

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

**DIOGO AUGUSTO BARROS PEREIRA**

**AVALIAÇÃO OBJETIVA DA QUALIDADE DE VÍDEO BASEADA NA  
RELEVÂNCIA DOS OBJETOS NOS QUADROS**

**DISSERTAÇÃO**

**CURITIBA**

**2011**

**DIOGO AUGUSTO BARROS PEREIRA**

**AVALIAÇÃO OBJETIVA DA QUALIDADE DE VÍDEO BASEADA NA  
RELEVÂNCIA DOS OBJETOS NOS QUADROS**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Mestre em Ciências” – Área de Concentração: Telemática.

Orientador: Alexandre de Almeida Prado Pohl

**CURITIBA**

**2011**

---

Dados Internacionais de Catalogação na Publicação

---

P436 Pereira, Diogo Augusto Barros  
Avaliação objetiva da qualidade de vídeo baseada na relevância dos objetos nos quadros / Diogo Augusto Barros Pereira. — 2011.  
111 p. : il. ; 30 cm

Orientador: Alexandre de Almeida Prado Pohl.

Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. Área de concentração: Telemática, Curitiba, 2011.

Bibliografia: f. 108-111

1. Vídeo digital. 2. Processamento de imagens. 3. Vídeos - Avaliação. 4. Engenharia elétrica – Teses. I. Pohl, Alexandre de Almeida Prado, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. III. Título.

CDD (22. ed.) 621.38833

---

Biblioteca Central da UTFPR, Campus Curitiba


**Título da Dissertação Nº 579:**


**“Avaliação Objetiva da Qualidade de Vídeo Baseada  
na Relevância dos Objetos nos Quadros”**

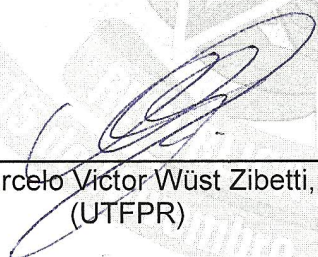
por

**Diogo Augusto Barros Pereira**


Esta dissertação foi apresentada como requisito parcial à obtenção do grau de MESTRE EM CIÊNCIAS – Área de Concentração: Telemática, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI – da Universidade Tecnológica Federal do Paraná – UTFPR – Campus Curitiba, às 9h do dia 28 de setembro de 2011. O trabalho foi aprovado pela Banca Examinadora, composta pelos professores:

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

  
Prof. Alceu de Souza Britto Junior, Dr.  
(PUCPR)

  
Prof. Marcelo Victor Wüst Zibetti, Dr.  
(UTFPR)

Visto da Coordenação:

  
Prof. Fábio Kurt Schneider, Dr.  
(Coordenador do CPGEI)

à Deus...

à minha esposa Bianca...

à minha família...

aos meus amigos...

## AGRADECIMENTOS

Inicialmente, agradeço à Deus por todas as graças recebidas, pela dádiva de poder concluir mais uma importante fase em minha vida, sob sua proteção e cuidado, não deixando em tempo algum o desânimo me abater.

Agradeço à minha esposa Bianca pela paciência, carinho e conforto em todos os momentos. Quando eu mais precisei, ela sempre esteve ao meu lado. Agradeço em especial a minha família, Ana Maria (mãe), Amarildo (pai), Mariana (irmã) e ao Tiago (irmão) por todo apoio e força, sempre com uma voz acolhedora, um ombro amigo, me motivando e dando forças para alcançar meus objetivos pessoais. A minha tia Terezinha que sempre rezou e torceu por meu sucesso. Ao meu tio Carlos, pelo meu primeiro computador.

Agradeço a todos que passaram por minha vida nestes dois anos e meio. Ao pessoal do laboratório LCD por todos os momentos bons e ruins, todas as piadas contadas e os momentos de *stress*. Não poderia esquecer de agradecer também a equipe de desenvolvimento DIDES-E1 da empresa Celepar por todo o apoio e motivação. Aos meus amigos Diego e Marcos pelas calorosas discussões sobre os diferentes trabalhos de mestrado na mesa do *shopping*. A amiga e colega de trabalhos Cristina, por todo auxílio e companheirismo nos diversos trabalhos feitos em conjunto.

Aos professores pelo conhecimento repassado, em especial, ao professor Alexandre Pohl (orientador) que transformou o estudante de iniciação científica em pesquisador e deu oportunidade para meu crescimento pessoal. Ao professor Paulo Stadzisz (Engenharia de Software - UTFPR) que quebrou os paradigmas e me ensinou o verdadeiro significado desta tão abrangente área. A professora Aurora Pozo (Inteligência Artificial - UFPR) pelo esforço e dedicação na correção dos artigos publicados. A professora Anelise (Redes Sem-Fio) pela excelente aula. Ao professor Hugo Vieira (Metodologia Científica - UTFPR) por me ensinar o significado de ciência, de pesquisa e o auxílio na utilização do *Latex* para confecção de trabalhos científicos. A professora Tânia Mezzadri (PDI - UTFPR) pela introdução de conceitos na área de pesquisa.

Também o meu agradecimento à CAPES, pela concessão da bolsa de mestrado.

Obrigado, obrigado e obrigado!

Tu te tornas eternamente responsável por aquilo que cativas.

(O Pequeno Príncipe - Antoine de Saint-Exupéry)

## RESUMO

PEREIRA, Diogo Augusto Barros. AVALIAÇÃO OBJETIVA DA QUALIDADE DE VÍDEO BASEADA NA RELEVÂNCIA DOS OBJETOS NOS QUADROS. 111 f. Dissertação – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2011.

Esta dissertação apresenta o estudo, a melhoria e a implementação de um algoritmo para avaliar a qualidade de vídeo baseado na relevância dos objetos nos quadros. A métrica usada para estimar a relevância tem como pré-requisito a segmentação do vídeo. O trabalho foi dividido em etapas distintas: (1) segmentação, (2) avaliação da segmentação e estimação da relevância, (3) geração de artefatos nos vídeos originais e (4) avaliação da qualidade de vídeo. Experimentos com diferentes artefatos comprovam a eficiência da métrica proposta em termos de correlação com a métrica subjetiva.

**Palavras-chave:** Vídeo Digital, Processamento de Imagens, Avaliação de Vídeo.



## ABSTRACT

PEREIRA, Diogo Augusto Barros. OBJECTIVE EVALUATION OF VIDEO QUALITY BASED ON THE RELEVANCE OF OBJECTS IN FRAMES. 111 f. Dissertação – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2011.

This work presents the study, improvement and implementation of an algorithm to evaluate the video quality based on the relevance of objects in the frames. The measurements used to estimate the relevance requires a segmentation of objects. The work was divided into the following sections: (1) segmentation, (2) evaluation of segmentation and estimation of relevance, (3) generation of artifacts in video original and (4) evaluation of video quality. Experiments with different artifacts prove the efficiency of the proposed metric in terms of correlation with subjective measures.

**Keywords:** Digital Video, Image Processing, Video Evaluation.

## LISTA DE FIGURAS

FIGURA 1	–	Representação dos principais elementos do olho humano. ....	18
FIGURA 2	–	Detalhes das camadas da retina .....	18
FIGURA 3	–	Sensibilidade Espectral dos Cones .....	19
FIGURA 4	–	Sensibilidade Espectral dos Bastonetes .....	20
FIGURA 5	–	Exemplificação do contraste simultâneo. ....	20
FIGURA 6	–	Representação do efeito de bandas de Mach .....	21
FIGURA 7	–	Ilustração do ruído gaussiano .....	22
FIGURA 8	–	Ilustração do ruído sal-e-pimenta .....	23
FIGURA 9	–	Representação de borramento. ....	24
FIGURA 10	–	Representação de blocagem. ....	25
FIGURA 11	–	Tipos de vizinhança .....	26
FIGURA 12	–	Exemplo de borda .....	27
FIGURA 13	–	Exemplos de elementos estruturantes. ....	28
FIGURA 14	–	Erosão .....	29
FIGURA 15	–	Exemplo de dilatação. ....	30
FIGURA 16	–	Resultado da etapa DifAR. ....	36
FIGURA 17	–	Resultado da etapa ERE. ....	37
FIGURA 18	–	Resultado da etapa ERT. ....	38
FIGURA 19	–	Resultado da etapa IR. ....	39
FIGURA 20	–	Exemplos de PSNR .....	47
FIGURA 21	–	Diagrama do SSIM .....	47
FIGURA 22	–	Tela Principal do Sistema de Avaliação. ....	53
FIGURA 23	–	Quadro 1 da sequência de vídeo Akiyo .....	54
FIGURA 24	–	Algoritmo para calcular a área .....	55
FIGURA 25	–	Exemplo de área .....	55
FIGURA 26	–	Exemplo de perímetro .....	56
FIGURA 27	–	Algoritmo para calcular a espessura .....	57
FIGURA 28	–	Elemento Estruturante utilizado no trabalho. ....	58
FIGURA 29	–	Algoritmo para calcular a espessura .....	58
FIGURA 30	–	Exemplo de espessura .....	59
FIGURA 31	–	Imagens utilizadas no experimento .....	60
FIGURA 32	–	Resultado da Segmentação .....	62
FIGURA 33	–	Diagrama do sistema de avaliação da qualidade de segmentação. ....	63
FIGURA 34	–	Algoritmo do Filtro de Sobel .....	67
FIGURA 35	–	Quadros com ruído branco gaussiano .....	74
FIGURA 36	–	Quadros com ruído sal-e-pimenta .....	76
FIGURA 37	–	Quadros com borramento .....	78
FIGURA 38	–	Quadro com blocagem .....	79
FIGURA 39	–	Algoritmo do MSE .....	80
FIGURA 40	–	Algoritmo do PSNR .....	80
FIGURA 41	–	Algoritmo do PSNR_Mod .....	82
FIGURA 42	–	Diagrama do sistema de avaliação da qualidade de vídeo. ....	82

FIGURA 43	–	Imagens utilizadas no experimento .....	85
FIGURA 44	–	Separação dos objetos na cena .....	86
FIGURA 45	–	Avaliação Independente .....	87
FIGURA 46	–	Adição de áreas do fundo no objeto 1 .....	88
FIGURA 47	–	Relevância individual dos objetos. ....	89
FIGURA 48	–	Relevância contextual dos objetos. ....	89
FIGURA 49	–	Avaliação Global da Segmentação. ....	90
FIGURA 50	–	Degradação com ruído branco gaussiano .....	91
FIGURA 51	–	Degradação com ruído branco gaussiano .....	92
FIGURA 52	–	Degradação com o artefato sal-e-pimenta .....	94
FIGURA 53	–	Degradação com emprego do artefato borramento linear .....	96
FIGURA 54	–	Degradação com emprego do artefato borramento matricial .....	97
FIGURA 55	–	Degradação com emprego do artefato blocagem .....	99
FIGURA 56	–	Degradação com emprego do artefato travamento .....	100
FIGURA 57	–	Correlação entre Av. Objetiva x MOS .....	104
FIGURA 58	–	Relação entre MOS x Predição da Av. Objetiva .....	105

## LISTA DE TABELAS

TABELA 1	– FO para o ACO puro com centros distantes. ....	84
TABELA 2	– FO para o ACO puro com centros aleatórios. ....	84
TABELA 3	– FO para o ACO com Busca Local e centros distantes. ....	84
TABELA 4	– FO para o ACO com Busca Local e centros distantes. ....	85
TABELA 5	– Avaliação Individual da Qualidade de Segmentação. ....	88
TABELA 6	– Avaliação do artefato ruído branco gaussiano ....	93
TABELA 7	– Avaliação do artefato sal-e-pimenta ....	95
TABELA 8	– Avaliação artefato borramento linear ....	98
TABELA 9	– Avaliação do artefato blocagem ....	98
TABELA 10	– Avaliação do artefato travamento ....	100
TABELA 11	– Comparação do desempenho das métricas objetivas. ....	103

