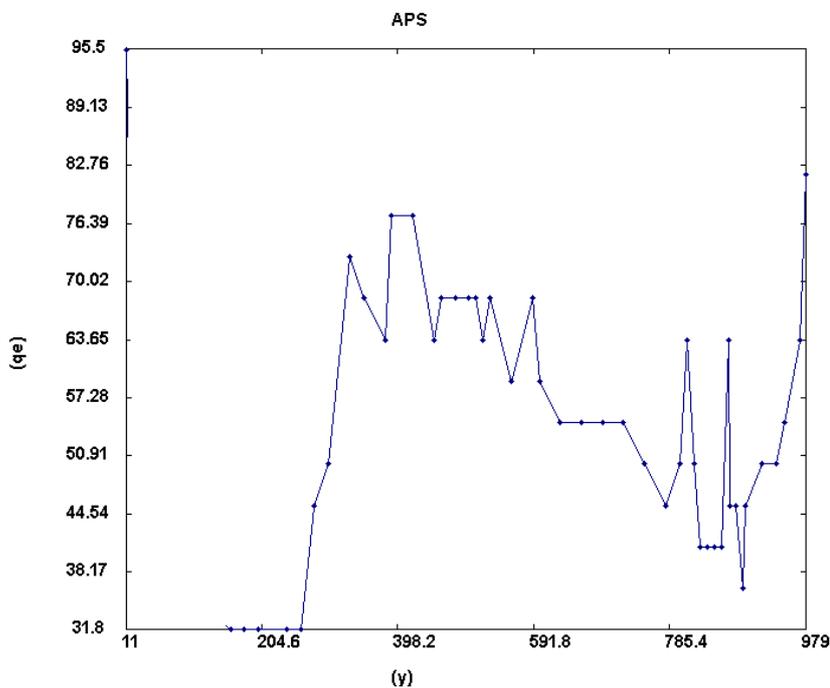


Figura 6 – Eficiência quântica relativa do pixel ativo.

Para uma medida mais confiável, a iluminação deve ser mantida constante, enquanto a saída é observada amostrando-se em intervalos regulares.

Os métodos comumente utilizados de medida de eficiência quântica para sensores de imagem, como os CCD, normalmente não levam em conta as não-linearidades introduzidas pelos circuitos lógicos de leitura de cada pixel, como no caso do APS.

Para uma medida mais confiável, a iluminação deve ser mantida constante, enquanto a saída é observada amostrando-se em intervalos regulares.

Para minimizar os efeitos das não linearidades de cada amplificador e circuitos de reset e leitura diferente para cada pixel, a leitura da saída do pixel deve ser observada continuamente e amostrada em intervalos de tempo regulares durante a integração das fotocargas.

Próximos Passos

Estamos trabalhando para minimizar as interferências observadas nas medidas realizadas e para tanto necessitaremos de algumas adaptações em nosso laboratório de medidas.

As medidas foram realizadas tomando como referência para normalização o próprio chip sob testes, o que pode ter mascarado alguns resultados, devido aos problemas de interferências. Para que os resultados sejam mais confiáveis, utilizaremos como referência um fotodiodo com especificações conhecidas para a comparação com os equipamentos testados.

O projeto para novas medidas foi iniciado de forma que possamos fazer a interface do APS com conversores A/D e microprocessador para aquisição e tratamento de imagens. Para a conversão A/D e lógica de habilitação será utilizada placa de aquisição da NI e interfaces com o Labview e demais equipamentos do laboratório de testes.

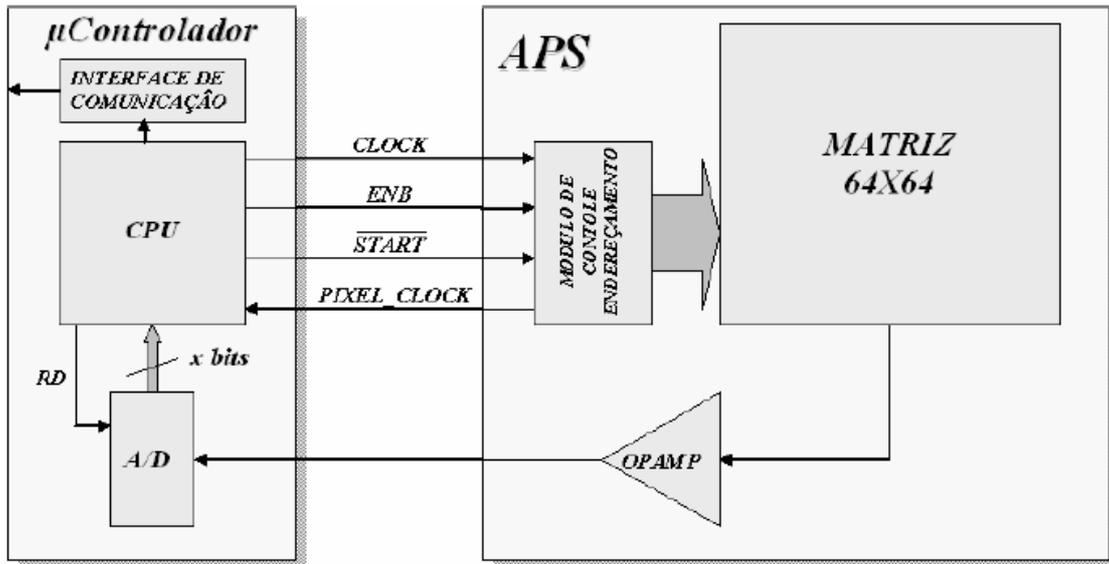


Figura 7 – Montagem para tratamento de imagens gerada pelo chip APS-UFRJ.

Para minimizar as interferências externas ao feixe óptico, deveremos equipar o laboratório com adaptadores para lâmpadas e para direcionamento e confinamento do feixe luminoso somente para os equipamentos sob teste, ao mesmo tempo em que evitemos que interferências externas não afetem o equipamento sob ensaio.

Como temos interesse na faixa do espectro NIR precisamos de um novo monocromador que trabalhe com comprimentos de onda acima de 1000nm. Para maior agilidade e precisão das medidas, além de interfaces com os outros equipamentos do laboratório, seria interessante que este monocromador disponha de interface GPIB e motor de passo controlado por esta interface para a seleção dos comprimentos de onda.

Referências

Boyd Fowler, Abbas El Gamal, David Yang, and Hui Tian; “A Method for Estimating Quantum Efficiency for CMOS Image Sensors”; Information Systems Laboratory, Stanford University

Mesquita Filho, A.C. et all; “Imageador APS”; LPC COPPE UFRJ e CCS UNICAMP.

Rocha Junior, M. F.; “Projeto e Caracterização de Imageadores APS Resistentes à Radiação”; Dissertação de Mestrado, UFRJ, 2007.

Testes APS – UFRJ

Atividades

Fase 1

Escolha dos componentes
Projeto da placa de testes
Confecção da placa de circuito impresso
Montagem dos componentes na placa de testes

Fase 2

Organização dos Equipamentos do Laboratório de testes

Fase 3

Testes elétricos
Testes ópticos

Resumo

Fase 1

O projeto da placa foi desenvolvido de forma que esta fosse bastante simples, barata e de rápida confecção por fabricante de pequeno porte. A placa foi projetada utilizando o software TraxMaker disponível nos laboratórios do ENE para que a confecção da placa de circuito impresso fosse realizada com o equipamento disponível no laboratório. Como este equipamento apresentou problemas técnicos, enviamos o projeto para ser executado em São Paulo.

A montagem/soldagem dos componentes da primeira versão da placa foi feita utilizando equipamentos de pequeno porte. Uma segunda placa de circuito impresso foi enviada para o laboratório para montagem com equipamento profissional, pois existe uma suspeita de contatos problemáticos gerando interferências/ruídos na primeira versão.

Fase 2

Equipamentos utilizados (vide detalhamento dos equipamentos no apêndice):

- para os testes elétricos: Labview integrado via GPIB a osciloscópio e fontes de tensão e geradores de função.
- para os testes ópticos: Lâmpada de vapor de mercúrio com respectiva fonte, monocromador, suportes.

Enquanto a placa era projetada e confeccionada, recebemos o monocromador H10 da Horiba-JobinYvon e preparamos uma bancada com suporte para este equipamento e para a lâmpada e para a placa de circuito impresso e demais equipamentos. Esta bancada foi montada de forma provisória para que pudéssemos testar equipamentos envolvidos e o APS-UFRJ de uma forma simplificada para validar o seu funcionamento.

Está sendo especificado um setup completo com equipamentos para direcionar o feixe óptico de forma mais precisa.

Fase 3

Os testes elétricos iniciaram com a validação da placa de circuito impresso.

Após esta verificação, utilizamos o APS-UFRJ e iniciamos as medidas dos componentes isolados, fotodiodos e transistores de teste. Nesta etapa geramos as curvas características dos componentes com auxílio de programas LabView utilizados no Laboratório.

As primeiras curvas características do fotodiodo foram levantadas em duas situações com o chip no escuro e com o chip com luz ambiente.

Os primeiros testes do APS levaram muito tempo, pois o primeiro chip estava com problemas (ou os causamos). Para isolarmos se o problema era na placa de circuito impresso ou no chip, fizemos diversos testes sem trocar o chip para evitar problemas com as outras amostras. Identificado que a placa estava sem problemas, trocamos o chip e conseguimos efetuar algumas medidas no APS.

Os testes ópticos foram feitos com o APS, o monocromador e a lâmpada e mercúrio, para cada comprimento de onda ajustado no monocromador, a corrente dos fotodiodos ou a tensão na saída do amplificador de pixel era medida. Com estes valores foram gerados os gráficos com os resultados.

Como a análise óptica está sendo feita sem um dispositivo de referência, alguns problemas que estamos detectando após as medidas, estão nos levando a repetir a mesma bateria de testes, atrasando a análise. Para evitar estes problemas, estamos verificando a possibilidade de montar novos equipamentos de forma definitiva.

Testes APS

Introdução

O trabalho realizado neste período compreendeu o estudo e preparação para os testes do APS desenvolvido na UFRJ, para que os testes fossem definidos foi feito um estudo das características e funcionalidades do chip protótipo operacional APS. Com base nas informações obtidas nos estudos, foi desenvolvida uma placa de circuito impresso para a execução dos testes, com a definição da placa estão sendo montadas as conexões com os equipamentos para a realização dos testes.

Características do APS desenvolvido na UFRJ.

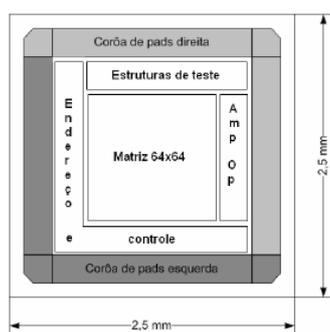
O chip que testaremos foi projetado por pesquisadores da UFRJ liderados pelo Prof. Antonio Mesquita, este chip é denominado protótipo operacional, uma vez que foram desenvolvidas duas versões de sensores APS na UFRJ.

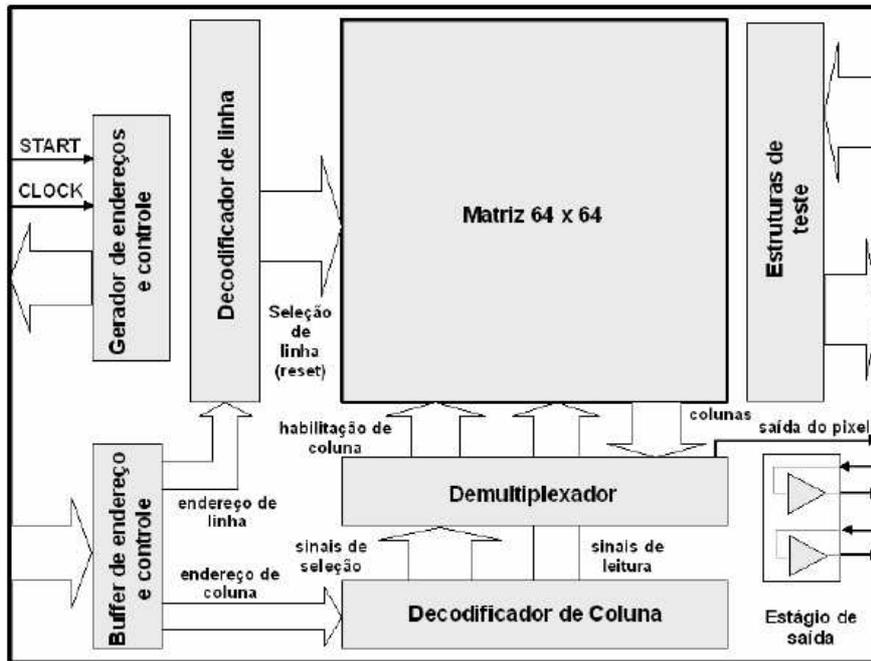
O primeiro chip denominado protótipo de testes continha várias estruturas de pixels e de transistores resistentes a radiação, que após a sua caracterização, as estruturas com melhores características elétricas, ópticas e de resistência a radiação deram origem ao protótipo operacional.