## LISTA DE SIGLAS

ACO	Otimização por Colônia de Formigas
AIR	Análise da Intersecção de Regiões Espaço-Temporais
ANSI	<i>American National Standards Institute</i>
CCP	<i>Pearson Linear Correlation Coefficient</i>
CCS	<i>Spearman Rank Order Correlation Coefficient</i>
CIF	<i>Common Intermediate Format</i>
DCT	<i>Discrete Cosine Transform</i>
DifAR	Diferença acumulativa de regiões
EE	Elemento Estruturante
ERE	Eliminação de regiões espacialmente
ERT	Eliminação de regiões espaço-temporalmente
FFT	<i>Fast Fourier Transform</i>
FO	Função Objetivo
FR	<i>full-reference</i>
IDE	<i>Integrated Development Environment</i>
IR	Intersecção entre regiões espaço-temporais
JDK	<i>Java Development Kit</i>
LHS	<i>Local Harmonic Strength</i>
MDRAV	Modelo para Detecção de Regiões de Atenção Visual
MM	Morfologia Matemática
MOS	<i>Mean Opinion Score</i>
nm	Nanômetros
NR	<i>no-reference</i>
PSNR	<i>Peak Signal to Noise Ratio</i>
RER	Registro espacial de regiões
RR	<i>reduced-reference</i>
SDSCE	<i>Single Double Stimulus for Continuous Evaluation</i>
SI	<i>Spatial Perceptual Information</i>
SSIM	<i>Structural SIMilarity</i>
SVH	Sistema Visual Humano
TI	<i>Temporal Perceptual Information</i>
VQEG	<i>Video Quality Experts Group</i>
VQM	<i>Video Quality Metric</i>
VSNR	<i>Visual Signal to Noise Ratio</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>14</b>
1.1 OBJETIVOS	15
1.2 ESTRUTURA DA DISSERTAÇÃO	16
<b>2 REFERENCIAL TEÓRICO</b>	<b>17</b>
2.1 SISTEMA VISUAL HUMANO	17
2.2 ARTEFATOS	21
2.2.1 Ruído Gaussiano	22
2.2.2 Ruído Sal-e-Pimenta	23
2.2.3 Borramento	23
2.2.4 Blocagem	24
2.2.5 Travamento	25
2.3 RELAÇÕES BÁSICAS ENTRE PIXELS	25
2.3.1 Vizinhaça	26
2.3.2 Conectividade	26
2.3.3 Adjacência	27
2.3.4 Borda	27
2.3.5 Morfologia Matemática Binária	27
2.4 SEGMENTAÇÃO	30
2.4.1 Otimização por Colônia de Formigas na Segmentação de Imagens	32
2.4.2 Segmentação utilizando MDRAV	35
2.5 AVALIAÇÃO DA QUALIDADE DE SEGMENTAÇÃO	39
2.6 QUALIDADE DE SEGMENTAÇÃO PELO MÉTODO INDEPENDENTE	41
2.6.1 Métrica Objetiva Independente Para Avaliação da Qualidade de Segmentação	41
2.7 ESTIMAÇÃO DA RELEVÂNCIA DO OBJETO	43
2.8 AVALIAÇÃO DA QUALIDADE DE IMAGENS E VÍDEOS	44
2.8.1 Erro Médio Quadrático	45
2.8.2 Relação Sinal-Ruído de Pico	46
2.9 SSIM	47
2.10 VQM	48
2.11 VSNR	49
2.12 TRABALHOS RELACIONADOS	50
<b>3 METODOLOGIA</b>	<b>52</b>
3.1 MATERIAIS	52
3.2 VÍDEO ORIGINAL UTILIZADO NOS TESTES	53
3.3 DESCRITORES DE REGIÕES	54
3.3.1 Área	54
3.3.2 Perímetro	55
3.3.3 Espessura	57
3.4 SEGMENTAÇÃO	60
3.4.1 Otimização por Colônia de Formigas na Segmentação de Imagens	60
3.4.2 Segmentação utilizando MDRAV	61

3.5	AVALIAÇÃO DA QUALIDADE DE SEGMENTAÇÃO	62
3.5.1	Avaliação Individual	63
3.5.1.1	Métrica Intra-Objeto	64
3.5.1.2	Métrica Inter-Objeto	68
3.5.1.3	Construção da Métrica de Avaliação Individual	68
3.5.2	Avaliação Global	69
3.6	ESTIMAÇÃO DA RELEVÂNCIA DO OBJETO	70
3.6.1	Composição da Métrica para Estimação da Relevância	72
3.7	GERAÇÃO DE VÍDEOS COM ARTEFATOS	73
3.7.1	Ruído Branco Gaussiano	73
3.7.2	Ruído sal-e-pimenta	75
3.7.3	Borramento	76
3.7.4	Blocagem	78
3.7.5	Travamento	79
3.8	AVALIAÇÃO DA QUALIDADE DE VÍDEO	80
3.9	MÉTRICA PROPOSTA	81
<b>4</b>	<b>RESULTADOS</b>	<b>83</b>
4.1	SEGMENTAÇÃO	83
4.1.1	Segmentação utilizando ACO	83
4.1.2	Segmentação por MDRAV	86
4.2	AVALIAÇÃO DA QUALIDADE DE SEGMENTAÇÃO	86
4.3	AVALIAÇÃO DA QUALIDADE DE VÍDEO	90
4.3.1	Ruído Branco Gaussiano	90
4.3.2	Artefato Sal-e-Pimenta	93
4.3.3	Borramento	95
4.3.4	Blocagem	98
4.3.5	Travamento	99
4.4	DISCUSSÃO DOS RESULTADOS	100
4.5	VALIDAÇÃO DA MÉTRICA PROPOSTA	101
<b>5</b>	<b>CONCLUSÃO</b>	<b>106</b>
	<b>REFERÊNCIAS</b>	<b>108</b>

## 1 INTRODUÇÃO

Nas últimas décadas a área de pesquisa em processamento, análise e transmissão de vídeo tem recebido atenção especial por parte dos pesquisadores. Devido ao avanço nas transmissões de conteúdo televisivo, com a exibição de vídeos em alta definição e a apresentação de conteúdos interativos, as técnicas de avaliação para medir a qualidade de vídeo se tornaram imprescindíveis. Por meio da utilização de ferramentas automáticas para o cálculo das avaliações, os usuários conseguem, por exemplo, distinguir os parâmetros de compressão e/ou transmissão que podem ser alterados e visualizar em tempo real o impacto destas mudanças sobre o vídeo.

Além disso, os vídeos podem sofrer diversos tipos de distorções em diferentes etapas (captura, processamento, etc.) resultando em artefatos, degradações e ruídos nos vídeos digitais. Quando isto ocorre, a qualidade do vídeo para o espectador poderá ficar comprometida.

Diante dessa perspectiva, torna-se fundamental o estudo e desenvolvimento de ferramentas que analisem a qualidade do vídeo. Tais ferramentas devem utilizar conceitos do sistema visual humano para medir a qualidade visual, pois são os usuários que determinam subjetivamente se um vídeo tem qualidade boa ou ruim. Porém, o sistema visual humano é complexo e, até o presente momento, nenhum sistema artificial inteligente conseguiu simulá-lo.

Avaliar a qualidade de vídeo tem como pré-requisito a adoção de métricas para capturas as informações do problema específico a ser resolvido. Por isso, torna-se fundamental o uso de métricas de avaliação para inúmeros tipos de aplicação, como por exemplo, o sistema de monitoramento de sinal da TV Digital. Ao verificar a transmissão de sinais de televisão, tanto analógica como digital, é possível visualizar as degradações que ocorrem nos vídeos. Como exemplo, cita-se: borramento e blocagem nos vídeos digitais e chuviscos nos analógicos.

As técnicas de avaliação da qualidade de vídeo são classificadas em: subjetiva e objetiva. A avaliação subjetiva utiliza os observadores humanos para atribuir notas aos vídeos e ainda são consideradas como uma forma confiável de avaliação. Posteriormente, com o valor das notas, é efetuado um cálculo estatístico cujo resultado é conhecido como MOS (*Mean Opinion Score*). Com isso, a avaliação torna-se uma tarefa cansativa, demorada e com alto custo



operacional. Em comparação, sendo considerada como solução destas desvantagens, existe a avaliação objetiva que atribui notas aos vídeos empregando modelos matemáticos que visam estimar a percepção visual de um espectador. Neste sentido, esta avaliação reduz o tempo despendido no processo e, algumas métricas podem ser computadas em tempo real, por se tratar de cálculos com baixo custo computacional. Todavia, várias métricas foram propostas, mas nenhuma, até agora, conseguiu se tornar um padrão de avaliação e substituir a avaliação subjetiva. Como exemplo, cita-se os seguintes trabalhos de métricas propostas (WANG; BOVIK, 2002), (PINSON; WOLF, 2004) e (CHANDLER; HEMAMI, 2007).

Por este motivo é interessante o desenvolvimento de métricas que avaliem de forma independente os diferentes objetos na cena, baseados em sua relevância. Com este algoritmo será possível ponderar de forma correta e eficiente os diferentes objetos atribuindo a eles pesos distintos fundamentados nas características que atraem mais atenção visual. Sendo assim, as métricas objetivas poderão prever o quanto tal distorção incomodará o telespectador ao ver o vídeo.

Este trabalho apresenta um conjunto de métricas que serão compostas em uma avaliação da qualidade de vídeo utilizando o conceito de relevância dos objetos na cena. A partir da imagem original é efetuada uma fase de pré-processamento dividida em duas etapas: 1) segmentação do vídeo; 2) geração do vídeo com artefatos obtidos por um gerador artificial. Os parâmetros de entrada para avaliação serão o vídeo original, o vídeo segmentado e o vídeo com artefatos. Após obter os parâmetros, na segunda fase é calculada a relevância dos objetos nas cenas que serão, posteriormente, utilizados pela métrica proposta como pesos.

## 1.1 OBJETIVOS

A resolução do problema da avaliação da qualidade de vídeo está centrada na utilização de conceitos do sistema visual humano, com o intuito de aumentar a correlação entre uma avaliação feita por um observador humano e uma avaliação objetiva. Com auxílio de métricas objetivas, que utilizam funções matemáticas, tenta-se obter uma nota para o vídeo. Esta tarefa não é fácil, principalmente pela alta complexidade e subjetividade quando o assunto é medir a qualidade visual percebida.

Portanto, o objetivo principal deste trabalho é melhorar uma métrica de avaliação de qualidade de vídeo por meio da adição de pesos ponderadores baseados na relevância dos objetos nos quadros, medida a partir de uma fonte de vídeo segmentado.

Como objetivos específicos:

- Descrever o estado da arte e discutir os conceitos básicos para auxiliar na compreensão das métricas utilizadas;
- Desenvolver e testar algumas técnicas de segmentação;
- Medir a qualidade de segmentação através da implementação de uma métrica proposta na literatura;
- Estimar a relevância dos objetos segmentados nos quadros;
- Utilizar uma métrica comumente usada na literatura para avaliar a qualidade de vídeo;
- Medir a qualidade de vídeo utilizando a métrica modificada ponderada baseada nos valores obtidos pela estimativa de relevância dos objetos;
- Aplicar as avaliações sobre vídeos com diferentes artefatos;
- Validar a métrica proposta comparando-a com os resultados obtidos por outras métricas objetivas e pelo cálculo de correlação entre elas e a avaliação subjetiva.

## 1.2 ESTRUTURA DA DISSERTAÇÃO

Este documento está organizado em 5 capítulos. O Capítulo 2 descreve os conceitos utilizados ao longo de trabalho. Fornece uma explicação sobre o complexo sistema visual humano e discute as métricas encontradas na literatura para avaliação da qualidade de segmentação e de vídeo. O Capítulo 3 descreve os materiais e métodos utilizados para o desenvolvimento do trabalho, desde a escolha da tecnologia utilizada na implementação dos algoritmos até o detalhamento das métricas utilizadas nas avaliações. Para facilitar trabalhos futuros, alguns algoritmos foram representados com o intuito de demonstrar como foram tratadas questões como limites do quadro. O Capítulo 4 apresenta os resultados obtidos tanto pela avaliação da qualidade de segmentação quanto pela avaliação da qualidade de vídeo. Nesse capítulo, foram ilustrados os resultados obtidos pela estimação dos objetos na cena indicando quanto cada objeto influencia o foco de atenção do observador. Para validar a métrica proposta é utilizada uma função logística para mapear os resultados obtidos pelas avaliações objetivas com a avaliação subjetiva. Em seguida, é exibido os valores de correlação entre as métricas objetivas com a método subjetivo, estimados pelos coeficientes de *Pearson* e de *Spearman* (TRIOLA, 2005). Por fim, o Capítulo 5 apresenta as conclusões e descreve os trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

O objetivo deste capítulo é apresentar uma série de conceitos e estudos bibliográficos que foram utilizados ao longo do trabalho.

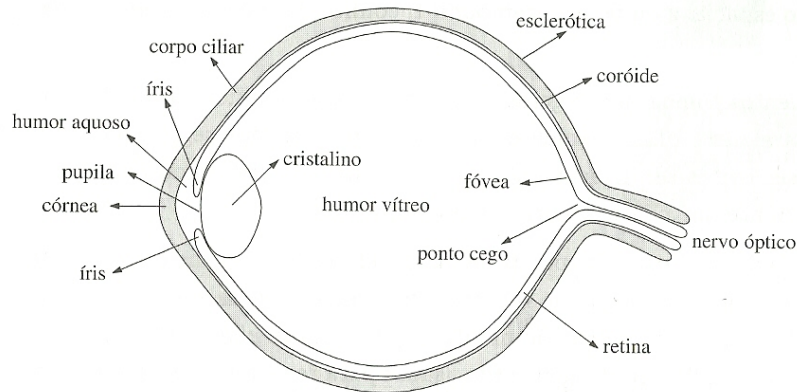
### 2.1 SISTEMA VISUAL HUMANO

Segundo (GONZALEZ; WOODS, 2010), apesar de o processamento digital de imagens e vídeos ser baseado em fórmulas matemáticas e estatísticas, a intuição e a análise humana desempenham um papel central na escolha de uma técnica em comparação com outra. Com esses fatos levantados se torna importante o início desta dissertação com a compreensão básica da percepção visual humana.

Define-se como sistema visual humano, SVH, o sistema responsável pela captação, interpretação e resposta de um observador humano a um estímulo visual. Diante desses argumentos, salienta-se a importância de se estudar o SVH para se entender as medidas de qualidade e relevância de um objeto em uma cena de vídeo e para construção de melhores ferramentas que serão aplicadas no processamento de imagens, em especial, quando o objetivo do processamento for agradar o ser humano.

A visão é uma das mais importantes capacidades sensoriais dos seres humanos. Ela é responsável pela percepção adequada do ambiente e apresenta diversas funções complexas, como por exemplo, reconhecimento, interpretação, localização e detecção de objetos no ambiente (PEDRINI; SCHWARTZ, 2008). O estudo da visão começa com o olho. O olho humano é um órgão sensorial complexo, cuja função é capturar as relações temporais e espaciais de objetos contidos em seu espaço visual, convertendo a imagem luminosa em sinais elétricos que serão processados pelo cérebro. Destaca-se que a entrada do estímulo luminoso acontece no primeiro elemento opticamente ativo no olho, a córnea. A córnea é responsável pela captação e concentração da luz agindo como uma lente e é através dela que acontece a recepção da luz do ambiente (CONCI et al., 2008).

A imagem formada pela captação da imagem de um objeto pelo sistema ocular e projeção sobre a retina é conhecida como imagem de retina (GONZALEZ; WOODS, 2010). A figura 1 representa os elementos principais contidos na óptica ocular.

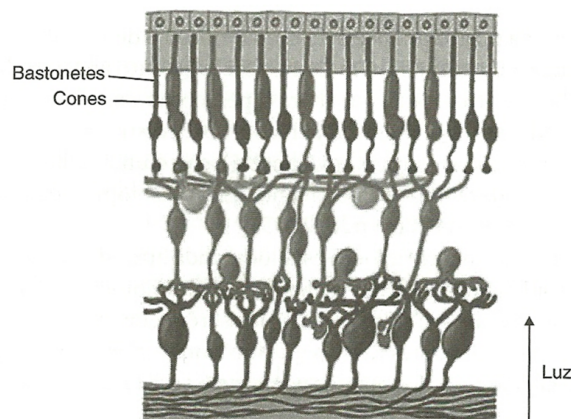


**Figura 1: Representação dos principais elementos do olho humano.**

**Fonte: (PEDRINI; SCHWARTZ, 2008)**

A retina é formada por cerca de cem milhões de sensores e ocupa uma área de aproximadamente mil nm. Cada sensor da retina transforma somente uma parte do estímulo luminoso da imagem para que o cérebro processe o sinal elétrico. Trata-se de uma membrana revestida por uma camada de tecidos nervosos, responsável pela codificação da imagem em sinais nervosos e a transmissão para o cérebro (CONCI et al., 2008).

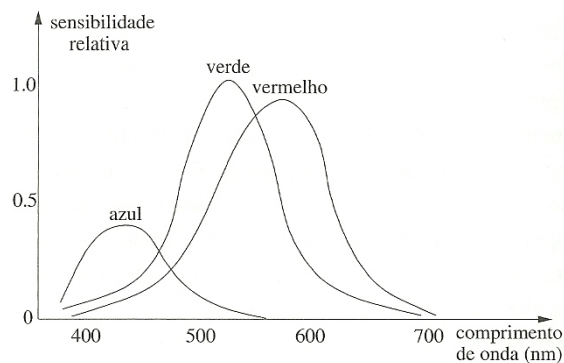
A retina é composta por células sensíveis à luz, entre elas, os cones e os bastonetes (PEDRINI; SCHWARTZ, 2008). Os cones e os bastonetes são representados na figura 2.



**Figura 2: Detalhes das camadas da retina**

**Fonte: (CONCI et al., 2008)**

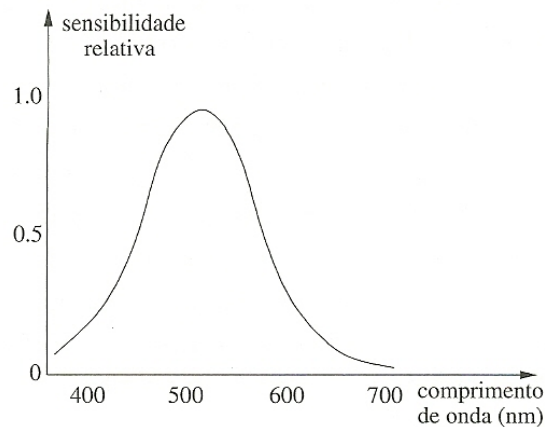
Os cones são responsáveis pela capacidade do olho em diferenciar os detalhes nas imagens e são altamente sensíveis à cor (PEDRINI; SCHWARTZ, 2008). Esta diferenciação de detalhes ocorre porque cada um dos cones está conectado à sua própria terminação nervosa. Os cones são em número de seis a sete milhões em cada olho e sua localização é na parte frontal do fundo do olho. Existem três tipos de cones e cada um deles é estimulado de acordo com o comprimento de onda das cores (CONCI et al., 2008). Os valores de máxima sensibilidade ocorrem nos comprimentos de onda de 430 (azul), 575 (verde) e 600 (vermelho) *nm*, e estão ilustrados na figura 3. A visão pelos cones é chamada de visão de luz clara ou visão fotópica (GONZALEZ; WOODS, 2010).



**Figura 3: Sensibilidade Espectral dos Cones**

**Fonte: (PEDRINI; SCHWARTZ, 2008)**

Os bastonetes são mais sensíveis à baixa intensidade de luz. Em comparação com os cones, os bastonetes estão em grande vantagem em relação ao seu número, cerca de setenta e cinco a cento e cinquenta milhões, distribuídos na superfície periférica da retina (PEDRINI; SCHWARTZ, 2008). Pelo fato de vários bastonetes estarem conectados a uma única terminação nervosa e distribuídos em uma ampla área, a quantidade de detalhes discerníveis por eles é reduzida. Destaca-se que os bastonetes não estão envolvidos na visualização de cores. O valor de máxima sensibilidade ocorre no comprimento de onda de 500 *nm* como pode ser visto na figura 4. A visão pelos bastonetes é chamada de escotópica ou visão da luz escura (GONZALEZ; WOODS, 2010).

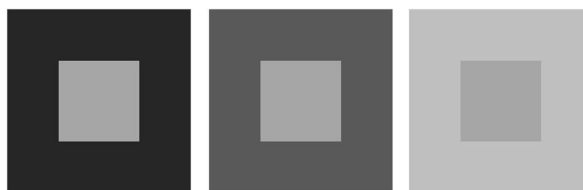


**Figura 4: Sensibilidade Espectral dos Bastonetes**

**Fonte: (PEDRINI; SCHWARTZ, 2008)**

O olho humano não responde de forma linear à intensidade luminosa e isto faz com que a sensibilidade dos olhos para diferenciar as intensidades não dependa somente da diferença de intensidade (WU; RAO, 2006). Para exemplificar isto, existem dois fenômenos que representam claramente que o brilho percebido não é somente uma função de intensidade, mas também do contraste simultâneo e das bandas de *Mach*.

O primeiro exemplo está relacionado ao fato do brilho percebido em uma região não depender simplesmente de sua intensidade, como mostra a figura 5. Sabe-se que, nesta imagem, todos os quadrados centrais possuem exatamente o mesmo nível de cinza. Porém, quando o fundo se torna mais claro eles parecem se tornar mais escuros (GONZALEZ; WOODS, 2010).

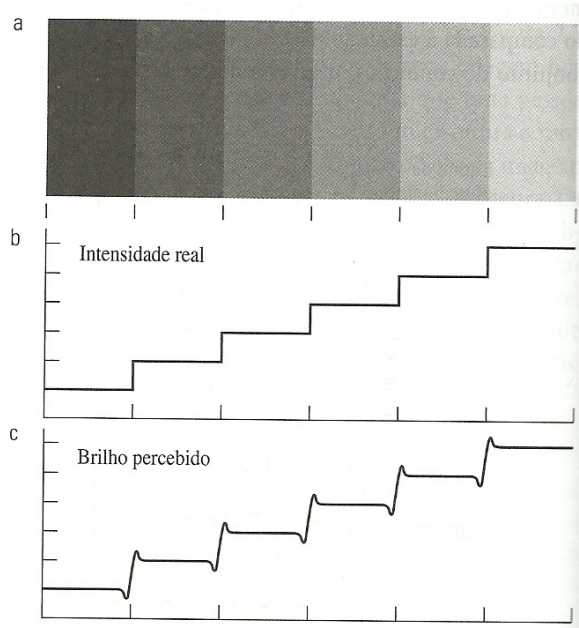


**Figura 5: Exemplificação do contraste simultâneo.**

**Fonte: Adaptado de (GONZALEZ; WOODS, 2010)**

O segundo exemplo é baseado no fato de o sistema visual tender a superestimar ou subestimar os contornos entre as regiões de diferentes intensidades. A figura 6 mostra um exemplo deste mecanismo. Apesar de o nível de cinza das linhas serem constantes, a percepção é que o padrão de brilho é fortemente alterado próximo às bordas (GONZALEZ; WOODS,

2010).



**Figura 6: Representação do efeito de bandas de Mach**

**Fonte: (GONZALEZ; WOODS, 2010)**

Neste ponto é necessário lembrar que é o observador humano quem melhor avalia a qualidade de uma imagem, de um vídeo e até mesmo a qualidade de uma segmentação. Logo, as técnicas utilizadas no processamento de imagens devem utilizar características do SVH.

Para finalizar esta seção, cita-se que o olho humano percebe as cores de um objeto através da luz refletida por ele e a utiliza para identificar e extrair informações em uma cena. Outros termos que serão citados nesta dissertação referem-se à luminância e ao brilho. A luminância mede a quantidade de energia que um observador percebe de uma fonte de luz. Já o brilho, é um descritor subjetivo da percepção da luz que é praticamente impossível de mensurar (GONZALEZ; WOODS, 2010).

## 2.2 ARTEFATOS

Imagens reais frequentemente sofrem degradações durante seu processo de aquisição, transmissão ou processamento. Artefato é um tipo de degradação que apresenta características peculiares. O artefato pode ser definido como uma variável aleatória  $z$ , caracterizada por uma função densidade de probabilidade  $p(z)$  (PEDRINI; SCHWARTZ, 2008).

Imagens/vídeos digitais que apresentam elevado número de pixel necessitam de uma

grande banda de transmissão. Para resolver este problema utiliza-se a técnica de compressão, cujo objetivo é diminuir a redundância dos dados transmitidos, porém é neste momento que pode ocorrer a perda de qualidade visual (STOLFI, 2008). A perda da qualidade acontece devido à aparição de artefatos que não estavam presentes na imagem/vídeo original e que tem origem após a compressão com perdas.

Nesta seção, são discutidos alguns exemplos de artefatos, entre eles, o ruído gaussiano, o ruído sal-e-pimenta, o borramento, a blocagem e o travamento.

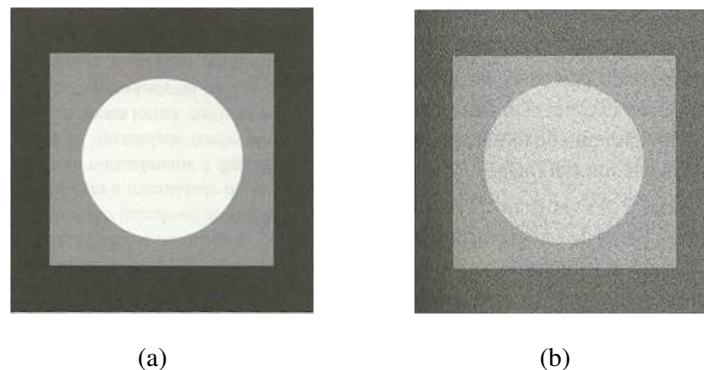
### 2.2.1 RUÍDO GAUSSIANO

O ruído Gaussiano é definido pela ocorrência de pixels com intensidades que variam conforme a distribuição Gaussiana. Diversas aplicações práticas utilizam-no por se tratar de uma boa aproximação da degradação que acontece em situações reais, sendo usado, por exemplo, para modelar o ruído gerado em um sistema de aquisição digital de imagens por componentes eletrônicos. Uma variável aleatória com distribuição Gaussiana tem uma densidade de probabilidade representada pela curva Gaussiana (PEDRINI; SCHWARTZ, 2008). No caso unidimensional, a função densidade de probabilidade é calculada pela equação ( 1).