Este item descreve com detalhes o projeto do protótipo operacional do imageador APS desenvolvido em tecnologia AMS 0,35 um C35B4.

O protótipo operacional é composto de quatro partes principais. Uma matriz APS 64x64 associada a uma lógica de endereçamento, um módulo de controle e endereçamento automático, um estágio de saída constituído de dois amplificadores operacionais com a mesma topologia, porém com layouts distintos, e duas estruturas teste destinadas à caracterização elétrica e óptica (uma matriz composta de 14 transistores do tipo n e 14 transistores do tipo p e outra matriz contendo 100 fotodiodos).

Os 4 blocos principais que compõem o circuito integrado foram construídos para operarem isoladamente. Esta estratégia de projeto além de facilitar significativamente a depuração do circuito, permite o funcionamento da matriz mesmo que haja avaria em um ou mais blocos do chip. Seguindo esta metodologia foi construído um anel de pads (pad-ring) dividido em duas partes independentes, denominadas de “coroa esquerda” e “coroa direita”. A planta baixa do circuito e o layout do anel de pads são mostrados na figura abaixo.



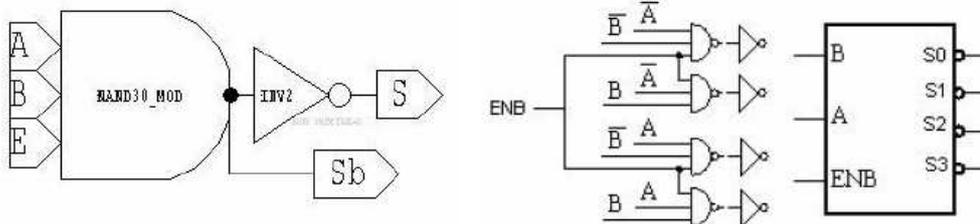


A matriz de pixels APS operacional contém 64x64 pixels. Cada pixel pode ser endereçado individualmente, seja para comandar a leitura do pixel, seja para forçar a inicialização (reset) da tensão do fotodiodo do pixel. Também é possível inicializar todos os pixels da matriz simultaneamente.

A lógica de endereçamento e controle permite que a matriz operacional seja endereçada tanto de modo estático quanto de maneira dinâmica e automática, através do gerador automático de endereços.

A seleção de cada pixel é feita habilitando o transistor de reset (seletor de linha – SL) da linha a qual o pixel pertence, por meio do sinal R_RST. Em seguida é ativado o transistor seletor de coluna (SC) através do sinal C_SEL, que irá conectar o transistor seguidor de fonte SF à saída do pixel (OUT), permitindo que o sinal trafegue através do restante do circuito até o estágio final de saída. Todos os outros pixels na mesma coluna, em linhas diferentes, não estão selecionados. Como consequência, o sinal de saída do transistor SF de apenas um pixel aparecerá naquela coluna.

A ativação do sinal R_RST dos pixels é feita via uma lógica de endereçamento de linha, e a ativação do sinal C_SEL é realizada através de uma lógica de endereçamento de coluna.



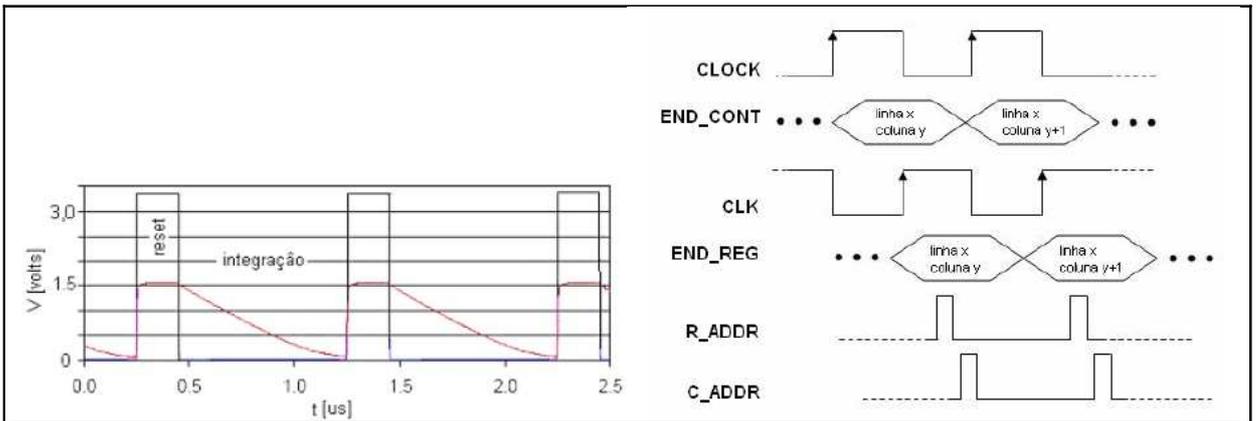
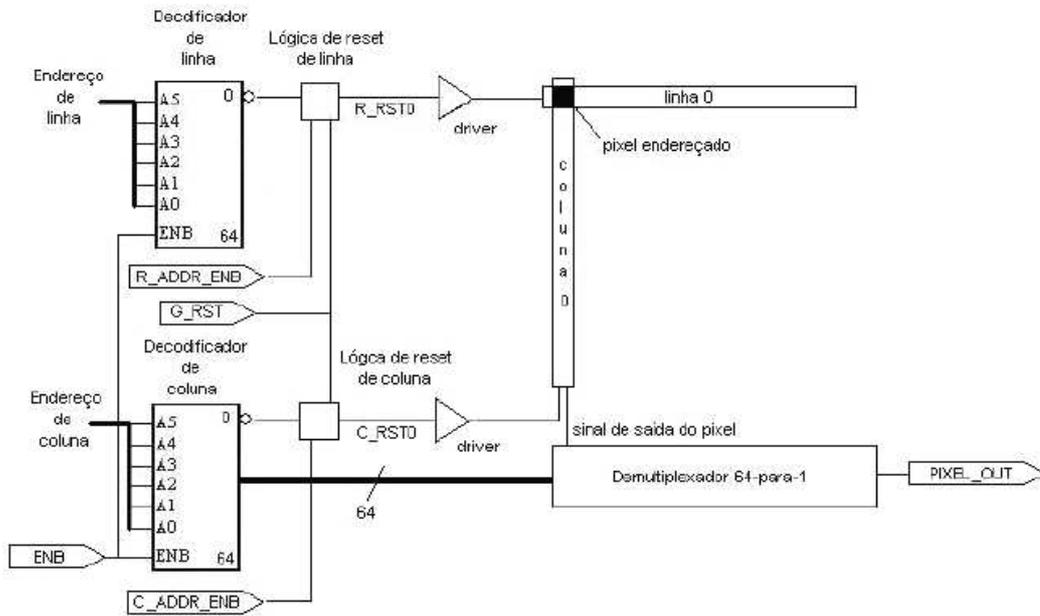
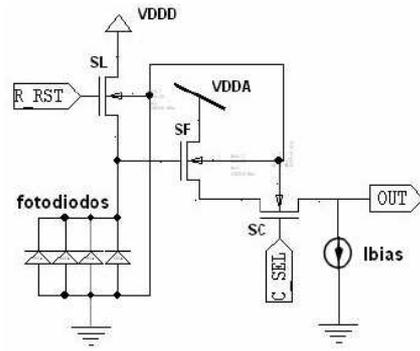
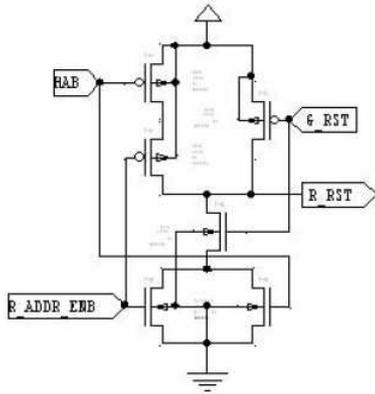
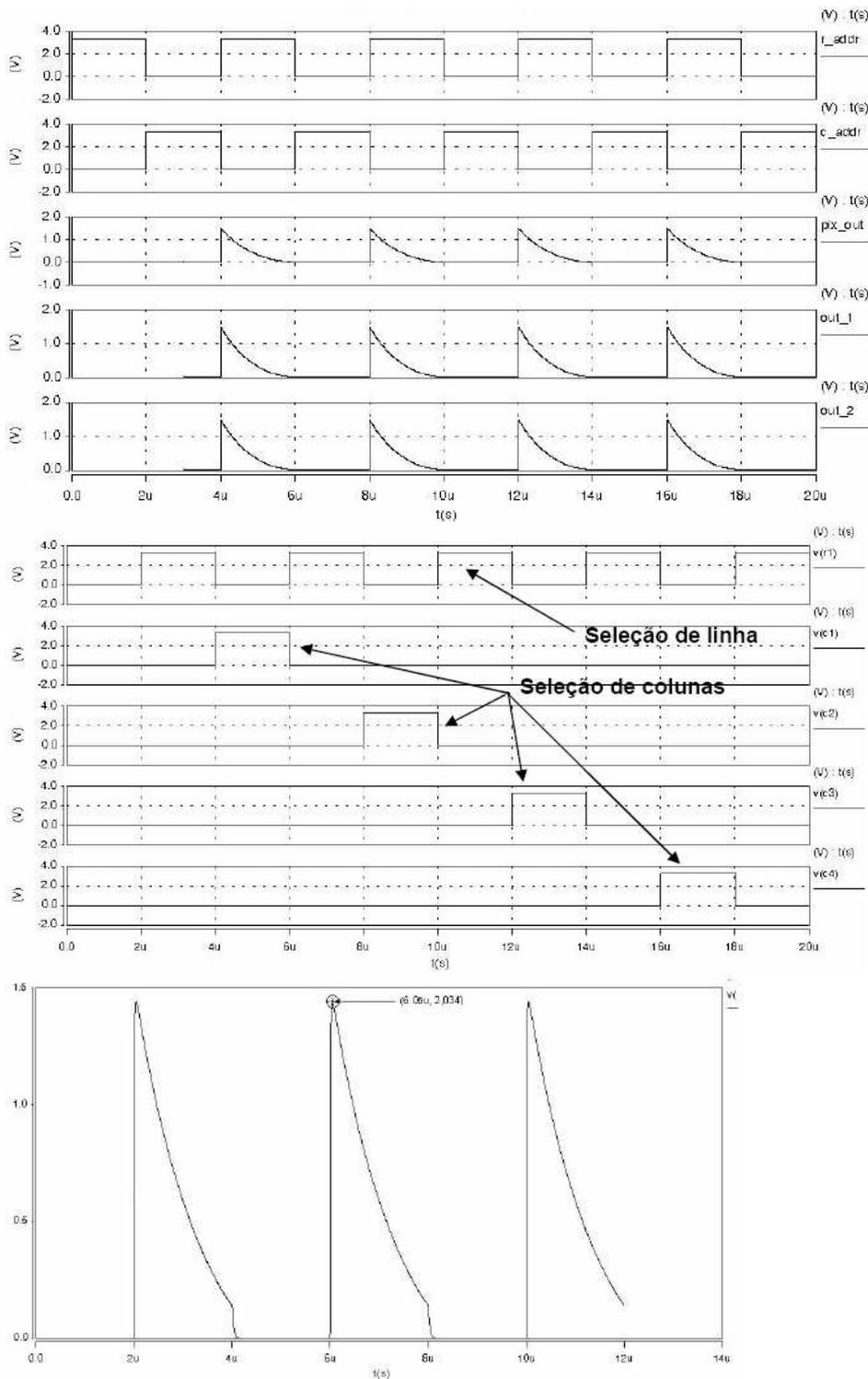


Diagrama típico do Pixel

Gerador de endereços



Simulação

O sinal de saída de cada coluna chega à saída global da matriz via um conjunto de chaves analógicas, as quais constituem um demultiplexador. Como a faixa dinâmica de saída do pixel não é próxima a V_{DD} (3,3 Volts), usam-se apenas transistores NMOS como chaves analógicas, reduzindo a capacitância total na saída. Estas chaves são comandadas, também, pela lógica de endereçamento de coluna.