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}, \quad (1)$$

onde  $\sigma^2$  é a variância e  $\mu$  é a média da variável  $z$ .

A figura 7 demonstra o aparecimento deste artefato em uma imagem digital.



**Figura 7: Ilustração do ruído gaussiano: (a) Imagem original, (b) Imagem com ruído gaussiano.**

**Fonte: Adaptado de (GONZALEZ; WOODS, 2010)**



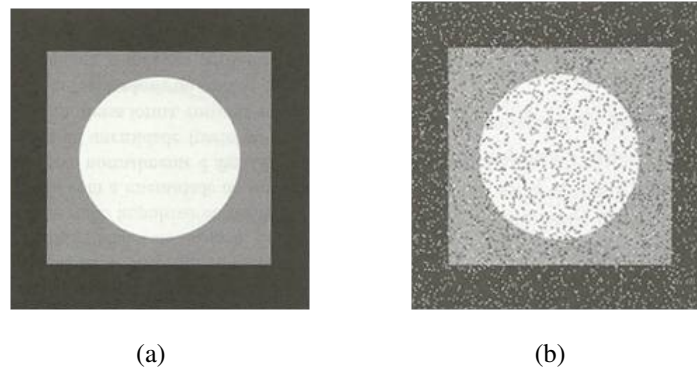
### 2.2.2 RUÍDO SAL-E-PIMENTA

O ruído sal-e-pimenta é originado devido a falhas nos pixels do sensor durante a captura sendo caracterizado como um ruído impulsivo. Um ruído impulsivo ocorre de forma aleatória apresentando pixels cujos valores de luminosidade diferem significativamente dos valores de seus pixels vizinhos. O ruído sal-e-pimenta é visível em imagens quando a degradação é dada pela ocorrência de pixels brancos e pretos (PEDRINI; SCHWARTZ, 2008). A função densidade de probabilidade é representada por:

$$p(z) = \begin{cases} P_a, & \text{para } z = a, \\ P_b, & \text{para } z = b, \\ 0 & \text{caso contrário.} \end{cases} \quad (2)$$

Se  $b > a$ , a intensidade  $b$  aparecerá como um pixel claro na imagem. Em contrapartida, o nível  $a$  aparecerá como um pixel escuro. Se nenhuma probabilidade for igual à zero, e se elas forem aproximadamente iguais, então os valores dos pixels se assemelharão aos grãos de sal-e-pimenta.

A figura 8 demonstra o aparecimento deste artefato em uma imagem digital.



**Figura 8: Ilustração do ruído sal-e-pimenta: (a) Imagem original, (b) Imagem com ruído sal-e-pimenta.**

**Fonte: Adaptado de (GONZALEZ; WOODS, 2010)**

### 2.2.3 BORRAMENTO

O borramento é causado por problemas de foco nas lentes, pelo tempo de exposição alta em conjunto com o movimento, entre outros. Este artefato aparece quando existe perda dos detalhes espaciais em regiões da imagem/vídeo com atividade espacial entre moderada e

alta, regiões de borda ou em áreas com textura. Sabe-se que o borramento está diretamente relacionado com a supressão dos coeficientes de ordem elevada da DCT nos macro-blocos de um quadro codificado (WU; RAO, 2006). Tais coeficientes representam os detalhes finos e as bordas de uma imagem.

É possível ver um exemplo de borramento na figura 9. Percebe-se, principalmente, a ocorrência deste artefato em regiões de alta intensidade, ou seja, na regiões de borda.



**Figura 9: Representação de borramento.**

#### 2.2.4 BLOCAGEM

A blocagem é causada pela representação inadequada do conteúdo de um bloco de pixels quando muitos coeficientes da DCT são descartados ou exageradamente quantizados. Este artefato é definido pela descontinuidade encontrada em uma imagem, principalmente, entre blocos adjacentes e no espaçamento dos blocos (WU; RAO, 2006). Está presente, na maioria dos casos, em cenas complexas ou com uma alta taxa de compressão. A principal causa do seu aparecimento procede da codificação em blocos e seu efeito é produzido a partir do espaçamento regular entre os blocos, cantos acentuados e de bordas induzidas nos limites. Segundo (SILVA, 1998), a blocagem é mais perceptível nas regiões homogêneas dos quadros.

A figura 10 exibe um exemplo do efeito da blocagem. Neste caso, a maior incidência está presente nas regiões de borda, como por exemplo, na face do humano.



**Figura 10: Representação de blocagem.**

**Fonte: (WU; RAO, 2006)**

### 2.2.5 TRAVAMENTO

O travamento é causado, na maioria das vezes, pela perda de informações ou pela ausência delas no receptor. Por exemplo, quando o sinal de recepção é baixo, quando há perda de pacotes ou quando a sincronização é afetada.

Na TV Digital, este artefato está presente na fase de transmissão ou processamento. Uma das práticas utilizadas para reduzir o efeito desta degradação, consiste na repetição do último quadro recebido pelo decodificador (KUROSE; ROSS, 2006).

### 2.3 RELAÇÕES BÁSICAS ENTRE PIXELS

Nesta seção serão discutidos importantes conceitos sobre as relações entre os pixels em uma imagem. Primeiramente, é necessário definir a imagem como sendo uma matriz bidimensional denotada por  $f$  formada por pixels  $f(i, j)$ . Outro padrão a ser seguido é a representação de pixels específicos como  $a$  e  $b$ .

### 2.3.1 VIZINHANÇA

Um pixel  $a$  presente na coordenada  $(i, j)$  tem quatro vizinhos horizontais e verticais cujas coordenadas são representadas por:

$$(i + 1, j), (i - 1, j), (i, j + 1), (i, j - 1). \quad (3)$$

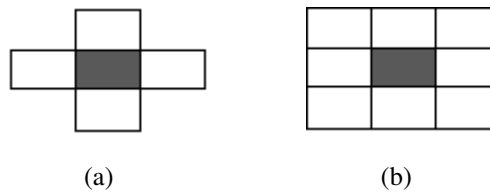
Esse conjunto de pixels, chamado de vizinhança-4 de  $a$ , é expresso pela notação  $N_4(a)$  e representado na figura 11 (a).

Os quatro vizinhos diagonais de  $a$ , representados por  $N_d(a)$ , são os pixels contidos nas coordenadas:

$$(i + 1, j + 1), (i + 1, j - 1), (i - 1, j + 1), (i - 1, j - 1). \quad (4)$$

A vizinhança-8 de  $a$  conhecida por  $N_8(a)$  e ilustrada na figura 11 (b), é definida como (PEDRINI; SCHWARTZ, 2008):

$$N_8(a) = N_4(a) \cup N_d(a). \quad (5)$$



**Figura 11: Tipos de vizinhança: (a) Vizinhança-4 e (b) Vizinhança-8**

### 2.3.2 CONECTIVIDADE

O objetivo da conectividade é determinar se duas regiões satisfazem algum determinado critério de similaridade, tais como, textura, intensidade de cor, e se são vizinhas dependendo do tipo de vizinhança.

Tal assunto é de suma importância, pois a conectividade será utilizada para estabelecer os limites dos objetos e componentes de regiões em uma imagem. Um exemplo de conectividade é dado quando, em uma imagem binária representada por pixels que possuem valores 0 ou 1, dois pixels que são vizinhos (vizinhança-4 ou vizinhança-8) somente serão considerados conexos se possuírem o mesmo valor (PEDRINI; SCHWARTZ, 2008).

### 2.3.3 ADJACÊNCIA

O elemento  $a$  é adjacente ao elemento  $b$  se, dependendo do tipo de vizinhança adotado, eles forem conexos. Os dois tipos principais de adjacência são (GONZALEZ; WOODS, 2010):

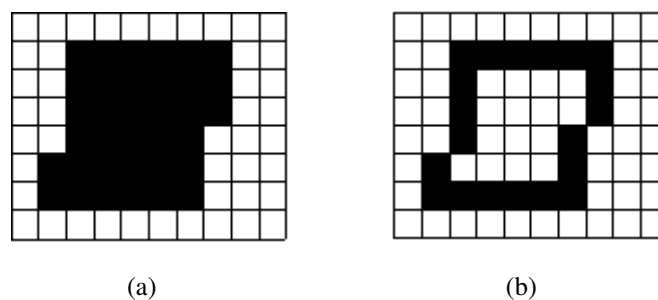
1. Adjacência-4: dois pixels  $a$  e  $b$  são adjacentes-4 se  $b$  estiver no conjunto  $N_4(a)$  (vizinhança-4).
2. Adjacência-8: dois pixels  $a$  e  $b$  são adjacentes-8 se  $b$  estiver no conjunto  $N_8(a)$  (vizinhança-8).

Em outras palavras, dois subconjuntos de pixels,  $P1$  e  $P2$ , são adjacentes se pelo menos um pixel de  $P1$  for adjacente a algum pixel de  $P2$  (PEDRINI; SCHWARTZ, 2008).

### 2.3.4 BORDA

A borda é um conjunto de pixels pertencentes ao elemento conexo  $P$ , em imagem bidimensional, que possui vizinhança-4 com um ou mais pixels externos a  $P$ . Em outras palavras, a borda é o conjunto de pontos no contorno do componente conexo (PEDRINI; SCHWARTZ, 2008).

A figura 12 mostra o exemplo de uma imagem binária (a) e de sua borda (b). Destaca-se que a borda foi obtida analisando os vizinhos através da vizinhança-4.



**Figura 12: Exemplo de borda: (a) Imagem Binária; (b) Borda**

### 2.3.5 MORFOLOGIA MATEMÁTICA BINÁRIA

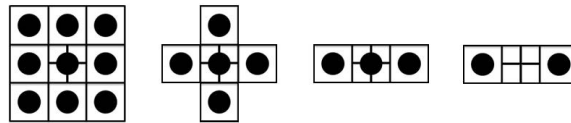
A morfologia matemática (MM) é uma metodologia utilizada para extrair os componentes das imagens que são necessários na representação e na descrição da forma de uma região, como por exemplo, as fronteiras e os esqueletos (GONZALEZ; WOODS, 2010). Trata-se de uma ferramenta poderosa na área de análise de imagens e vem sendo usada em um grande

número de aplicações, tais como, extração e afinamento de bordas e buscas de padrões entre os objetos.

A MM também é utilizada para analisar a forma dos objetos. A análise da estrutura geométrica a partir de malhas retangulares conhecidas como elementos estruturantes (EE) é o principal objetivo da MM e é através deles que ela atua sobre as imagens digitais (SOILLE, 1999).

Para (FACON, 1996) o EE é um conjunto completamente definido e conhecido (forma e tamanho), o qual é comparado, a partir de uma transformação, ao conjunto desconhecido da imagem. O resultado desta transformação permite avaliar o conjunto conhecido.

O elemento estruturante é utilizado pela MM para atuar sobre as imagens digitais e geralmente apresenta a forma de malha retangular. Ele é representado com respeito a uma origem local, normalmente no centro da malha. Neste trabalho, a origem do elemento estruturante está ilustrada por uma cruz. A figura 13 apresenta quatro exemplos típicos de EE, com origem no centro do elemento.



**Figura 13: Exemplos de elementos estruturantes.**

**Fonte: Adaptado de (GONZALEZ; WOODS, 2010)**

A morfologia matemática possui duas operações básicas, a erosão e a dilatação. A erosão apresenta a característica de remover da imagem um pixel que não atenda a um dado padrão. Já a dilatação altera uma determinada área de acordo com um dado padrão (FACON, 1996).

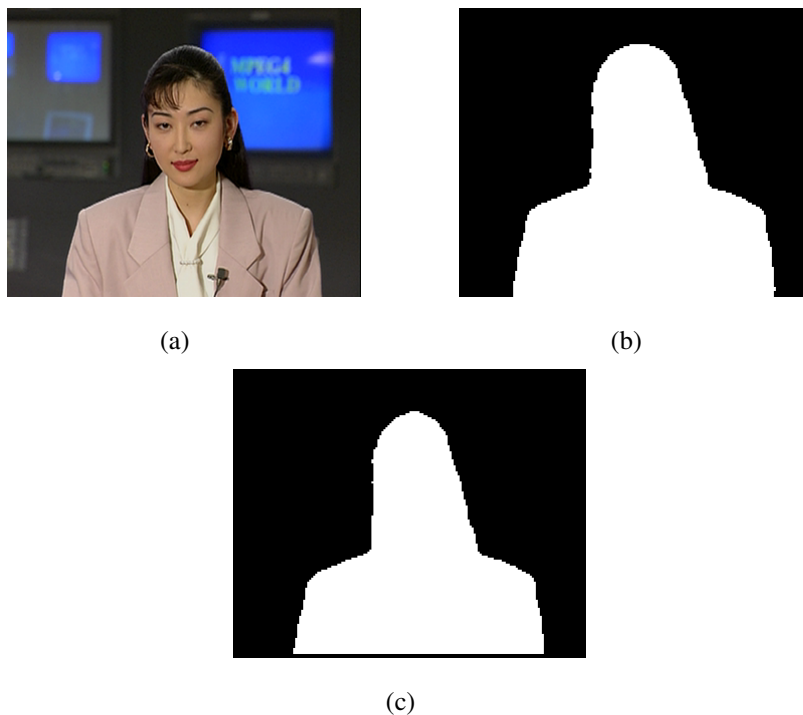
A erosão é uma das operações fundamentais e primitivas da morfologia matemática, na qual muito dos algoritmos morfológicos se baseiam (GONZALEZ; WOODS, 2010). Um conjunto  $X$  erodido pelo elemento estruturante  $B$  é denotado por  $\varepsilon_B(X)$  é definido como a posição dos pontos  $x$  tal que  $B$  está incluído em  $X$  quando sua origem está localizada em  $x$  (SOILLE, 1999):

$$\varepsilon_B(X) = \{x \mid B_x \subseteq X\}. \quad (6)$$

Essa equação representa o elemento estruturante  $B_x$  denotado pelo elemento estrutu-

rante  $B$  centrado no pixel  $x$  que deverá deslizar sobre a imagem  $X$ . De acordo com a equação (6), a erosão é calculada comparando o EE com a vizinhança de  $x$ . Se os *pixels* apresentarem a mesma configuração do EE, o pixel origem é preservado, senão ele será considerado um pixel irrelevante.

Para exemplificar a aplicação desta técnica, é possível ver na figura 14 o afinamento do objeto na imagem binária. Ressalta-se que a figura 14(c) ilustra a execução de cinco passos da aplicação da operação erosão sobre a imagem segmentada (14(b)). O resultado é perceptível, principalmente, nos limites do quadro.

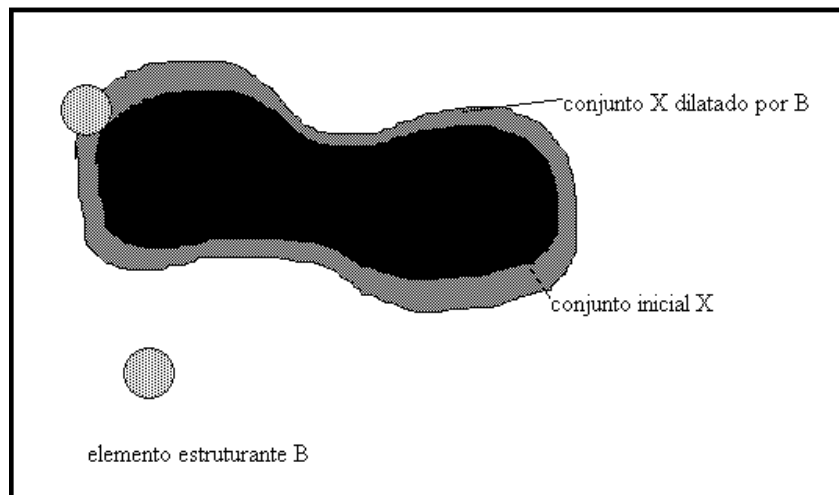


**Figura 14: Exemplo de Erosão: (a) Original, (b) Imagem Segmentada, (c) 5 iterações.**

A dilatação binária é uma transformação morfológica que combina dois conjuntos através da adição. O resultado desta adição será uma imagem “engordada”. Tal técnica é definida por:

$$A \oplus B = c \parallel c = a + b, a \in A, b \in B, \quad (7)$$

onde  $A$  representa a imagem original e  $B$  é o elemento estruturante. A composição do elemento estruturante define a natureza específica da dilatação e neste caso a imagem é expandida. Um exemplo da aplicação desta técnica pode ser visto na figura 15.



**Figura 15: Exemplo de dilatação.**

**Fonte: (FACON, 1996)**

## 2.4 SEGMENTAÇÃO

Entende-se como segmentação o processo de subdividir uma imagem em regiões de interesse que apresentem características similares. Trata-se de uma ferramenta valiosa na extração de informações e na compressão de dados, e tem como objetivo separar as regiões em grupos e destacá-los dos demais. Por exemplo, a segmentação que visa identificar as queimadas na Amazônia, deve agrupar as regiões com elevado calor ou cobertas por fumaça. Na maioria das aplicações, a segmentação é um estágio prévio de processamento para reconhecimento ou identificação das imagens/vídeos (CONCI et al., 2008). Salienta-se que este estágio não é trivial, principalmente, devido a necessidade de precisão nos resultados para o sucesso ou o fracasso dos procedimentos da análise computadorizada (GONZALEZ; WOODS, 2010).

Portanto, um método de segmentação que consiga identificar de forma correta a forma, a topologia e a localização dos objetos é um atributo essencial para que a análise computadorizada seja confiável. Como resultado da segmentação, os objetos são identificados pelas suas características geométricas e topológicas, como por exemplo, área, perímetro e textura. Estas características são utilizadas nas análises.

A subdivisão da imagem deve obedecer ao nível de detalhamento, que é dependente da resolução do problema. Ou seja, o processo de segmentar deve ser parado ao instante que os objetos ou regiões forem detectados. Neste instante surge um dos problemas da segmentação, a definição do nível de detalhamento para identificar os elementos procurados. Por isso, deve-se primeiramente estimar o risco envolvido em aumentar a probabilidade de se obter uma



segmentação precisa. Por exemplo, na área de controle industrial, é possível ter algum controle sobre o ambiente (iluminação). Em outras situações não é possível ter este controle, com isso há um aumento significativo na complexidade da segmentação (GONZALEZ; WOODS, 2010). Outras dificuldades encontradas estão no fato das imagens apresentarem fronteiras ou bordas das regiões sem nitidez e muitas vezes irregulares e imprecisas. Além disso, a existência de movimento ou sombras modifica os aspectos analisados e tende a apresentar uma falsa impressão ao redor da região a ser segmentada (CONCI et al., 2008).

Segundo (LUCCA, 1998), existem dois modos de fazer uma segmentação: (1) Manual: Baseado no reconhecimento visual das feições, um especialista compara as características apresentadas pelos elementos (cor, textura) e consegue separar os objetos; (2) Automático: Através das propriedades espectrais e espaciais, um software procura simular esse especialista para reconhecer regiões homogêneas.

Segmentar de forma automática uma imagem de modo a processar um número de objetos, provavelmente com diferentes formas e tamanhos e em diferentes posições, é uma tarefa complexa, principalmente em imagens que apresentam ruídos. Imagens com artefatos podem levar as técnicas de segmentação a distorcer as formas dos objetos, complicando seu reconhecimento (PEDRINI; SCHWARTZ, 2008).

A maioria dos algoritmos de segmentação utilizam critérios de reconhecimento de diferenças mensuráveis entre regiões ou de similaridades. Na primeira categoria, a ideia é particionar uma imagem (quadro de um vídeo) com base nas mudanças abruptas de intensidade de pixel, caracterizadas pela presença de linhas, pontos isolados e limites (ou bordas). Nesta categoria espera-se que os limites das regiões sejam suficientemente diferentes entre si e em relação ao fundo do quadro para possibilitar a detecção da fronteira com base nas discontinuidades locais. A segmentação baseada nas bordas é um exemplo desta categoria.

Na segunda categoria, as principais abordagens estão baseadas na divisão de uma imagem em regiões similares através de um conjunto de características predefinidas, como por exemplo, níveis de tons, cores e texturas (GONZALEZ; WOODS, 2010). A escolha do conjunto de características deve estar relacionada ao domínio da aplicação com o objetivo de produzir um conjunto de regiões homogêneas (PEDRINI; SCHWARTZ, 2008). A limiarização, divisão e fusão de regiões e o crescimento de região são exemplos de métodos desta categoria. Tais métodos estão fundamentados na separação de uma imagem em regiões que sejam semelhantes respeitando um conjunto de critérios predefinidos.

As técnicas de segmentação, tratadas a seguir, se enquadram na categoria de similaridade.

### 2.4.1 OTIMIZAÇÃO POR COLÔNIA DE FORMIGAS NA SEGMENTAÇÃO DE IMAGENS

Cita-se a técnica de particionamento como um dos principais métodos aplicados no problema de agrupamento (clusterização). Tal técnica tem como objetivo agrupar os elementos em *clusters* (grupos) (BRAMER, 2007). Inicialmente, é definida a quantidade de grupos que serão formados. Posteriormente, são selecionados  $m$  objetos que formarão os  $n$  centros iniciais dos grupos sem membros. Segue-se para a escolha dos centros dos objetos, ou seja, cada objeto será atribuído a um grupo. Para avaliar a qualidade de um agrupamento, uma das maneiras é medir a similaridade entre os objetos de cada grupo, através de uma função objetivo (FO). Neste trabalho, o problema de agrupamento é modelado como um problema de otimização com o objetivo de encontrar uma atribuição de  $m$  objetos com  $v$  atributos para um dos  $g$  grupos de forma que a raiz quadrada da soma dos quadrados das diferenças entre o valor de pixel de cada objeto e o seu centro seja minimizada. Tal função objetivo é conhecida como distância euclidiana e pode ser vista na equação (8). Destacam-se as seguintes restrições para o problema:

1. Cada objeto deve pertencer a um único grupo;
2. Cada cluster deve possuir no mínimo um objeto.

$$\text{Minimizar } J = \sum_{i=1}^m \sum_{j=1}^g w_{i,j} \sqrt{\sum_{v=1}^n (x_{iv} - c_{jv})^2}, \quad (8)$$

onde  $x_i$  é o vetor do objeto  $i$ ;  $c_j$  é o vetor do centro do cluster  $j$  e  $c_j \in R^n$ ;  $x_{iv}$  é o valor do atributo  $v$  do objeto  $i$ ;  $c_{jv}$  é o valor do atributo  $v$  do centro do cluster  $j$ , de modo que este valor seja a média dos valores dos atributos  $v$  dos objetos no cluster  $j$ ;  $w_{ij}$  é o peso associado ao valor de  $x_i$  com  $c_j$ , sendo que  $w_{ij}$  receberá o valor 1 se o objeto  $i$  estiver agrupado no cluster  $j$  e 0 caso contrário. É importante destacar que o problema aplicado na segmentação de imagens consiste em analisar os pixels de cada imagem e criar agrupamentos de pixels similares. Ressalta-se que cada pixel é considerado como um objeto e os valores dos pixels foram normalizados.

Para resolver o problema do agrupamento, uma das formas é a de utilizar sistemas inteligentes, como por exemplo, a meta-heurística ACO (Otimização por Colônia de Formigas). Esta meta-heurística foi proposta por (DORIGO, 1992),(DORIGO et al., 1996) como uma técnica baseada no comportamento de insetos sociais visando resolver os problemas de otimização com a utilização de uma população de soluções. A técnica imita a maneira como as formigas reais buscam, a partir de seu ninho, os seus alimentos através do caminho mais curto. As tarefas são passadas para a colônia de formigas sem que haja uma unidade centralizadora de

informações. Com isso, o único meio delas se comunicarem e decidir qual caminho percorrer, consiste em trilhas compostas por uma substância química conhecida por feromônio. Enquanto uma formiga isolada se move aleatoriamente no espaço de soluções, outra que encontre a trilha de feromônio deixada no seu ambiente tende a segui-la, reforçando ainda mais a trilha com seu próprio feromônio (DORIGO et al., 1996), (DORIGO, 1994).

Um sistema de busca inteligente surge através deste comportamento cooperativo e auto-adaptativo das formigas artificiais e é através dele que as formigas coletam de forma paralela as informações necessárias sobre a estrutura do problema, probabilisticamente decidem qual caminho percorrer e constroem um conjunto de soluções. Vale à pena destacar que essas informações necessárias para a tomada de decisão das formigas podem ser informações específicas do problema em estudo, concentração de feromônio e o valor da função heurística.

O algoritmo de ACO utilizado no presente trabalho se baseia no algoritmo proposto por (KAO; CHENG, 2006) para agrupamento de dados. Uma técnica de busca local é aplicada juntamente com o algoritmo ao final de cada solução construída por uma formiga na tentativa de melhorar sua qualidade. Ao final do processo a saída do algoritmo é dada pela melhor solução e os centros destas.