A aquisição de imagens utilizando a matriz APS operacional deve obedecer a uma seqüência de operações de ordem bem definida. Inicialmente deve-se comandar a inicialização (reset) dos pixels, que pode

ser global ou por linha e, após um intervalo de tempo, é realizada a leitura da tensão de saída do pixel endereçado. O endereçamento da matriz pode ser executado remotamente via um dispositivo externo, ou de forma automática utilizando o mecanismo de endereçamento automático interno.

O ciclo de operação do pixel ocorre em três fases: fase de reset ou inicialização, fase de integração e fase de leitura.

- Fase de reset: A inicialização do pixel é efetuada com a aplicação da tensão de alimentação positiva (VDD) no terminal porta (gate) do transistor SL. Carregando, assim, as capacitâncias de junção dos fotodiodos.
- Fase de integração: A desativação do transistor SL, ou seja, retirada da tensão de 3,3 Volts da porta do transistor, deixando o catodo dos fotodiodos eletricamente isolados, dá início ao período de integração. Nesta situação é criada uma carga foto-gerada devido à incidência de luz (fótons) na superfície dos fotodiodos.
- Fase de leitura: Ao selecionar o transistor SC a tensão do pixel irá seguir para a saída, através de transistor SF, polarizado pela fonte de corrente externa ao pixel. O intervalo de tempo entre o reset e a leitura do pixel é variável. Usualmente, o instante de leitura do pixel ocorre na metade do período de integração.

O circuito de endereçamento e controle é composto por três componentes principais: a lógica de endereçamento de linha, a lógica de endereçamento de coluna e o gerador automático de endereços.

O endereçamento de linha da matriz é feito através de um decodificador digital de 6 entradas e 64 saídas (6x64). Este decodificador é construído a partir do cascadeamento de decodificadores 2x4 com entrada de habilitação (ENB).

A lógica de reset ou seleção de linha é composta por 3 transistores PMOS e 3 transistores NMOS. Ela permite tanto a habilitação de uma única linha, através dos sinais HAB, ativo em nível lógico “0” e R_ADDR_ENB, ativo em “0”, como a inicialização simultânea de todos os pixels da matriz, por meio do sinal de reset global G_RST, ativo em nível lógico zero.

Cada sinal de saída R_RST, associado à uma determinada linha da matriz, só será ativado, se o sinal R_ADDR_ENB estiver ativo (nível lógico “0”) e o sinal HAB correspondente à esta linha estiver habilitado (nível lógico “0”). Caso o sinal G_RST (reset global) seja acionado, em nível lógico “0”, todas as linhas da matriz serão “resetadas” em conjunto.

Como cada sinal de saída R_RST está associado a uma linha, ou seja, 64 pixels, a capacitância total da linha somada a capacitância das interconexões, torna-se significativa. Este problema é resolvido utilizando um estágio lógico adicional de drivers, capazes de fornecer mais corrente que uma porta lógica convencional, completando, assim, o circuito de endereçamento de linha da matriz.

O endereçamento de coluna pode ser feito de maneira similar ao que foi feito para as linhas. Exceto que, neste caso, as saídas das colunas são conectadas a um demultiplexador analógico de 64 entradas e 1 saída (64x1) acionado pelas 64 saídas do decodificador digital ativas em nível lógico alto (“1”).

Cada sinal de saída C_RST, associado à uma determinada coluna da matriz, só será ativado, se o sinal C_ADDR_ENB estiver ativo (nível lógico “0”) e o sinal HAB, uma das saídas do decodificador de coluna, correspondente à esta coluna, estiver habilitado (nível lógico “0”).

Caso o sinal G_RST (reset global) seja acionado (nível lógico “0”), todas as colunas da matriz serão “resetadas” em conjunto. O circuito de decodificação de coluna também é dotado de um estágio adicional de drivers com o propósito de aumentar a capacidade carga das portas lógicas convencionais.

O endereçamento da matriz pode ser realizado de modo remoto, empregando dispositivos externos, ou de maneira automática utilizando o gerador de endereços do próprio circuito integrado. Este gerador é composto por três contadores de 4 bits totalizando 12 bits de endereçamento, 6 bits para endereçar as linhas e outros 6 bits destinados ao endereçamento das colunas da matriz. Cada contador de 4 bits é constituído por 4 flip-flops do tipo toggle posicionados de maneira seqüencial.

O funcionamento do contador de 12 bits depende apenas de dois sinais de controle: o sinal CLOCK e o sinal START. O sinal START é responsável pela inicialização de todos os flip-flops do contador, levando suas saídas ao nível lógico “0”. A partir daí, a saída de cada flip-flop só mudará de estado, caso os flip-flops anteriores já tenham assumido todas as combinações binárias possíveis. Todas as trocas de estado dos flip-flops são comandadas a partir de transições positivas do relógio básico do contador através do sinal CLOCK.

O gerador de endereços é responsável, também, pela geração dos sinais de controle R_ADDR e C_ADDR, que têm o propósito de habilitar, respectivamente, a linha e a coluna que serão acessadas.

O circuito de saída é composto por dois amplificadores operacionais de três estágios com entrada diferencial que funcionam de modo independente, o que confere um grau de redundância ao estágio de saída do chip. Estão configurados para operarem como buffers com o propósito de aumentarem a capacidade de carga do sinal de saída do pixel.

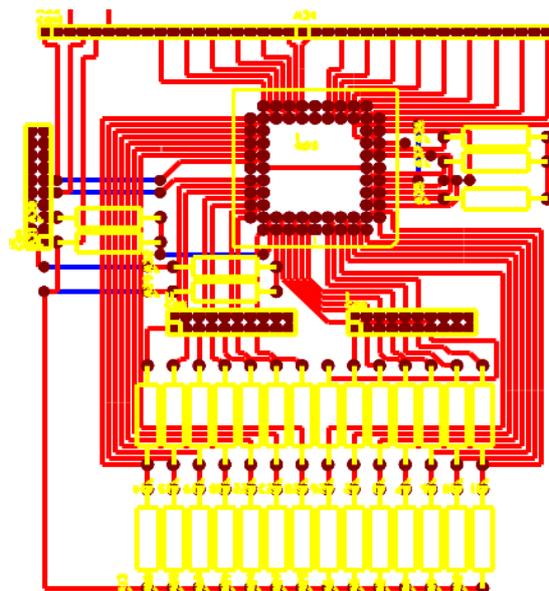
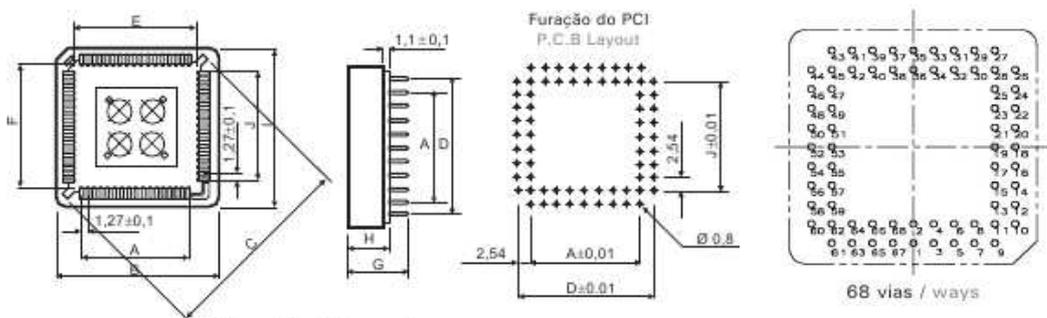
O protótipo do imageador APS operacional é dotado de duas estruturas de testes. Uma matriz de transistores MOS contendo uma sub-matriz com 14 transistores do tipo n e outra com 14 transistores de tipo p, destinados à caracterização elétrica do circuito integrado. A segunda estrutura é uma matriz composta de 100 fotodiodos idênticos aos utilizados na composição da matriz principal. Estes elementos têm o propósito de atuar na caracterização óptica do chip.

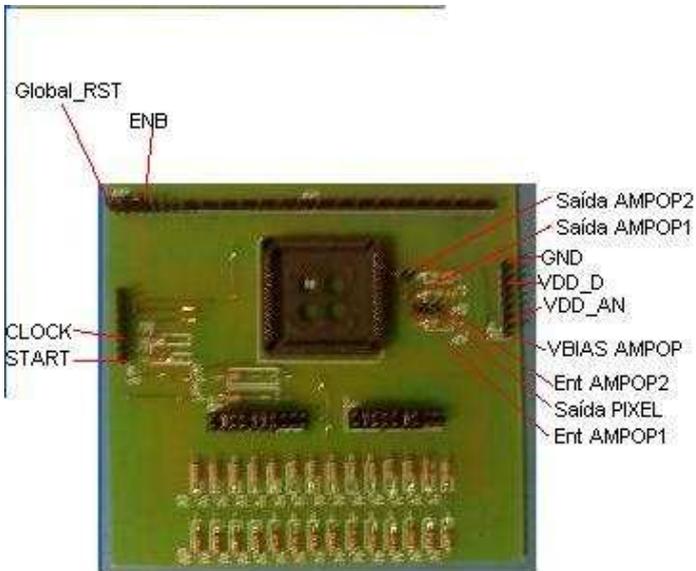
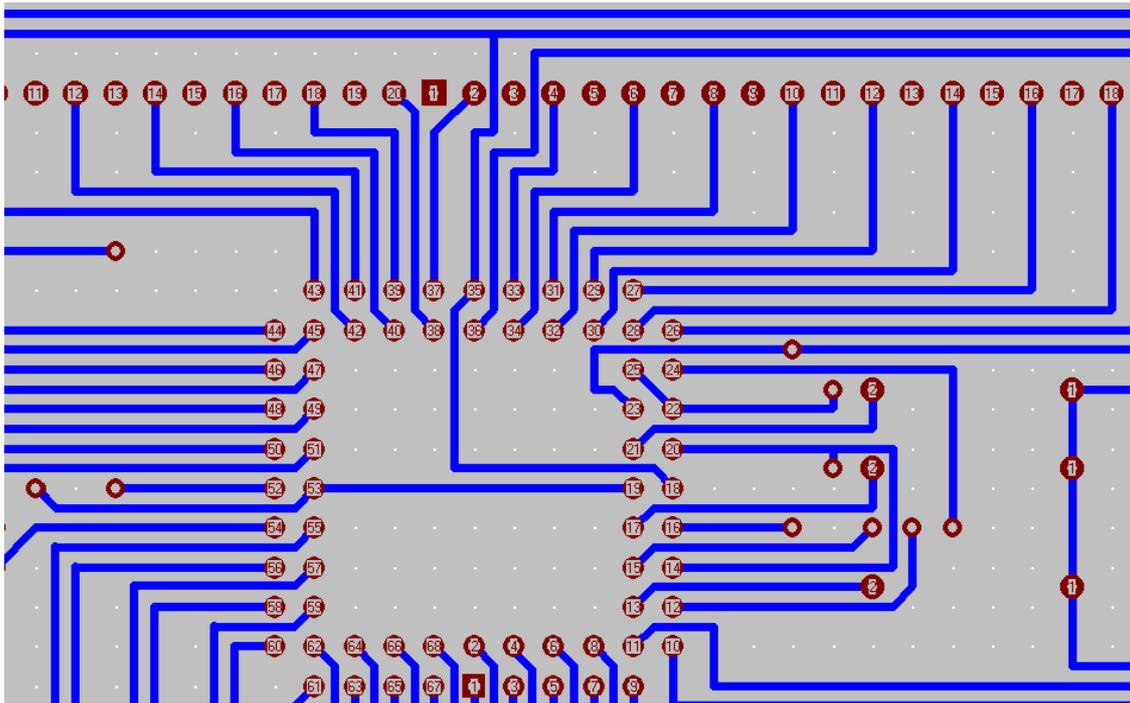
Especificação e Projeto de placa de circuito impresso para testes APS.

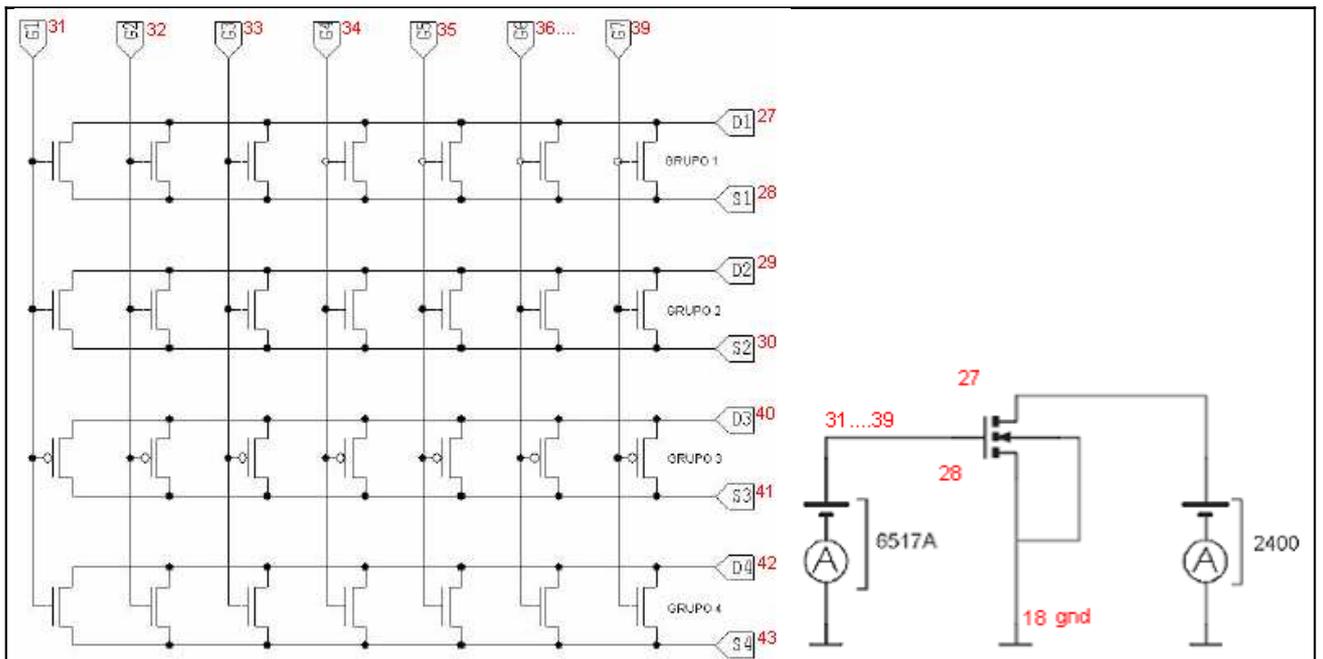
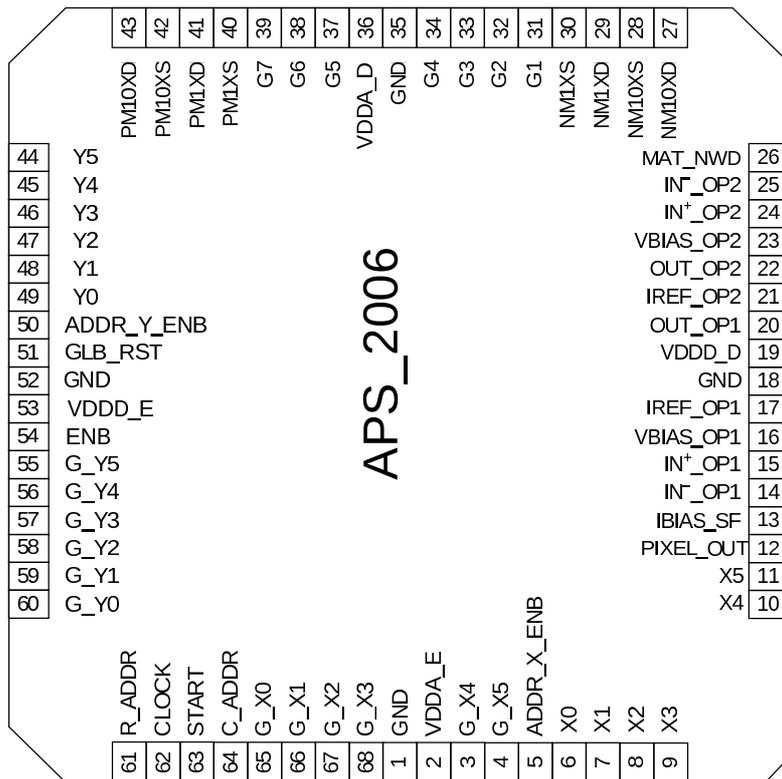
A placa de circuito impresso foi projetada levando em consideração o soquete escolhido e as características dos equipamentos de teste a serem utilizados e suas interfaces.

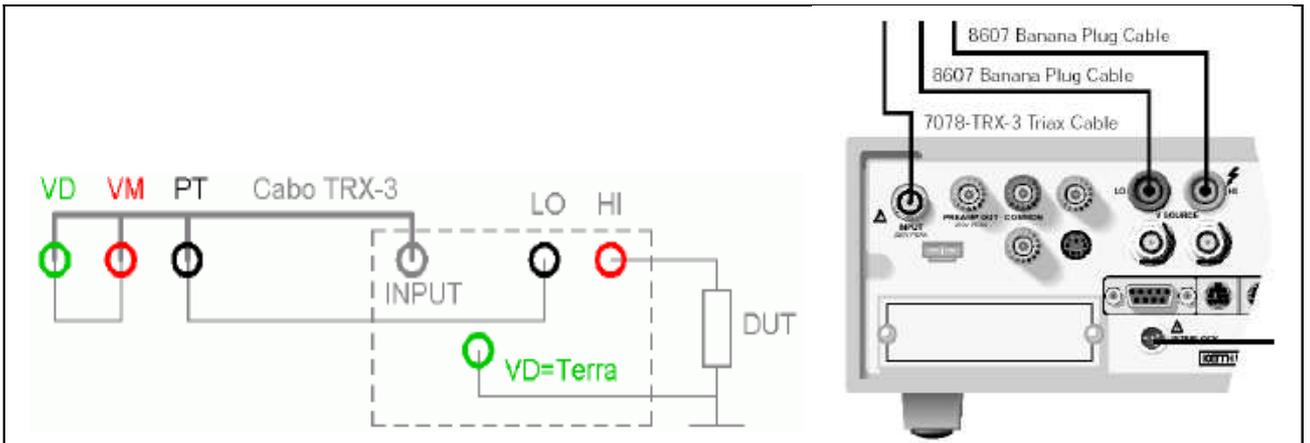
Foi utilizado soquete PLCC68 com as especificações conforme figuras abaixo. Os demais componentes da placa servem para adaptação das tensões do Analisador Lógico para o Chip APS, bem como para a polarização dos circuitos analógicos do chip. As interfaces do chip com o analisador lógico serão feitas com barras de terminais do tipo 3M2520, estes mesmos terminais servirão para instalação de jumpers para os testes em modo manual (endereçamento de cada pixel do APS) ou automático com o endereçamento do chip feito internamente ao chip.

Para a execução do layout da placa de circuito impresso foi utilizado o software Traxmaker disponível no ENE.

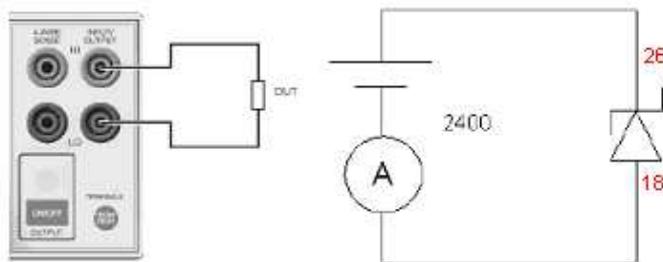
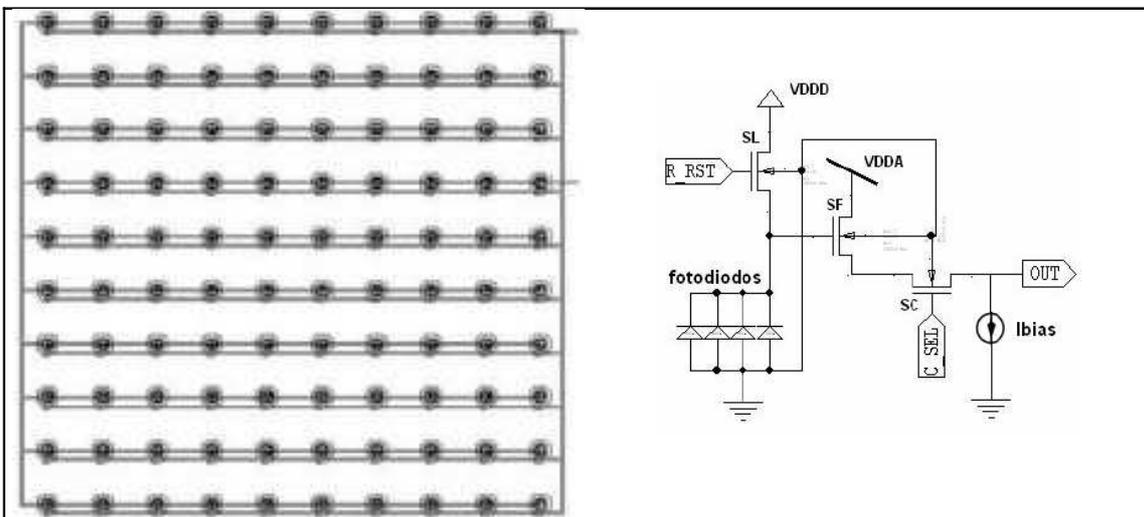


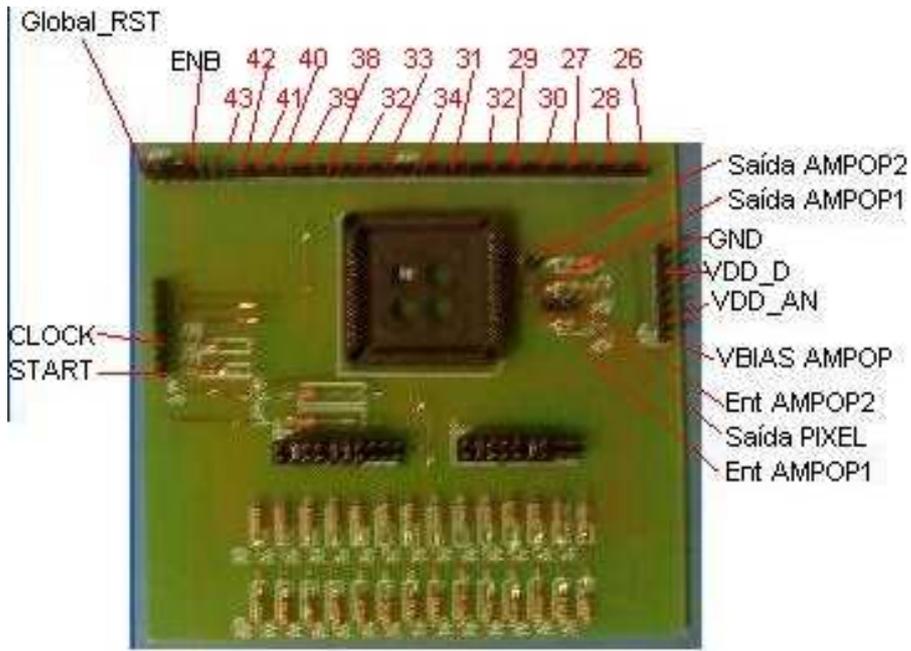






A caracterização elétrica dos transistores foi realizada de acordo com os procedimentos definidos para a extração dos parâmetros específicos do modelo EKV em (ENZ et al., 1995). Os parâmetros extraídos foram: a tensão de limiar, o fator de corpo, a transcondutância ou fator ganho, o coeficiente de redução de mobilidade, o coeficiente de modulação do comprimento do canal e os parâmetros relativos às dimensões do canal, descritos mais adiante.