Os passos executados pelo algoritmo são: a cada movimento, as formigas artificiais selecionam um pixel ainda não agrupado e o adiciona a uma solução parcial considerando a intensidade de feromônio e a informação heurística. Esta última indica o desejo de atribuir um pixel a um cluster em particular e seu valor é obtido calculando o inverso da distância euclidiana entre o pixel corrente e cada centro de cluster existente. Cada formiga transporta uma matriz de centros e a atualiza após cada pixel ser agrupado. Para selecionar um cluster  $j$  para um pixel  $i$ , uma dentre duas estratégias podem ser adotadas dependendo do resultado da equação (9): (a) Intensificação: permite que as formigas se movam de forma gulosa para um centro cujo produto do nível de feromônio e heurística seja mais alto (ver equação (9)); (b) Exploração: permite que uma formiga escolha um centro de forma estocástica (ver equação (10)). Quanto mais promissor um centro for, maior será a probabilidade de este ser escolhido.

$$j = \begin{cases} \max_{j \in N_i} [\tau(i, j)]^\alpha [\eta^k(i, j)]^\beta & \text{se } q \leq q_0 \\ S & \text{caso contrário,} \end{cases} \quad (9)$$

onde  $q_0$  representa uma prioridade atribuída a priori,  $q$  é uma probabilidade gerada aleatoriamente,  $N_i$  é o conjunto de  $g$  clusters disponíveis e  $S$  é escolhido de acordo com a equação (10) sendo que  $P^k(i, j)$  é a probabilidade de atribuir o pixel  $i$  ao cluster  $j$ ,  $[\tau(i, j)]^\alpha$  é expresso como a intensidade de feromônio na ligação  $(i, j)$  indicando o quão útil essa atribuição foi no pas-

sado,  $[\eta^k(i, j)]^\beta$  indica o valor da função heurística para a formiga  $k$  calculada por  $\eta^k = \frac{1}{d^k(i, j)}$  onde  $d^k(i, j)$  representa a distância euclidiana entre o pixel  $i$  e o centro  $j$  de uma formiga  $k$  e é dado por  $d^k(i, j) = \sqrt{\sum_{v=1}^n (x_{iv} - c_{jv}^k)^2}$ .  $\alpha$  e  $\beta$  são constantes positivas utilizadas para indicar a influência das concentrações de feromônio e da heurística empregada, respectivamente.

$$p^k(i, j) = \frac{[\tau(i, j)]^\alpha [\eta^k(i, j)]^\beta}{\sum_{j=1}^g [\tau(i, j)]^\alpha [\eta^k(i, j)]^\beta}. \quad (10)$$

A cada escolha de um centro para um pixel, a informação de cada formiga é atualizada e o centro do cluster cujo pixel foi adicionado é recalculado. Uma vez atribuído todos os pixels a um dos centros de clusters existentes, isto é, uma vez construída uma solução, calcula-se o valor da função objetivo (FO) de cada formiga  $J_k$  (ver equação (8)). Nesse momento, a técnica de busca local proposta no presente trabalho é aplicada com o intuito de melhorar a solução encontrada pela formiga.

Os seguintes passos são executados pela busca local:

1. Percorrem-se todos os pixels  $m$  da solução final construída pela formiga  $k$ ;
2. Compara-se o pixel  $m$  com cada um dos centros de clusters para verificar com qual dos centros ele mais se assemelha (FO);
3. Se o pixel  $m$  for mais semelhante ao centro de cluster para o qual não foi atribuído pela formiga  $k$ , realiza-se a troca para o novo centro. Caso contrário, o pixel continua no cluster para o qual a formiga  $k$  havia atribuído;
4. Se a busca local já percorreu todos os pixels, o algoritmo vai para o passo 5, senão volta para o passo 2;
5. Após realizar todas as trocas necessárias, calcula-se a FO; Se a nova solução encontrada pela técnica de busca local for melhor que a solução da formiga ( $J_k$ ), a nova solução será aceita e as variáveis do ACO serão substituídas pelas novas variáveis da busca local. Caso contrário, o ACO permanece com os dados da formiga  $k$  inalterados.

Após todas as formigas de cada iteração construírem suas soluções e aplicá-las na busca local, as  $k$  soluções são ordenadas através dos valores de suas funções objetivos  $J_k$ . A melhor solução é considerada a melhor solução da iteração corrente. Esta solução é comparada à melhor solução encontrada até o momento e a melhor delas será a nova solução melhor. Nesse momento, a regra de atualização global é aplicada e somente as formigas elites (melhor solução)

poderão adicionar feromônio no final de cada iteração. O feromônio é atualizado da seguinte forma:  $\tau_{ij}(t+1) = (1 - \rho)\tau_{i,j}(t) + \sum_{k=1}^K \Delta\tau_{ij}^k$ . Onde: (a)  $\rho$  é a taxa de evaporação de feromônio,  $0 < \rho < 1$ ; (b)  $t$  é o número da iteração; (c)  $K$  é o número de formigas elites; (d)  $\Delta\tau_{ij}^k$ , diferente do trabalho de (KAO; CHENG, 2006) é determinado pelo inverso da função objetivo  $Jk$  dividido pelo número de objetos existentes ( $m$ ):  $\Delta\tau_{ij}^k(t) = \frac{1}{j^k/m}$  se o pixel  $i$  for atribuído ao centro  $j$ . Quando o algoritmo exceder o número máximo de iterações pré-definido ou quando não houver mais melhora na solução após 10 iterações consecutivas, ele é finalizado e a melhor solução encontrada e os centros desta são retornados.

#### 2.4.2 SEGMENTAÇÃO UTILIZANDO MDRAV

A segmentação utilizando o modelo para detecção de regiões de atenção visual (MDRAV) em vídeo digital foi proposto por (BEZERRA, 2011) e tem como objetivo principal detectar as regiões de atenção visual. As etapas do modelo são:

1. Diferença acumulativa de regiões (DifAR);
2. Eliminação de regiões espacialmente (ERE);
3. Eliminação de regiões espaço-temporalmente (ERT);
4. Intersecção entre regiões (IR).

A etapa de diferença acumulativa provê o conteúdo dos blocos para o quadro central de uma janela de quadros. Este conteúdo é obtido a partir de operações executadas entre o quadro atual ( $F_t$ ) e os outros quadros a sua esquerda ( $F_{t-1}, F_{t-2}$ ) e a sua direita ( $F_{t+1}, F_{t+2}$ ) na janela de quadros. Tais operações são efetuadas através da utilização da equação (11), que representa a probabilidade de mudança ( $P$ ) ou não das intensidades.

$$P_{i,i+j}(x,y) = \begin{cases} \frac{(N+1-j)}{(2\sum_{k=1}^N K)} & \text{se } |F_i(x,y) - F_{i+1}(x,y)| > \tau \\ 0 & \text{caso contrário,} \end{cases} \quad (11)$$

sendo  $\tau$ , definido em 1,  $j$  é a quantidade de quadros distante do quadro central,  $N$  é o número de quadros à esquerda ou à direita do quadro central.  $P$  armazenará uma mudança entre dois pixels, se para todo  $j \in Z^*$  e  $-N \leq j \leq N$ ,  $\|F_i(x,y) - F_{i+1}\|$  for maior que  $\tau$ . A equação(12) exhibe a função para calcular  $P$  sobre uma janela de quadros.

$$P^{jq}(x,y) = \sum_{j=1}^N (P_{i,i-j}(x,y) + P_{i,i+j}(x,y)), \quad (12)$$

onde  $P^{jq}$  estará entre os limites  $[0;1]$ . A equação 12 pode ser estendida para calcular utilizando um bloco com dimensão  $C$  por  $L$ , que resultará na probabilidade de mudança em nível de bloco ( $PB^{jq}$ ), exibida na equação (13).

$$PB^{jq}(x,y) = \frac{1}{CL} \sum_{col=0}^C \sum_{lin=0}^L P^{jq}(x+col,y+lin), \quad (13)$$

sendo  $x$  e  $y$  as coordenadas iniciais do quadro,  $col$  e  $lin$  são os índices do bloco.  $B^{DifAR}$ , conteúdo de cada bloco de  $F_i$ , oriundo das operação da DifAR é composto com base em sua  $PB^{jq}$  e é obtido por:

$$B^{DifAR}(x,y) = \begin{cases} B_{F_i}(x,y) & \text{se } PB^{jq}(x,y) \geq \lambda_p \\ \phi & \text{caso contrário,} \end{cases} \quad (14)$$

onde  $B(x,y)$  é um bloco de  $F_i$  que pertence a uma  $R^{tf}$  (região de primeiro plano),  $\lambda_p$  é o valor probabilístico para que ocorra uma mudança no bloco (neste caso definido em 0,5),  $\phi$  representa a ausência do plano de fundo em um quadro. O resultado desta etapa é ilustrado na figura 16.



**Figura 16: Resultado da etapa DifAR.**

A segunda etapa, ERE, tem como objetivo eliminar os blocos em um quadro evitando a formação de falsas regiões. Esta etapa é aplicada aos quadros resultantes da etapa de DifAR,

fornecendo novos blocos chamados de  $B^{ERE}$ , conforme:

$$B^{ERE}(x,y) = \begin{cases} B(x,y)^{ref} & \text{se } B^{DifAR}(x,y) = \phi \text{ e } N8_{tf} \geq \lambda_{tf} \\ \phi & \text{se } B^{DifAR}(x,y) \neq \phi \text{ e } N8_{\phi} \geq \lambda_{\phi} \\ B^{DifAR}(x,y) & \text{caso contrário,} \end{cases} \quad (15)$$

onde  $B^{ref}$  é o bloco do quadro de referência,  $B^{DifAR}(x,y)$  é o bloco de um quadro resultante da etapa DifAR, as variáveis  $\lambda_{tf}$  e  $\lambda_{\phi}$  são os valores limites para a quantidade de vizinhança-8 ( $N8$ ). Por fim,  $N8_{tf}$  e  $N8_{\phi}$  são valores entre  $[0;8]$  que representam a quantidade de blocos vizinhos que estão contidos em  $R^{tf}$  ou que são vazios ( $\phi$ ). Em (BEZERRA, 2011), para tais variáveis foram atribuídos os valores 8, 7 e 5, ou seja, a etapa ERE foi executada 3 vezes. O resultado desta etapa é exibido na figura 17.



**Figura 17: Resultado da etapa ERE.**

A terceira etapa, ERET, se diferencia da anterior pelo fato de adicionar o fator temporal. Isto se faz necessário, para que ocorra a eliminação efetiva de quadros não pertencentes ao conteúdo do bloco. Tal etapa é calculada pela equação (16).

$$B_{F_i}^{ERT}(x,y) = \begin{cases} \phi & \text{se } B_{F_i}^{ERE}(x,y) \neq \phi \text{ e } N8_{\phi} \geq \lambda_{\phi} \\ & \text{e } (B_{F_{i-1}}^{ERE}(x,y) = \phi \text{ ou } B_{F_{i+1}}^{ERE}(x,y) = \phi) \\ B_{F_i}^{ref}(x,y) & \text{se } B_{F_i}^{ERE}(x,y) = \phi \text{ e } N8_{tf} \geq \lambda_{tf} \\ & \text{e } (B_{F_{i-1}}^{ERE}(x,y) \neq \phi \text{ ou } B_{F_{i+1}}^{ERE}(x,y) \neq \phi) \\ B_{F_i}^{ERE}(x,y) & \text{caso contrário,} \end{cases} \quad (16)$$

sendo  $B^{ERE}$  o bloco do quadro ( $F_i$ ), anterior ( $F_{i-1}$ ) e seguinte ( $F_{i+1}$ ), resultado da etapa ERE. A figura 18 demonstra o resultado obtido nesta etapa.



**Figura 18: Resultado da etapa ERT.**

A quarta e última etapa, denominada IR, tem como objetivo filtrar os pixels que pertencem a regiões de primeiro plano  $R_{tf}$ , e que estejam percentualmente representados dentro de uma janela de quadros. Tal etapa possui três passos que são aplicados sobre os resultados obtidos na etapa anterior:

1. Registro espacial de regiões (RER);
2. Registro da persistência temporal de RER ( $RER^{jq}$ );
3. Análise da IR (AIR).

O primeiro passo ( $RER$ ) verifica se o pixel pertence a uma região de primeiro plano do quadro da janela em análise. Seu cálculo é representado por:

$$RER(x,y) = \begin{cases} 1 & \text{se } F(x,y) \neq \phi \\ 0 & \text{caso contrário,} \end{cases} \quad (17)$$

O segundo passo, aqui denominado  $RER^{jq}$ , aplica o  $RER$  sobre todo espaço de tempo avaliado, registrando a persistência temporal de um pixel pertencente ao primeiro plano. O cálculo de  $RER^{jq}$  é obtido por:

$$RER^{jq} = \sum_{i=1}^{T_{jq}} RER_i(x,y). \quad (18)$$



O objetivo do terceiro passo é analisar o IR sobre a persistência do pixel espaço-temporalmente, calculado por:

$$F_i^{IR}(x,y) = \begin{cases} F_i^{ref} & \text{se } RER^{jq}(x,y) > \lambda_p \cdot T_{jq} \\ \phi & \text{caso contrário,} \end{cases} \quad (19)$$

onde  $\lambda_p$  é valor limite probabilístico para que ocorra uma mudança no bloco. Este valor está sendo multiplicado por  $T_{jq}$ , para se obter o limiar de IR. O resultado desta etapa é exibido na figura 19.