Pinagem completa

SINAL	TIPO	DESCRIÇÃO	DIREÇÃO	ATIVACÃO	PINO
Y0 – Y5	DIG	Endereço de linha	entrada	-	49 a 44
X0 – X5	DIG	Endereço de coluna	entrada	-	6 a 11
ADDR_Y_ENB	DIG	Habilita endereço de linha	entrada	0	50
ADDR_X_ENB	DIG	Habilita endereço de coluna	entrada	0	5
ENB	DIG	Sinal de habilitação global	entrada	1	54
GLB_RST	DIG	Reset global	entrada	0	51
GY0 a GY5	DIG	Gerador – endereço de linha	saída	-	60 a 55
GX0 a GX3	DIG	Gerador – endereço de coluna	saída	-	65 a 68
GX4 - GX5	DIG	Gerador – endereço de coluna	saída	-	3 – 4
R_ADDR	DIG	Gerador - habilita endereço de linha	saída	-	61
CLOCK	DIG	Gerador - clock	entrada	-	62
START	DIG	Gerador - reset	entrada	-	63
C_ADDR	DIG	Gerador - habilita endereço de coluna	saída	-	64
PIX_OUT	ANA	Saída do pixel	saída	-	12
REF_MUX	VDC	Sinal de referencia do pixel	entrada	-	13
OP1_IN-	ANA	Entrada inversora do amplificador 1	entrada	-	14
OP1_IN+	ANA	Entrada não inversora do amplificador 1	entrada	-	15
OP1_VB	VDC	Tensão de polarização do amplificador 1	entrada	-	16
OP1_REF	VDC	Referencia do amplificador 1	entrada	-	17
OP1_OUT	ANA	Saída do amplificador 1	saída	-	20
OP2_REF	ANA	Referencia do amplificador 2	entrada	-	21
OP2_OUT	ANA	Saída do amplificador 2	saída	-	22
OP2_VB	VDC	Tensão de polarização do amplificador 2	entrada	-	23
OP2_IN+	ANA	Entrada não inversora do amplificador 2	entrada	-	24
OP2_IN-	ANA	Entrada inversora do amplificador 2	entrada	-	25
MATRIZ_NDW	VDC	Matriz de fotodiodos de teste	entrada		26
NM10XD	ANA	Dreno da matriz de transistores NMOS interdigitados	entrada		27
NM10XS	ANA	Fonte da matriz de transistores NMOS interdigitados	entrada		28
NM1XD	ANA	Dreno da matriz de transistores NMOS	entrada		29
NM1XS	ANA	Source da matriz de transistores NMOS	entrada		30
G1	ANA	Porta 1 das matrizes de transistores	entrada		31
G2	ANA	Porta 2 das matrizes de transistores	entrada		32
G3	ANA	Porta 3 das matrizes de transistores	entrada		33
G4	ANA	Porta 4 das matrizes de transistores	entrada		36
G5	ANA	Porta 5 das matrizes de transistores	entrada		37
G6	ANA	Porta 6 das matrizes de transistores	entrada		38
G7	ANA	Porta 7 das matrizes de transistores	entrada		39
PM1XS	ANA	Fonte da matriz de transistores PMOS	entrada		40
PM1XD	ANA	Dreno da matriz de transistores PMOS	entrada		41
PM10XS	ANA	Fonte da matriz de transistores PMOS interdigitados	entrada		42

SINAL	TIPO	DESCRIÇÃO	DIREÇÃO	ATIVAÇÃO	PINO
PM10XD	ANA	Dreno da matriz de transistores PMOS interdigitados	entrada		43
GND	VDC	Terra de sinal	entrada	-	1, 18, 34 e 52
VDDA_E	VDC	VDD analógico da coroa esquerda	entrada	-	2
VDDA_D	VDC	VDD analógico da coroa direita	entrada	-	35
VDDD_E	VDC	VDD digital da coroa esquerda	entrada	-	53
VDDD_D	VDC	VDD digital da coroa direita	entrada	-	19

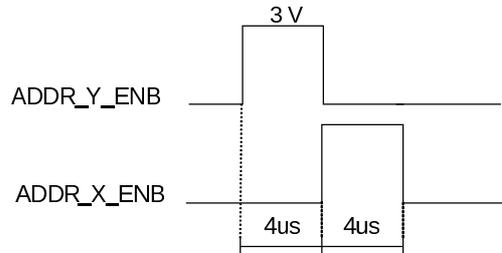
1 - Sinal ENB pino em V_{DD}

2 - Sinal GLB_RST em V_{DD}

3 – Polarizar os espelhos de corrente dos seguidores de fonte do pixel - I_{BIAS} entre 1 e 2 μA)

4 - Alimentar os sinais V_{DD} com 3,3 V e aterrar os sinais GND. Notar a diferença entre VDD analógico (VDDA) e digital (VDDD).

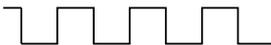
Após a configuração das chaves deverão ser gerados os sinais de ADDR_Y_ENB e ADDR_X_ENB com as seguintes características:



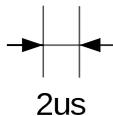
Monitorar o sinal de saída do pixel no pino 12 (PIX_OUT) (precisa de um buffer analógico para visualizar no osciloscópio)

Ligações para Endereçamento Automático

1. Conectar as entradas Y0 a Y5 às saídas G_Y0 a G_Y5 respectivamente.
 2. Conectar as entradas X0 a X5 às saídas G_X0 a G_X5 respectivamente.
 3. Conectar a entrada ADDR_Y_ENB a saída R_ADDR e a entrada ADDR_X_ENB a saída C_ADDR.
 4. Gerar sinais de CLOCK e START segundo o diagrama:
 5. Manter os sinais de ENB, GLB_RST na configuração anterior.
- Note que os sinais ADDR_Y_ENB e ADDR_X_ENB agora são gerados automaticamente.

clock  500kHz, 50% duty-cycle

start  Em qualquer instante - operação assíncrona



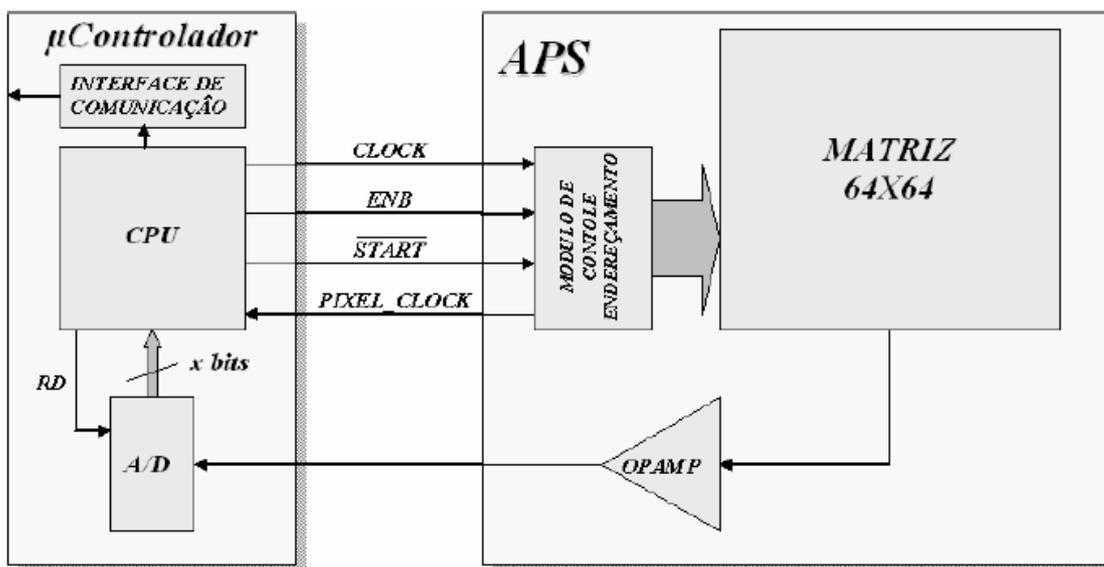
Resumo da pinagem:

Sinal	Pino
Y0 a Y5	49 a 44
G_Y0 a G_Y5	60 a 55
X0 a X5	6 a 11
G_X0 a G_X3	65 a 68
G_X4 a G_X5	3 a 4
ADDR_Y_ENB	50
R_ADDR	61
ADDR_X_ENB	5
C_ADDR	64
CLOCK	62
START	63

Bibliografia

- [1] LIMA, Kátia Goretti, “ESTRUTURAS APS RESISTENTES À RADIAÇÃO PARA APLICAÇÕES ESPACIAIS”. Dissertação de Mestrado, UFRJ, 2006.
- [2] ROCHA JUNIOR, Milton Ferreira, “PROJETO E CARACTERIZAÇÃO DE IMAGEADORES APS RESISTENTES À RADIAÇÃO”. Dissertação de Mestrado, UFRJ, 2007.
- [3] TRAXMAKER, Ferramenta de Software – Projeto PCI.

ANEXO



Do lado direito o Chip UFRJ e do lado esquerdo o controle.

A saída do A/D pode ir direto para memória, endereçado pela CPU.

Verificação do monocromador Horiba-Jobin-Yvon H10.

Introdução

O laboratório de medidas ópticas para caracterização de dispositivos (LTSD) dispõe dos seguintes equipamentos:

- Lâmpada de Mercúrio 50W PHYWE 08147.00
- Monocromador H10 – Horiba Jobin Yvon
- Laser He-Ne @ 633nm JDS Uniphase 1135-P
- Medidor de Potência Thorlabs PM100 com detector de silício S130A (400-1100nm)

A lâmpada de mercúrio utilizada foi emprestada de outro laboratório e é utilizada em experiências com menor necessidade de resolução de comprimento de onda, esta lâmpada não apresenta sua resposta em comprimento de onda.

Os dispositivos a serem caracterizados e os dispositivos detectores de referência do laboratório estavam apresentando picos em certos comprimentos de onda e precisávamos identificar a origem destes picos (figura 1). Para esta identificação, isolamos os componentes e verificamos as calibrações do monocromador e pesquisamos as características de lâmpadas semelhantes a que dispúnhamos.

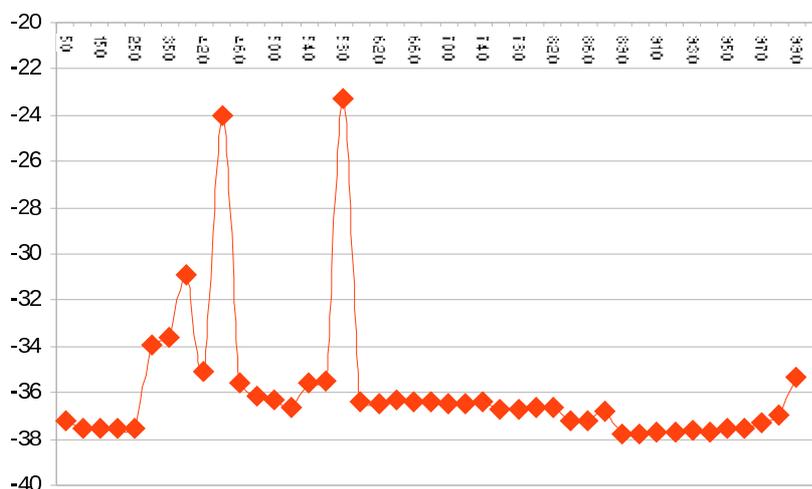


Figura 1 – Caracterização detector de silício no LTSD

As figuras 2 e 3 apresentam curvas típicas para lâmpadas de mercúrio e de xenônio para comparação. Na figura 2 podemos verificar que a lâmpada de mercúrio apresenta picos em determinados comprimentos de onda compatíveis com os observados no experimento da figura 1.

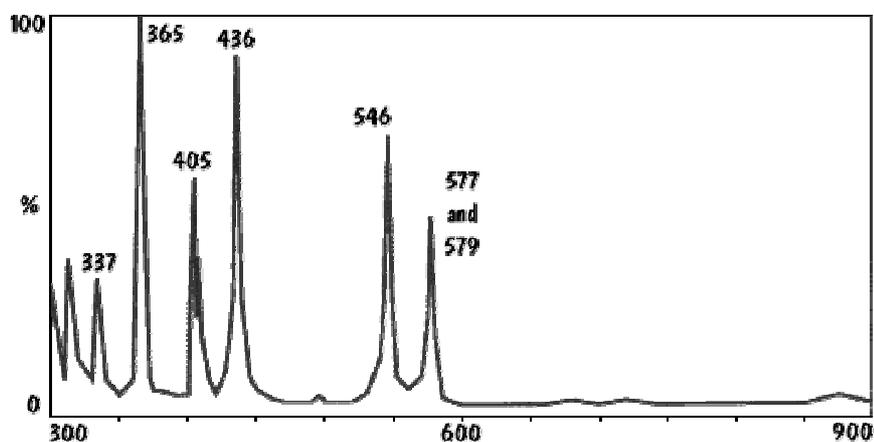


Figura 2 – Lâmpada típica de mercúrio

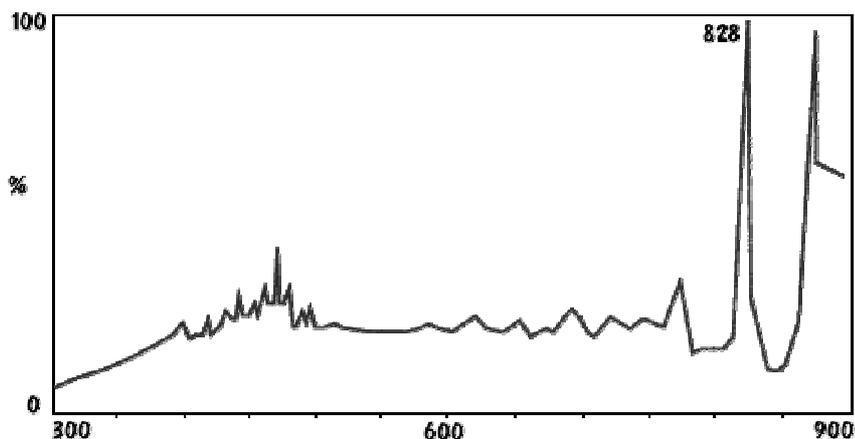


Figura 3 – Lâmpada típica de xenônio

Verificação do H10 com laser de He-Ne @ 633nm (JDS Uniphase 1135-P)

Na figura 4, a verificação da potência na saída do monocromador para um laser de He-Ne de comprimento de onda 633nm. Podemos verificar que neste comprimento de onda o monocromador está calibrado com uma pequena diferença de 1nm, que pode ser devido a erro no mostrador do monocromador, bem como do laser, mas a variação é aceitável para as medidas em curso.