**Figura 19: Resultado da etapa IR.**

## 2.5 AVALIAÇÃO DA QUALIDADE DE SEGMENTAÇÃO

A segmentação de uma imagem em objetos ou regiões de interesse (plano de fundo e primeiro plano) é uma tarefa complexa e subjetiva. A proposta desse trabalho, em termos de avaliação, é medir a qualidade das técnicas de segmentação estudadas.

Existem dois métodos para avaliar a qualidade de segmentação: os métodos analíticos e os métodos empíricos (ZHANG, 1996). Os métodos analíticos visam avaliar o desempenho do algoritmo de segmentação, levando em consideração os aspectos de complexidade e propriedades da implementação do algoritmo. Em contrapartida, os métodos empíricos têm como principal objetivo avaliar os resultados obtidos pela segmentação.

O objetivo desta seção é discutir a qualidade da segmentação com o emprego dos métodos empíricos. Tais métodos podem ainda ser divididos em dois grupos: métrica empírica de similaridade (relativa) e métrica empírica de discrepância (independente) (GASPAR, 2009).

A avaliação da qualidade de segmentação de forma relativa dá-se pela utilização de uma referência, como por exemplo, “*ground-truth*” ou uma segmentação dita ideal, para dar uma nota indicando a qualidade da segmentação. Enquanto a avaliação relativa utiliza uma referência, a avaliação independente é baseada nas características dos objetos segmentados, como por exemplo, na forma, na posição ou no contraste entre as regiões de interesse.

Sabe-se que, em muitos casos, a métrica de avaliação relativa obtém melhores resultados comparada com a avaliação independente. Isto acontece devido à existência de uma segmentação “ideal” para o cenário onde será utilizada a segmentação. Logo, se um determinado algoritmo de segmentação apresenta bons resultados para este cenário, acredita-se que o mesmo funcione bem para outros conteúdos desse mesmo tipo. A condição necessária é ter disponível uma referência que poderá ser feita de forma manual, a fim de fornecer um resultado bom para a segmentação. Tal processo de segmentação manual é dependente do tipo de aplicação, pois pode haver requisitos de segmentação diferentes para os diferentes cenários. Como exemplo, podem-se citar dois tipos diferentes de aplicações: câmeras de vigilância e composição de uma nova cena. Nas câmeras de vigilância, caso alguma região do fundo seja selecionada como objeto isso pode não ser um problema, desde que o objeto de interesse esteja visível, em comparação, quando o objetivo é compor uma nova cena é inaceitável que os objetos segmentados apresentem regiões do fundo (GASPAR, 2009).

É importante destacar que a maioria das aplicações que utilizam algoritmos de segmentação é destinada aos observadores humanos. Logo, é subentendido que a avaliação da qualidade deve ser modelada para atender, tanto quanto possível, as características do SVH. Com isso, os resultados alcançados devem prover uma maior adequação ao que se obtém com um observador humano. Em (OSBERGER et al., 1998), os autores mencionam diversos fatores que influenciam a atenção humana, tais como, tamanho do objeto, forma, contraste, brilho, orientação, cor, primeiro plano, fundo, localização e movimento. Esses fatores podem e devem ser usados na avaliação da qualidade de segmentação, ou seja, através destas características deve-se ponderar cada fator de acordo com sua relevância visual.

A próxima seção discutirá a avaliação independente, pois é ela que será utilizada para avaliar a qualidade de segmentação neste trabalho. Esta escolha foi feita devido ao fato da avaliação relativa depender da existência de uma segmentação de referência, muitas vezes feita de forma manual. Como o trabalho visa uma avaliação independente da referência, esta é a abordagem mais adequada.

## 2.6 QUALIDADE DE SEGMENTAÇÃO PELO MÉTODO INDEPENDENTE

Nesta seção são abordadas as métricas mais relevantes encontradas na literatura para avaliação independente, onde não há uma segmentação de referência para ser comparada. O foco é a discussão dos parâmetros espaciais (forma) e temporais (movimento).

### 2.6.1 MÉTRICA OBJETIVA INDEPENDENTE PARA AVALIAÇÃO DA QUALIDADE DE SEGMENTAÇÃO

A avaliação da segmentação empregada nesta dissertação fundamenta-se nos trabalhos de (ERDEM et al., 2004) e (CORREIA; PEREIRA, 2003), apresentados a seguir.

Em (ERDEM et al., 2004), os autores discutem uma série de medidas para avaliação independente, tais como a métrica intra-quadro de diferença de cor ao longo das fronteiras dos objetos e uma métrica inter-quadro que faz a diferenciação do histograma de cores e medidas de movimento.

A medida intra-quadro de diferença de cor é obtida através da comparação entre os pixels internos e externos do objeto ao longo de sua fronteira. Supõe-se que as fronteiras do objeto coincidem com os limites de cores, de modo que apenas os pixels internos deverão ser significativamente diferentes, quando comparados aos pixels externos.

O conjunto de pixels usados é obtido pelo desenho de linhas curtas no sentido limite do objeto, com tamanhos iguais do interior para a parte externa do objeto. Esta medida é dada por:

$$d_{CB}(t) = 1 - \frac{1}{K_t} \sum_{i=1}^{K_t} d_{CB}(t;i), \quad (20)$$

$$d_{CB}(t;i) = \frac{\|Col_0^i(t) - Col_1^i(t)\|}{\sqrt{3} \times 255^2}, \quad (21)$$

onde  $K_t$  é o número de linhas normais desenhadas na fronteira do objeto estimado no quadro  $t$ ,  $Col_0^i(t)$  é a média de cor calculada na vizinhança externa do pixel  $i$ th e  $Col_1^i(t)$  é a média de cor da vizinhança interna do  $i$ -ésimo pixel. Com esta métrica, um resultado ruim de segmentação deverá atingir o valor limite de  $d_{CB}(t)$ , que neste caso é representado por 1.

Já a métrica inter-quadro da diferença do histograma de cores é feita supondo que o histograma de cores de um objeto segmentado específico seja estático de quadro a quadro. Seguindo esta suposição, a métrica retorna a medida de variação de histograma de cores do objeto segmentado ao longo do tempo. Os autores (ERDEM et al., 2004) propõem uma forma

para calcular a diferença entre histogramas do objeto nos tempos  $t$  e  $t - 1$  como segue:

$$d_{x^2}(H_{t-1}, H_t) = \sum_{j=1}^B \frac{[H_{t-1}(j) - H_t(j)]^2}{H_{t-1}(j) + H_t(j)}, \quad (22)$$

onde  $H_t$  e  $H_{t-1}$  são, respectivamente, os histogramas de cores do objeto segmentado no instante  $t$  e  $t - 1$  e  $B$  é o número de colunas do histograma.

Outra abordagem para a métrica intra-quadro de diferença entre cores é dada pela medida de movimento. Ao invés de comparar a cor somente entre os pixels internos e externos da borda do objeto, esta métrica propõe calcular a diferença entre os vetores de movimento em torno destes pixels. Este cálculo é considerado como:

$$d_M(t) = 1 - \frac{\sum_{i=1}^{K_t} d_M(t; i)}{\sum_{i=1}^{K_t} \eta_i}, \quad (23)$$

$$d_M(t; i) = d(v_0^i(t), v_1^i(t)) \cdot \eta_i, \quad (24)$$

$$\eta_i = Rel(v_0^i(t)) \cdot Rel(v_1^i(t)), \quad (25)$$

onde  $v_0^i(t)$  e  $v_1^i(t)$  são, respectivamente, a média dos vetores de movimento calculados em torno do pixel  $i$ th externo e interno,  $d(v_0^i(t), v_1^i(t)) \cdot \eta_i$  denota a distância média entre dois vetores de movimento e  $Rel(v(t))$  é a confiabilidade do vetor  $v(t)$  em um determinado ponto  $p(t)$ .

Em (CORREIA; PEREIRA, 2003), os autores propõem métricas para a avaliação independente, ou seja, sem referência. As métricas podem ser subdivididas em dois grupos: homogeneidade intra-objeto e disparidade inter-objeto.

O grupo de métricas intra-objeto, cujo objetivo é medir a homogeneidade interna dos objetos, são fundamentadas nas características espacial e temporal. As características consideradas são regularidade de forma (circularidade, alongamento, compacidade), uniformidade espacial (variância da textura, informação perceptual espacial), estabilidade temporal (tamanho, posição, variância da textura, circularidade, alongamento, compacidade, informação temporal perceptual e estabilidade de criticidade) e uniformidade de movimento (variância do vetor de movimento e criticidade).

O segundo grupo, disparidade inter-objeto, é uma medida de comparação entre objetos vizinhos. As métricas propostas para o cálculo da disparidade incluem o contraste local entre vizinhos, e mais quatro métricas baseadas na diferença de vizinhança dos objetos. Essas quatro métricas são: diferença da uniformidade de movimento, diferença da regularidade de forma, diferença da estabilidade temporal e diferença da uniformidade espacial.

Por fim, destaca-se que neste trabalho o método proposto por (CORREIA; PEREIRA, 2003) será utilizado para avaliar as segmentações. Na seção 3.5 serão discutidas as vantagens e os cálculos envolvidos para obtenção da nota final da avaliação da qualidade de segmentação.

## 2.7 ESTIMAÇÃO DA RELEVÂNCIA DO OBJETO

Esta seção será dedicada ao entendimento de como se estima a relevância de um objeto em uma cena de vídeo. É a relevância que dita o valor de ponderação utilizado pela técnica de avaliação de vídeo proposta.

As cenas de vídeo normalmente exibem diversos objetos de interesse e, por isso, é importante que a segmentação atinja não somente um, mas sim vários objetos. Logo, sabe-se que o algoritmo de segmentação pode trazer mais de um objeto segmentado de interesse, tornando necessário o estudo da relevância de cada objeto na cena a fim de obter uma medida de qualidade de segmentação mais coerente. Neste ponto, é importante dizer que, na grande maioria das vezes, nem todos os objetos presentes em uma cena de vídeo tem o mesmo nível de relevância (GASPAR, 2009).

A relevância de um objeto está relacionada diretamente com a importância do objeto na cena para um observador humano e é baseado fortemente na atenção visual. A atenção visual é um ramo de estudo complexo, pois existem inúmeros fatores que fazem com que, por exemplo, alguns objetos atraiam mais atenção do que outros. Uma das formas de determinar a atenção visual de forma automática é dada pelas características dos objetos, tais como: contraste, forma, cor, tamanho e movimento (OSBERGER et al., 1998).

Esta dissertação aborda dois principais métodos encontrados na literatura (ITTI et al., 1998) e (CORREIA; PEREIRA, 2000). Em (ITTI et al., 1998), os autores propuseram um modelo para atenção visual de um observador humano através do desenvolvimento de um mapa de saliência que representa áreas de maior atenção visual, baseado nas características dos objetos, tais como, cor, intensidade e orientação.

Por sua vez, (CORREIA; PEREIRA, 2000) apresentaram uma combinação de várias métricas elementares baseadas nas características do SVH, formando uma métrica única de relevância contextual, onde cada classe de métricas elementares é ponderada de acordo com seu impacto na captação da atenção visual.

A relevância dos objetos pode ser medida de duas formas: individualmente e contextualmente. A relevância individual é utilizada quando, por exemplo, se quer compor uma nova cena e deseja-se saber qual a relevância individual do objeto segmentado. Em contrapartida, a

relevância contextual, é útil para medir a qualidade geral da segmentação de um determinado quadro. Na seção 3.6 são descritos os cálculos para se obter a relevância individual e contextual dos objetos na cena.

## 2.8 AVALIAÇÃO DA QUALIDADE DE IMAGENS E VÍDEOS

A qualidade de uma imagem é fortemente dependente do tipo de aplicação. Imagens são usadas, por exemplo, tanto para televisão e jogos eletrônicos quanto aplicações para área de medicina (aplicações que necessitam de uma grande resolução) (GASPAR, 2009). É possível notar através destes exemplos que os tipos e graus de degradação que uma imagem pode sofrer, em geral, são bem diferentes.

Diante desse problema, é necessário medir a qualidade das imagens e vídeos. Esta avaliação pode ser feita de duas maneiras: objetiva e subjetiva. Na avaliação subjetiva, após o observador dar a nota para a qualidade, é feito um estudo estatístico com os dados coletados e o resultado é uma nota conhecida como MOS (*Mean Opinion Score*) ((ITU-R, 2002)). A MOS é usada para avaliar a qualidade percebida e, segundo (WANG et al., 2004), é a única maneira apropriada de se quantificar a qualidade, pois são os próprios espectadores que podem melhor avaliar a percepção visual do vídeo.