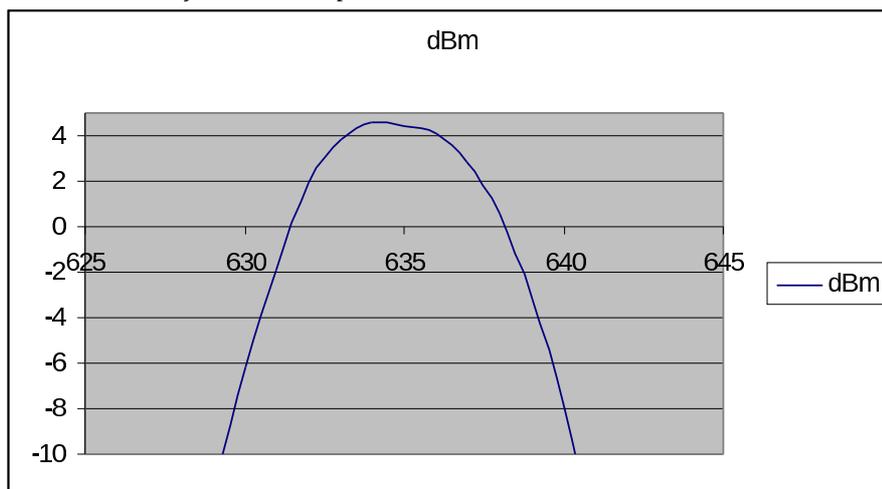


Figura 4 – Verificação do monocromador com laser de 633nm.

Resultados

Identificamos que a causa dos picos se deve à lâmpada de mercúrio utilizada e que o monocromador está calibrado. Poderemos eliminar a influência destes picos utilizando os resultados do nosso dispositivo sob teste em relação ao dispositivo de referência, utilizando a mesma lâmpada para ambas as medidas.

Referências Bibliográficas

- Marston, Neil, Solid-state imaging: a critique of the CMOS sensor, *Thesis submitted for the degree of Doctor of Philosophy, University of Edinburgh*, Edinburg, November 1998.
- IBIS5-A-1300 1.3M Pixel Dual Shutter Modo CMOS Image Sensor Data sheet, Cypress semiconductor Corporation, San Jose CA.
- IBIS4-1400 14 M Pixel Rolling shutter CMOS Image Sensor Data sheet, Cypress semiconductor Corporation, San Jose CA
- IBIS4-6600 High resolution 6.6 M Pixel Rolling shutter CMOS Image Sensor Data sheet, Cypress semiconductor Corporation, San Jose CA
- IBIS4-1300 1.3M Pixel Rolling shutter CMOS Image Sensor Data sheet, Cypress semiconductor Corporation, San Jose CA
- OV7648 Color CMOS VGA/OV7148 B&W CMOS VGA Camera Chip, Datasheet, OmniVision Technologies, Sunnyvale, CA, 2004.
- OV7649 Color CMOS VGA/OV7149 B&W CMOS VGA Camera Chip, Datasheet, OmniVision Technologies, Sunnyvale, CA, 2004.
- OV7660/OV7161 B&W CMOS VGA Camera Chip, Datasheet, OmniVision Technologies, Sunnyvale, CA, 2004.
- HDCS-1020, HDCS-2020 CMOS Image Sensors, Data Sheet, Agilent Technologies, Inc, Palo Alto, CA, 2001.
- Agilent CMOS Monochrome Imagen Sensor ADCS-1120 (CIF) and ADCS-2120 VGA, Data Sheet, Agilent Technologies, Inc, Palo Alto, CA, 2001.
- Agilent ADCS-1021 ADCS-2021 CMOS Imagen Sensor, Data Sheet, Agilent Technologies, Inc, Palo Alto, CA, 2001.
- Agilent ADCS-1121 ADCS-2121 CMOS Monochrome Imagen Sensor, Data Sheet, Agilent Technologies, Inc, Palo Alto, CA, 2001.
- V. Bockaert, *Color Filter Array*, *Digital Photography Review*, http://www.dpreview.com/learn/?/Glossary/Camera_System/Color_Filter_Array_01.htm
- P. B. Catrysse, W. Suh, S. Fan, M. Peeters, *One-mode model for patterned metal layers inside integrated color pixels*, *Optics Letters*, Vol. 29, No. 9, May 2004.
- A. Dickinson, B. Ackland, E. S. Eid, D. Inglis, E.R. Fossum, *Standard CMOS active pixel image sensors for multimedia applications*, 16th Conference on Advanced Research in VLSI, 1995.
- El-Sayed Eid, Suat U. Ay, Eric R. Fossum, *Design of Radiation Tolerant CMOS APS System-on-a-chip Image Sensors*, *Aerospace Conference Proceedings, IEEE*, page(s): 4-2005- 4-2011 vol.4, 2002.

C Trabalho IEEE-CNCC 2010

A seguir é reproduzido o trabalho “CMOS Image Sensor Device for Objective Evaluation of Video Quality in Mass Distribution Networks” publicado e apresentado no IEEE-CNCC 2010 (Consumer Communications & Networking Conference) em Las Vegas em janeiro de 2010.

CMOS Image Sensor Device for Objective Evaluation of Video Quality in Mass Distribution Networks

Márcio L. Graciano, Jr., Alexandre R. S. Romariz, José C. da Costa
Dep. Electrical Engineering
University of Brasilia
Brasilia, Brazil

Abstract— This work aims to present a device for the objective evaluation of video quality from the user perspective, in mass content distribution networks. This simple and cost-effective device with an integrated image sensor and an embedded processor intends to be network pervasive while minimizing the network overhead, allowing quality measurements over the network with greater control of the content distribution throughout the video chain, from content producers, distributors and to consumers.

Keywords- CMOS; Image sensor; APS; video quality

I. INTRODUCTION

With the current technological development and of telecommunications networks and audiovisual content, digital video is becoming more complex and is available in multiple formats. There are a lot of possibilities for source and channel coding and compression, transmission media and decoding in the receptor. The quality assurance of these technologies is not easily available to the network users.

The evaluation of video quality as well as the efforts for the establishment of standards in that field [1]-[8] has taken into account technical oriented quality aspects of the distribution chain of video, using large and expensive measure devices, applicable to use in laboratories and in the field instead of in the user side. Quality assessment close to the users actually would require the employment of a large number of devices to capture information, which could increase the costs for quality evaluation.

In the proposed scheme, we use video cameras with CMOS active pixel image sensors, APS, which capture test patterns embedded in the videos, and a processing software embedded in each user's decoder.

To measure the quality of distributed content, a subliminal image, uncorrelated to the original video, is embedded in the video so that human observers cannot see it naturally [9]-[11]. However the device can recover testing patterns and calculate objectively the quality of video distributed to the user.

The test pattern will be inserted in the generation of video so that it can be tested throughout the distribution chain, from encoding and transmission to decoding. The test pattern analysis will be used to estimate the possible impact on video quality.

II. DEVICE DESCRIPTION

The device for measuring video quality consists of an encoder inserting new information as test patterns modulated over the original video, at an early point in the distribution chain, soon after its creation/generation and before any processing or coding, so that one can measure the quality end to end, taking into account all the problems of encoding, processing, quantization, compression and others. This new information embedded is not correlated with the original video, simplifying modulation and may be applied to any type of video.

The information embedded in the generation of the video is designed so that the user reception, regardless of network type or the display type, would be possible detect and estimate video quality using our image sensor.

The proposed device for quality assessment from the user's perspective is built using a SoC with APS image sensor and a processing unit. This SoC is a development work in the area of smart integrated nano- and microelectronic systems, which enable the creation of autonomous electronic smart sensor networks, and embedded and self-configurable systems for telecommunications [12][13].

These SoCs can be built into the video receiver's remote control, so that for the assessment of video quality one just needs to point the remote control towards the TV set in the same way as one would change the channel.

The signals received by the APS from the TV monitor, will be sent to the processing unit of the embedded SoC application that calculates video quality. These devices are decoupled from the network, not affecting its performance and avoiding extra costs for the operators. This system can extract pre-determined parameters that indicate the type of problem in the distribution and the more common artifacts in the video distribution chain.

III. RESULTS

The video quality verification uses an apparently simple and flexible reference circuit allowing multiple configurations. This circuit is composed of three blocks: a source, a channel and a receiver, which can be separately parameterized. In the first block, source coding, MPEG encoders were used and the parameters for coding rates and/or transmission error rates

were chosen to deliver quality video to the typical situations encountered on TV scenarios, while getting a good quality distribution for different scenes. The transmission section used coded sequences and transmission rates from 512 kb/s to 50 Mb/s.

Quality measurements were carried out employing sequences with various contents, covering a range of typical applications for TV, like sports, music and video clips. We also selected these because they display a diverse set of features, such as flat surfaces, complex textures, camera movement and objects, faces and landscapes, and high complexity of encoding. In this work the same test sequences of 10 seconds used in previous tests of VQEG [14][15] were adopted.

To validate results obtained with the image sensor, a complete reference system VQM standardized by ANSI and ITU-T [8][16][17] was used. Figure 1 shows a correlation of 0.79 between the results obtained with the VQM and our proposed system, which indicates a good starting result.

IV. CONCLUSIONS

We have presented an overview of an APS sensor and processing device to evaluate video quality in mass distribution networks. This simplified test approach has a good correlation with the model of reference VQM.

The processing of the captured image generates a metric that has a good correlation with the VQM model. This application is near real time, has low computational complexity, which makes it economically feasible for the evaluation of quality in massive distribution networks, with quality meters placed close to users, in order to evaluate the quality of the images provided to them.

Further work will include new test patterns and more robust pattern modulation of the video under test, improving the quality and identification of artifacts responsible for the degradation of the video.

ACKNOWLEDGMENTS

The authors acknowledge the support of CNPq (Brazilian Council for Science and Technology Development) and NAMITEC Millenium Institute.

REFERENCES

[1] A. Pessoa, A. Falcão, A. Faria, and R. Lotufo, "Video quality assessment using objective parameters based on image segmentation," *IEEE Int. Telecommunications Symposium 98*, 1, Brazil, 1998, pp. 498-503.

[2] S. Wolf and M. Pinson, "Video quality measurement techniques," NTIA Report 02-392, June 2002. Available: www.its.bldrdoc.gov/n3/video/documents.htm

[3] ITU-R Recommendation BT.601, "Encoding parameters of digital television for studios," International Telecommunication Union.

[4] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," International Telecommunication Union.

[5] ITU-T Recommendation J.143, "User requirements for objective perceptual video quality measurements in digital cable television," International Telecommunication Union.

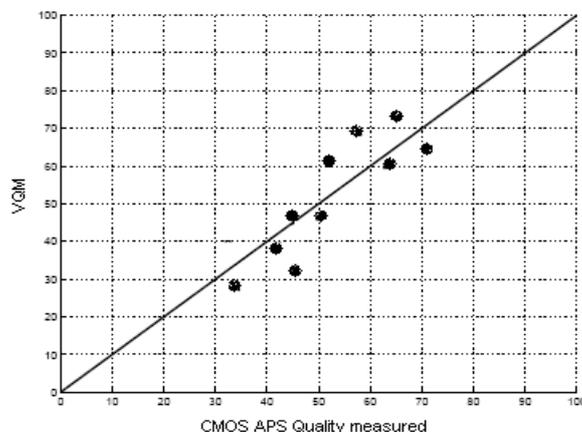


Figure 1 – Results from VQM versus CMOS APS Video Quality Device.

[6] ITU-T Recommendation P.930, "Principles of a reference impairment system for video," International Telecommunication Union, 1996.

[7] ITU-T Recommendation BT.500-8, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, 1998.

[8] ANSI T1.801.03 – 2003, "American National Standard for Telecommunications – Digital transport of one-way video signals – Parameters for objective performance assessment," American National Standards Institute, 2003.

[9] H. T. Sencar, M. Ramkumar, A. N. Akansu, "Data hiding fundamentals and applications – Content security in digital multimedia," Elsevier Academic Press, 2004.

[10] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3/4, pp. 313–336, 1996.

[11] Fabien A. P. Petitcolas, Ross J. Anderson and Markus G. Kuhn, "Information hiding-a survey. Proceedings of the IEEE," July 1999 Special Issue "Identification and protection of multimedia information".

[12] Brazil Namitec Project, 2008. Available <http://ccspsc-010.ccs.unicamp.br/namitec/item.php?id=102#>.

[13] L. H. Braga; S. Domingues; M. F. Rocha; L. B. Sá; F. S. Campos; F. V. Santos; A. Mesquita; M. V. Silva; J. W. Swart, "Layout techniques for radiation hardening of standard CMOS active pixel sensors," *Proceedings of the 20th annual conference on Integrated circuits and systems design*, 2007.

[14] VQEG subjective test plan (phase 1), "Video Quality Experts Group," 1999, available <http://www.vqeg.org>.

[15] VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment." 2000, Available <http://www.vqeg.org>.

[16] S. Wolf, M. Pinson, S. D. Voran, and A. A. Webster, "Objective quality assessment of digitally transmitted video," *Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, Victoria, BC, Canada, pp. 477-82 vol. 9-10 May 1991.

[17] M. Pinson and S. Wolf, "Video quality metric software, Version 2," NTIA Software/Data Product SD-03-396, Volumes 1-5, Oct. 2002. Available: www.its.bldrdoc.gov/n3/video/vqmssoftware.htm

[18] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An objective video quality assessment system based on human perception," *Proc. SPIE Conference on Human Vision, Visual Processing, and Digital Display IV*, San Jose, CA, USA, pp. 15-26, 1-4 February 1993.

D Trabalho IJCA-2013

A seguir é reproduzido o trabalho “Methodology for Objective Evaluation of Video Broadcasting Quality using a Video Camera at the User’s Home” publicado no International Journal of Computer Applications (0975 – 8887) Volume 63– No.16, February 2013.

Methodology for Objective Evaluation of Video Broadcasting Quality using a Video Camera at the User's Home

Marcio L. Graciano
Dep. of Electrical Engineering
University of Brasilia
Campus Darcy Ribeiro,
Brasilia, Brazil

Alexandre R. S. Romariz
Dep. of Electrical Engineering
University of Brasilia
Campus Darcy Ribeiro,
Brasilia, Brazil

Jose Camargo da Costa
Dep. of Electrical Engineering
University of Brasilia
Campus Darcy Ribeiro,
Brasilia, Brazil

ABSTRACT

In this work, a methodology for objective evaluation of the quality of video programs, without reference, recording these programs in the users' residence using a video camera is presented. The methodology is based on the use of a digital watermark embedded in the original program. The watermark is invisible to the user, but capturable by the video camera. The recorded video is handled by specific software that evaluates the watermark degradation. The measure of degradation of this watermark is used to estimate the quality of the video broadcasting system. A case study is presented to validate the methodology. The results of video quality metrics using this methodology were compared to a standardized full reference metrics and the linear correlation between these metrics was superior to 93%, which indicates a high convergence. The result of video quality metrics were also compared to a pixel based difference metrics, PSNR (Peak Signal to Noise Ratio) and the linear correlation was superior to 99%.

Keywords

video quality, quality metrics, human visual system, modulation transfer function

1. INTRODUCTION

The use of digital video has increased in recent years. Although there have been great advances in compression and transmission techniques, impairments are often introduced along the several stages of a communication system. The visibility and annoyance of these impairments are directly related to the quality of the received/ processed video. For many applications, such as broadcasting, it is important to have a good estimate of the quality of the material being received [10][11].

There is an ongoing effort to develop video quality metrics that are able to detect impairments and estimate their annoyance as perceived by human viewers [1][5][14][16][18][21][23][24]. Most successful video quality metrics are Full Reference (FR) metrics. These metrics estimate the quality of a video by comparing original and impaired videos. Requiring the reference video becomes a serious impediment in many real-time applications. In these cases, it becomes essential to develop ways of blindly estimating the quality of a video using a No-Reference (NR) video quality metric. NR metrics, unfortunately, has a lower performance than FR what makes their use in real applications quite difficult [4].

One possible approach to estimate the quality of video signals without requiring the reference is to use a data hiding or watermarking system. In this approach, a digital mark is embedded into the original video frames before the compression and transmission stages. At the receiver, the mark is extracted and a measure of the degradation of the mark is used to estimate the quality of the test video. This type of metric has the advantage of being fast and not requiring the use of the original video [4]. In this work, our goal is to develop a methodology to implement an objective quality metric based on a watermarking technique [8].

Our approach is different than other approaches in the literature [1][5][14][15][16][18][21][23][24] in the following aspects: First, we acquire the watermarked video using a simple consumer electronic video camera placed in the room where the video is being played (i.e. the video screen is filmed with the video camera in order to capture the broadcasted program). Then, the captured video is processed in order to recover the watermark. Afterwards, a quality metrics function analyzes and determines the watermark level of degradation.

The video quality verification will be done as consumers evaluate quality at their homes using their human visual system (HVS). The methodology was developed in order to replace HVS by the camera and the video processing unit. The evaluation of the quality from this HVS-like system will be done using the contrast sensitivity function (CSF) of the human eye and the Modulation Transfer Function (MTF) of the optical system. Each consumer can verify quality in their environment without interfering with the broadcaster distribution equipment.

This approach has minimum interference in distribution systems and has several advantages when we take into account the heterogeneous environment of broadcasting networks. These networks have a wide range of equipment and technologies used in the distribution of video content.

In order for the metrics to accurately evaluate quality, it is necessary to compensate for different users' environment parameters using the MTF.

2. QUALITY METRICS DEFINITION

In this work, a digital mark is embedded into the reference video frames before the compression and transmission stages. At the receiver, the mark is extracted and a measure of the degradation of the mark is used to estimate the quality of the video received. These operations are done in a processing unit containing a set of programs that extract, decode and analyze

the mark in order to evaluate its quality. These programs were entirely developed in this work.

In this work, those programs run on a PC and the video camera is a model currently available in the consumer market. The processing tasks can be eventually made by a compatible platform (a mobile device like a smart phone or a dedicated device with processing capabilities and a video camera).

The following steps describe the process to extract and decode the digital watermark [13][19]:

—At the reception environment of the consumer, a video camera films the screen where the user watches the video and generates a RAW file.

—The frames of RAW file are separated, generating a frame sequence.

—Each frame is then normalized to a fixed size [13].

Some performance tests were done to choose the best normalization frame size to decode the mark. The performance parameters were: small processing time, low processing and ability to decode the mark using several frames resolutions. The best size for use in SD (standard definition) and HD (high definition) video was 512x512.

—The normalized frames were processed by edge detection. This process finds several areas where the marks could have been inserted [13].

—Each area is tested by a watermarking detection algorithm [13].

The video quality measure is performed by correlating the watermark retrieved at the user's environment and the watermark inserted in the content producer (Eq. (1) [13]).

$$N_c = \frac{\sum_{i=1}^n w(i)\hat{w}(i)}{\sqrt{\sum_{i=1}^n w^2(i)}\sqrt{\sum_{i=1}^n \hat{w}^2(i)}} \quad (1)$$

where N_c is the normalized correlation between marks, n is the mark length, w is the mark inserted during video production and \hat{w} is the mark recovered in the consumer environment. In this work 64 bit vectors were chosen as marks (after some preliminary tests with other vector lengths).

The correlation threshold ($N_{cT} = 0.6$) was chosen so that the mark was invisible to the human observer and identifiable by the processing algorithm. The choice was based on subjective tests [10]. A normalized correlation N_c above N_{cT} indicates the watermark presence.

—The highest N_c value from each area in a given frame becomes N_{cframe} .

—A fixed number of video frames (Num_{frames}) was chosen to calculate the overall N_{cvideo} .

The choice of a large number of frames increases the accuracy of the watermark retrieval. A large number of frames also cause the processing to become slower. An optimum value of Num_{frames} equal to 50 was chosen.

The N_{cframe} values were plotted against the corresponding Num_{frames} . A normal distribution was fitted to the obtained curve. This step removes the N_{cframe} outliers and the 95% percentile is then calculated resulting in the N_{cvideo} .

—This N_{cvideo} is then converted to the same scale of SQF (Subjective Quality Factor) [12]. Eq. (2) shows this scale conversion:

$$N_{cvideoQ} = 100 \times \frac{N_{cvideo}^{0.6}}{(N_{cvideo} - 0.4) \times 0.6} \quad (2)$$

The SQF (subjective quality factor) Eq. (3) [12] is calculated using the MTF_T from Eqs. (5) and (6) presented in next section. The SQF is an objective value related with human eye CSF. The SQF was empirically verified in a well conducted observer study of perceived sharpness [9][12].

$$SQF = 100 \times \frac{\int_{v^-}^{v^+} MTF_T(v_d) d(\ln v_d)}{\int_{v^-}^{v^+} d(\ln v_d)} \quad (3)$$

where v_d is the spatial (position in space) frequency in cycles per degree at the retina and the limits of integration are $v^- = 3$ and $v^+ = 12$ cycles per degree [12].

—Using the $N_{cvideoQ}$ value calculated from the watermark recovery Eq. (2) and the SQF value calculated using the optical system MTF Eq. (3), the overall quality metrics FQ Eq. (4) can be evaluated.

$$FQ = a \times N_{cvideoQ} + b \times SQF \quad (4)$$

To find the coefficients a and b , it was performed a nonlinear least-squares data fitting of Eq. (4) to the differential mean opinion scores (DMOS) obtained from subjective tests conducted [10]. The values calculated are $a = 0.501$ and $b = 0.389$.

These values were calculated using the DMOS values according to quality metrics definition methodology steps [6].

3. OPTICAL SYSTEM MTF

The optical system MTF combines all components in the optical path of the system, among which we can highlight the image sensor, the lens, the distances involved between the components and the target image displayed on the video monitor [2]. Figure 1 shows a typical diagram for the optical system MTF. Eq. (5) shows the MTF calculation for the diagram in frequency domain.



Fig. 1: Optical system MTF diagram

$$MTF_T = MTF_{monitor} \times MTF_{env} \times MTF_{len} \times MTF_{sensor} \quad (5)$$

where MTF_T is the total optical system MTF, $MTF_{monitor}$, MTF_{env} , MTF_{len} and MTF_{sensor} are the MTFs due to monitor, environment, camera lens and camera sensor respectively.

Table 1. Camera and monitor parameters

	Parameter	Value
Camera Canon D10	sensor size	6.17x4.55mm
	pixels quantity	4000x3000
	pixels size	1.54x1.52µm
Monitor Samsung P2270HN	diagonal size	22" (55.88cm)
	pixels quantity	1920x1080
	pixels size	0.248x0.248mm
Environment	visualization distance	100cm

The frequency domain is chosen, because the transfer functions' theory for optical systems can be used, which

makes the MTF calculation simpler and faster [2][7]. In the frequency domain (Fourier transform), the product of the MTFs of the various systems' components gives the optical system MTF.

From Eq. (5) we derive a simplified equation taking in account a term due to MTF_{env} and other due to MTF_{elect} relating to all electronic equipment used, Eq. (6).

$$MTF_T = MTF_{elect} \times MTF_{env} \quad (6)$$

where $MTF_{elect} = MTF_{monitor} \times MTF_{lens} \times MTF_{sensor}$, from Eq. (5).

Using the MTF_T obtained with the slanted-edge method [3], it is possible to obtain the values of the SQF from Eq.(3) and the FQ from Eq. (4). The MTF_T allows the calculation of MTF_{elect} values or MTF_{env} values if we fixed one of them while calculating the other.

The $MTF_{elect} = MTF_{elect1}$ fixed value can be calculated, because all electronic equipment do not change. Estimation of MTF_{elect1} could be done averaging measurements of MTF_T while maintaining the value MTF_{env} constant for the test environment.

MTF_T and FQ measurements were done with equipment setup presented in Table 1, where the video camera was mounted fixed on a tripod at distance of 100 cm from the video monitor.

4. RESULTS AND DISCUSSION

In this section we present a case study for validation of the methodology.

In this work we are keeping the camera and monitor parameters fixed as shown in Table 1. Those camera's and monitor's parameters results in MTF_{elect1} to be used to calculate the MTF_{env} .

Three different video sequences were selected (cactus, crowd and BQTerrace). These sequences are commonly used in video quality tests [20]. The watermarks were inserted and then, from each watermarked sequence, other sequences were generated applying degradations indicated in column 3 of Table 2. After that step, seventeen (17) test video sequences were available.

Using these sequences and the same equipment of Table 1 the new MTF_{env} can be calculated from the constant MTF_{elect1} previously calculated.

The evaluation of the video quality metrics was done comparing the results obtained by our metrics with another standardized FR metrics known as VQM (Video Quality Metric) [17] available in the literature. The graphic in figure 2 and Table 2 show the correlation between our metrics FQ and the VQM metrics.

Each numbered point in the graphic of Figure 2 represents a pair relating VQM quality to FQ quality for the same video sequence from Table 2. The straight line represents the correlation least-squares line. The metrics VQM and FQ were normalized to show results from 0 to 1, where values presented near 0 means bad quality and values near 1 means excellent quality for both metrics. The distribution of points on the graph is compatible with the types of degradation presented in Table 2. The videos encoded with higher bit rates of 30Mbps and 10Mbps have better quality while others encoded with lower bit rates or with some kind of packet loss have lower quality. These quality values are in agreement with the subjective tests carried out according to [10]. The

results of VQM metrics for videos which present the same type of degradation have values very close with little excursion on the horizontal axis of Figure 2.

The points 3 and 4 plotted by VQM do not match quality by degradation as the others points in Figure 2. The quality of video coded with low bit rate should be plotted to the left. The FQ metrics plots those points in low quality region at the bottom, in accordance with another videos coded with low bit rate.

Our FQ metrics provides a little variation compared to VQM (more excursion on the vertical axis of Figure 2). This range of values follows SQF quality values [9]. The FQ results obtained are consistent with the different types of degradation shown in Table 2. Through our experiment it was demonstrated that our proposed metrics is closely related to VQM in terms of prediction accuracy. A high correlation exceeding 93% (linear correlation, $r=0.9388$) was observed between these two metrics.

The graphic of Figure 3 represents a pair relating PSNR to FQ quality for the same video sequence from Table 2. The linear correlation of our metrics FQ with PSNR is equal to $r=0.9908$.

5. CONCLUSIONS

In this work, a new methodology using optical characteristics of the user environment to evaluate video quality based on a watermarking technique was proposed.

The methodology usage was checked presenting a case study for NR video quality metrics. Our proposed methodology using optical parameters allows utilization in various user environments and technologies of video broadcasting transmission and distribution. The performance of the objective video quality metrics was evaluated comparing with other standardized metrics [17] and the bit difference metrics PSNR.

A correlation of 99% between our metrics and PSNR and a correlation of 93% between our metrics and VQM were attained. As a result one can see that our no-reference metrics was successfully used in place of a full-reference one, which implies that a straight-ahead, almost real-time, low cost, high quality video evaluation methodology was developed and is now available.

A further refinement to improve robustness can consider using quaternion Fourier domain for calculation of watermarks [22].

6. ACKNOWLEDGEMENTS

The authors acknowledge the support of CNPq (Brazilian Council for Science and Technology Development) and INCT-NAMITEC.

7. REFERENCES

- [1] S. Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. In Andrew B. Watson, editor, Digital Images and Human Vision, pages 179–206, Cambridge, Massachusetts, 1993. MIT Press.
- [2] J. B. DeVelis and G. B. Parrent. Transfer function for cascaded optical systems. *J. Opt. Soc. Am.*, 57:1486–1490, 1967.
- [3] M. Estriebeau and P. Magnan. Fast mtf measurement of cmos imagers using iso 12233 slanted edge methodology. In *SPIE Detectors and Associated Signal Processing*, volume 5251, pages 243–251, 2004.

- [4] M. C. Q. Farias, M. Carli and S. K. Mitra. Objective video quality metric based on data hiding. *IEEE Transactions on Consumer Electronics*, 51:983–992, 2005.
- [5] M. C. Q. Farias and S. K. Mitra. No-reference video quality metric based on artifact measurements. In *IEEE International Conference on Image Processing*, volume 3, pages 141–144, 2005.
- [6] M. C. Q. Farias and S. K. Mitra. A methodology for designing no-reference video quality metrics. In *Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pages 1–6, 2009.
- [7] J.W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill Physical and Quantum Electronics Series, 1968.
- [8] M. L. Graciano, A. R. S. Romariz and J. C. Costa. Cmos image sensor device for objective evaluation of video quality in mass distribution networks. In *IEEE 7th Consumer Communications and Networking Conference (CCNC)*, pages 1–2, 2010.
- [9] E. M. Grainger and K. N. Cupery. An optical merit function (sqf) which correlates with subjective image judgments. *Photographic Science and Engineering*, 16:221–230, 1972.
- [10] ITU-R. Recommendation BT.500-13, chapter Methodology for the subjective assessment of the quality of television pictures. Recommendations of the ITU, Radiocommunication Sector, 2012.
- [11] ITU-T. Final report from the video quality experts group (VQEG) on the validation of objective models of video quality assessment, volume 4, chapter COM 9-80-E, approved for release at VQEG meeting, 2000.
- [12] B. W. Keelan. Objective and subjective measurement and modeling of image quality: a case study. In *SPIE Applications of Digital Image Processing XXXIII*, volume 7798, pages 779–815, 2010.
- [13] L. Li, B. Guo and L. Guo. Rotation, scaling and translation invariant image watermarking using feature points. *The Journal of China Universities of Posts and Telecommunications*, 15:82–87, 2008.
- [14] W. Lin and C.C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312, 2011.
- [15] H. Loukil, M. H. Kacem and M. S. BouhleL. A new image quality metric using system visual human characteristics. *International Journal of Computer Applications*, 60(6):32–36, 2012.
- [16] A. K. Moorthy and A.C. Bovik. Visual quality assessment algorithms : What does the future hold? *International Journal of Multimedia Tools and Applications, Special Issue on Survey Papers in Multimedia by World Experts*, 51(2):675–696, 2011.
- [17] M. Pinson and S. Wolf. *Video Quality Measurement Users Manual*. NTIA Handbook HB-02-01, 2002.
- [18] M.H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50:312–322, 2004.
- [19] S. Poongodi and B. Kalaavathi. Comparative study of various transformations in robust watermarking algorithms. *International Journal of Computer Applications*, 58(11), 2012.
- [20] F. De Simone, L. Goldmann, J.S. Lee and T. Ebrahimi. Towards high efficiency video coding: Subjective evaluation of potential coding technologies. *Journal of Visual Communication and Image Representation*, 22(8):734–748, 2011.
- [21] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment - Phase II. Tech. Report, 2003.
- [22] X. Wang, C. Wang, H. Yang, and P. Niu. A robust blind color image watermarking in quaternion fourier transform domain. *Journal of Systems and Software*, 86(2):255–277, 2013.
- [23] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 2004.
- [24] S. Winkler and P. Mohandas. The evolution of video quality measurement: From psnr to hybrid metrics. *IEEE Transactions on Broadcasting*, 54:660–668, 2008.

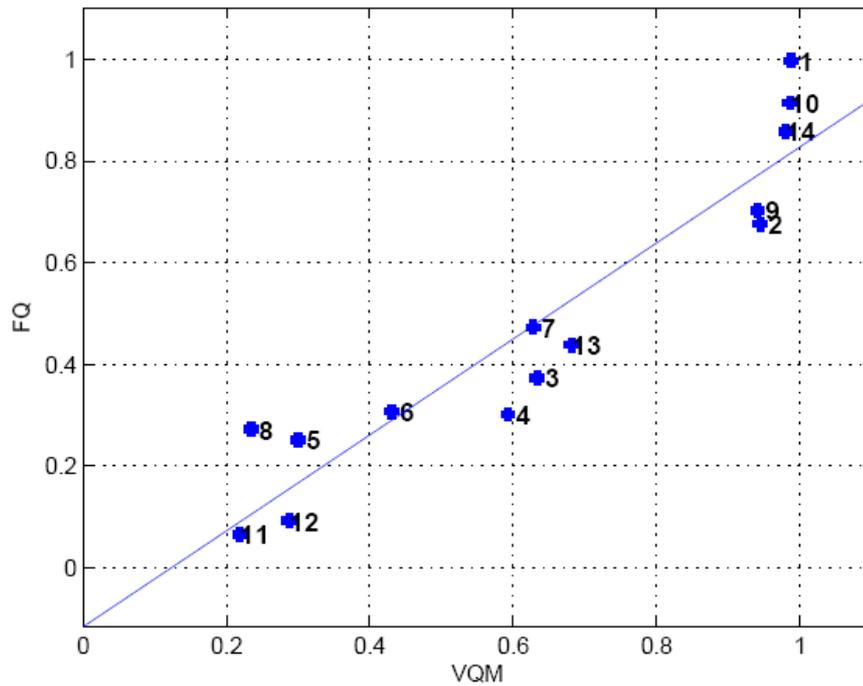


Fig 2: FQ Metrics results compared with VQM [17]

Table 2. Metrics Results used to generate Figure 2 and Figure 3

Sequence	video name	degradation	PSNR	VQM	FQ
1	BQTerrace	video coded H.264 bitrate 30Mbps	48.947	0.987	0.9977
2	BQTerrace	video coded MPEG2 bitrate 10Mbps	46.6079	0.9447	0.6756
3	BQTerrace	video coded MPEG2 bitrate 1Mbps	32.4961	0.6336	0.3734
4	BQTerrace	video coded H.264 bitrate 300kbps	31.6951	0.5927	0.301
5	cactus	video coded H.264 packet loss rate 10%	30.5056	0.3004	0.251
6	cactus	video coded MPEG bitrate 1Mbps	33.6848	0.4307	0.3071
7	cactus	video coded H.264 packet loss rate 1%	37.1515	0.6281	0.4729
8	cactus	video coded H.264 bitrate 300kbps	30.8936	0.2345	0.2718
9	cactus	video coded MPEG bitrate 10Mbps	43.3981	0.9408	0.7022
10	cactus	video coded H.264 bitrate 30Mbps	49.4828	0.9858	0.9142
11	crowdrun	video coded H.264 bitrate 300kbps	25.7758	0.2186	0.065
12	crowdrun	video coded H.264 packet loss rate 10%	27.0725	0.2876	0.0918
13	crowdrun	video coded H.264 packet loss rate 1%	34.1506	0.6813	0.4378
14	crowdrun	video coded H.264 bitrate 30Mbps	45.1649	0.9797	0.8584

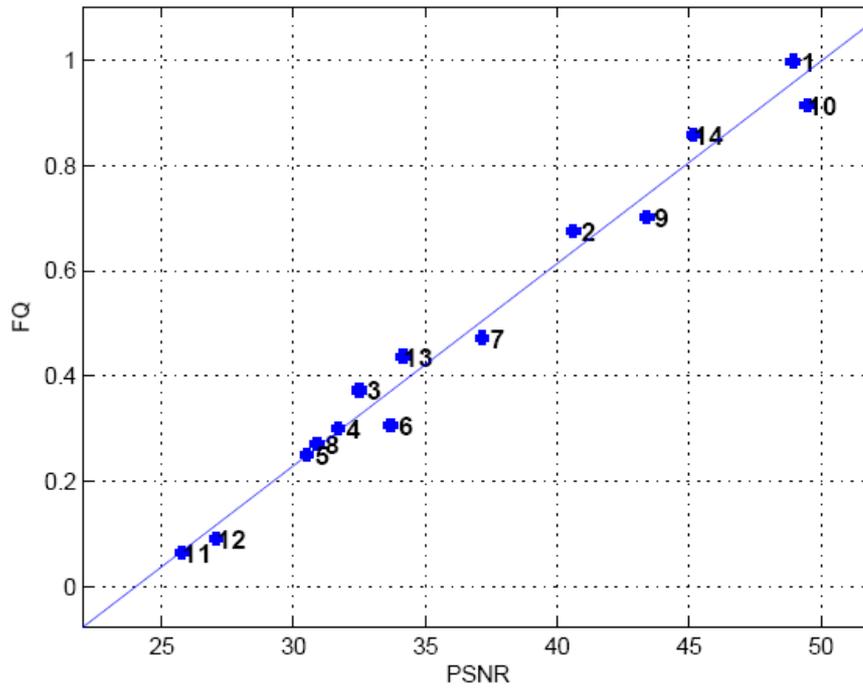


Fig 3: FQ Metrics results compared with PSNR