Seguindo a mesma linha de pensamento, (KOUMARAS et al., 2007) concluem que o MOS é seguramente o método mais preciso para uma avaliação de qualidade, e assim, vários estudos o utilizam para avaliar subjetivamente a qualidade. Para esta avaliação, a (ITU-R, 2002) disponibiliza uma série de recomendações que especificam alguns procedimentos para as avaliações subjetivas, entre elas: distância do telespectador à imagem, método utilizado na aplicação dos testes, características importantes a serem preenchidas na ficha do avaliador, número de telespectadores por teste, forma de ser exibido o vídeo e até mesmo a iluminação do local onde será aplicado o teste. A P.910 da (ITU-T, 2008) é outra recomendação onde são apresentadas sugestões, tais como, análise estatística dos resultados, quantidade de participantes, arranjo das condições de visualização, instruções e treinamento dos espectadores. Porém, o método subjetivo requer um elevado tempo, pois necessita de um grande número de espectadores e por isso torna-se muito dispendioso para ser implantado e mantido por um longo período (DURANT, 1998).

Por outro lado, a avaliação objetiva propõe alcançar uma nota próxima a que seria atribuída pelo observador humano por meio de funções matemáticas. Esta avaliação, preferencialmente, deve usar elementos do SVH. Uma das principais vantagens de utilizá-la é a

diminuição do tempo gasto em sua realização.

A avaliação objetiva da qualidade de vídeo digital pode ser subdividida em três tipos: sem referência ( NR - *no-reference*), com referência reduzida ( RR - *reduced-reference*) e com referência total ( FR - *full-reference*)(WU; RAO, 2006),(KOUMARAS et al., 2007). As métricas sem referência tentam prever a qualidade do vídeo com base apenas na imagem final recebida, ou seja, da mesma forma que um telespectador avalia sem saber como era aquela imagem na sua forma original. O preço que se paga por esta característica é a complexidade e a dificuldade de se implementar esta métrica (WANG; BOVIK, 2006). Ainda existem poucos estudos nesta vertente e não se conhece alguma métrica que simule de forma coerente o SVH do espectador (KOUMARAS et al., 2007), (RICHARDSON, 2003).

O segundo tipo de avaliação objetiva estudada é a métrica com referência reduzida, onde algumas informações do vídeo original são utilizadas. Ou seja, quando o vídeo é transmitido, junto com ele são anexadas algumas informações da imagem referência. Como exemplo, a técnica de referência reduzida, conhecida como LHS - *Local Harmonic Strength*, utiliza a perda harmônica e o ganho harmônico (GUNAWAN; GHANBARI, 2003). A perda harmônica e o ganho harmônico são informações obtidas com base nas bordas detectadas, e com os coeficientes da FFT - *Fast Fourier Transform*. Tais valores são calculados tanto para o vídeo usado como referência quanto para o degradado e os resultados são comparados para geração da nota da avaliação da qualidade do vídeo (GUNAWAN; GHANBARI, 2007). A avaliação com RR é menos complexa quando comparada com a avaliação sem referência.

Na avaliação com referência total a base dos cálculos é dada pela comparação entre a imagem recebida e a imagem enviada. As principais vantagens desta métrica é a baixa complexidade dos cálculos e a facilidade na implementação das técnicas, mas apresenta a desvantagem de se necessitar da referência no ponto de avaliação.

Neste trabalho são empregadas algumas técnicas com referência total baseadas em medidas de similaridade ou discrepância entre imagens. Tais métricas estão entre as que mais são utilizadas na literatura.

### 2.8.1 ERRO MÉDIO QUADRÁTICO

Para calcular o PSNR é necessário, primeiramente, medir o erro médio quadrático (MSE, *Mean Squared Error*) do vídeo. O MSE é calculado como a soma do quadrado das diferenças de cada pixel do quadro referência e do quadro degradado, dividido pela quantidade de pixels do quadro, expresso como:

$$MSE = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x,y) - g(x,y)]^2, \quad (26)$$

onde  $M$  e  $N$  representam a largura e altura do quadro, respectivamente. Com esta métrica, quanto menor o valor do MSE, melhor será a imagem degradada comparada com a referência. Para medir a qualidade do vídeo, calcula-se a média aritmética sobre os instantes avaliados.

## 2.8.2 RELAÇÃO SINAL-RUÍDO DE PICO

A relação sinal-ruído de pico (PSNR, *Peak Signal to Noise Ratio*) é usada para avaliar a diferença global entre dois quadros. O cálculo do PSNR é dado por:

$$PSNR = 10 \times \log_{10} \frac{L_{max}^2}{MSE}, \quad (27)$$

$$L_{max} = 2^n - 1 \quad (28)$$

onde  $L_{max}$  é o valor máximo de intensidade de cinza que pode ser assumida por um pixel e  $n$  é quantidade de bits por pixel. Normalmente, para imagens representadas por oito bits ( $n = 8$ ),  $L_{max}$  é igual a 255. Diferentemente da métrica MSE, quanto maior o valor de PSNR, mais a imagem degradada se aproxima da referência. O PSNR é expresso em *decibel*. Quando o objetivo é medir a qualidade do vídeo, o PSNR é o resultado da média aritmética das avaliações para todos os instantes.

Por fim, destaca-se que o PSNR é uma medida quantitativa da distorção que apresenta algumas limitações, como por exemplo, não considerar limitações do SVH e não possuir um alto nível de correlação quando comparado com a métrica subjetiva. Em outras palavras, um valor de PSNR particular não necessariamente se compara com a qualidade observada por um avaliador. Ou seja, nem sempre o PSNR tem uma equivalência direta com a qualidade perceptual. A fim de exemplificar essa afirmação, a figura 20(b) mostra uma imagem degradada da figura 20(a), onde somente o fundo da imagem está borrado. Tal imagem possui um PSNR de 27,7 dB comparado com a original. Visivelmente é possível verificar que a maioria dos avaliadores dariam notas melhores para esta em comparativo, por exemplo, com a imagem 20(c), que apresenta um PSNR de 28,3 dB. Mesmo apresentando valores menores de PSNR, a imagem 20(b) quando comparada com a imagem 20(c) é avaliada como melhor pelo fato do foco de atenção (a face da violinista) estar mais nítido, contradizendo a nota dada pela métrica.



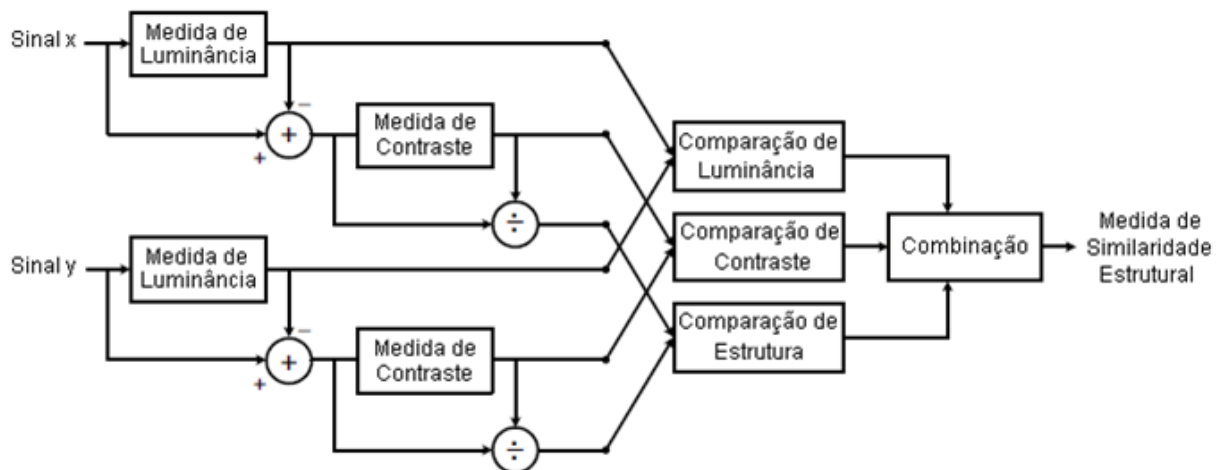


**Figura 20: Exemplos de PSNR: (a) original, (b) 27,7 dB, (c) 28,3 dB.**

**Fonte: Adaptado de (RICHARDSON, 2003)**

## 2.9 SSIM

O índice de similaridade estrutural (SSIM) é aplicado sobre a luminância, contraste e a estrutura da imagem. Foi proposto por (WANG; BOVIK, 2002) e seu objetivo é melhorar a correlação com as medidas subjetivas através da proposta de uma métrica baseada na similaridade estrutural (WANG et al., 2004). Sua principal característica é avaliar a percepção de qualidade em imagens estáticas. A figura 21 mostra o diagrama do modelo.



**Figura 21: Diagrama do SSIM**

**Fonte: Adaptado de (WANG et al., 2004)**

O cálculo do SSIM entre duas imagens  $f(i, j)$  e  $g(i, j)$  é ilustrado na equação (29).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (29)$$

$$C_1 = (K_1L)^2, \quad (30)$$

$$C_2 = (K_2L)^2, \quad (31)$$

onde  $\mu$  é a média,  $\sigma^2$  representa a variância e  $2\sigma$  é a covariância,  $L$  representa a faixa dinâmica dos valores de pixels (255, quando o pixel é representado por 8 bits),  $K_1$  é igual a 0,01 e  $K_2$  é igual 0,03. Tais valores foram definidos empiricamente pelos autores e estes são bem sensíveis quando variados.

Ressalta-se que as três características base para cada bloco 8x8 das imagens são:

1. Comparação entre a luminância das duas imagens ( $l(f, g)$ );
2. Comparação do contraste entre as duas imagens ( $c(f, g)$ );
3. Comparação entre a estrutura das duas imagens ( $s(f, g)$ ).

O resultado obtido pela métrica é um valor entre 0 e 1, sendo 1 o valor ideal. Tal métrica não é executada pixel a pixel, como as anteriores, mas geralmente em blocos de dimensão 8x8. Segundo (WANG et al., 2004), para reduzir a complexidade computacional, sugere-se a utilização apenas em um subconjunto de possíveis janelas.

O SSIM também é utilizado para avaliar a qualidade de vídeo, para isto, o índice é calculado pela média dos resultados obtidos dos quadros analisados.

## 2.10 VQM

O VQM (*Video Quality Metric*) é o algoritmo de avaliação da qualidade de vídeo proposto por (PINSON; WOLF, 2004). Testes com referência completa feitas na Fase II do VQEG (*Video Quality Experts Group*) (VQEG, 2003), comprovaram seus bons resultados em termos de correlação (0,938), uma das mais altas obtidas no teste. Por isso, tal métrica foi selecionada como padrão nacional para avaliação objetiva da qualidade de vídeo pela ANSI (*American National Standards Institute*). A métrica considera as características perceptuais para medir a qualidade do vídeo, como por exemplo, o embaçamento, os ruídos, a distorção de bloco de cor, o movimento não natural e a atenuação.

Seu cálculo é dividido em 4 etapas distintas: calibração, extração de características de qualidade, cálculo dos parâmetros de qualidade, e, por fim o cálculo da métrica. Na primeira

etapa é feita a calibração do vídeo para que este apresente a mesma série cronológica e espacial do vídeo referência. Na segunda etapa é calculada, através de funções matemáticas, a extração de um conjunto de características relacionadas às mudanças temporais, espaciais e propriedades de crominância, ocorridas entre o vídeo com degradação e a referência. A terceira etapa consiste em comparar as características obtidas na etapa anterior entre o vídeo referência e o degradado. A última etapa mede a qualidade do vídeo através de uma combinação linear dos resultados obtidos nas etapas anteriores (PINSON; WOLF, 2004).

O algoritmo usa uma combinação linear entre 7 variáveis, sendo quatro obtidas pelo gradiente espacial aplicado sobre a luminância, dois são baseados nas características extraídas pelas componentes de crominância e o último é estimado pelo contraste e pela informação temporal da luminância (WANG, 2006). Destaca-se que quanto menor o valor obtido pela métrica melhor será a qualidade do vídeo.

## 2.11 VSNR

O VSNR (*Visual Signal to Noise Ratio*) é o método de avaliação que analisa a influência da percepção do ruído em áreas uniformes nas imagens, desenvolvido por (CHANDLER; HEMAMI, 2007). Seu objetivo é quantificar a fidelidade visual de imagens degradadas. Seu cálculo é dividido em 2 fases:

1. Detecção dos limiares (*threshold*) do contraste usando a luminância. Caso o quadro avaliado não apresente artefatos, o algoritmo retorna infinito ( $\infty$ ), indicando uma perfeita fidelidade quando comparado com uma referência.
2. Caso a avaliação extrapole o valor de limiar determinado na primeira fase, é estimada a fidelidade visual com base em uma medida de contraste percebido e em uma medida de grau em que as distorções interrompem a precedência do vídeo.

Como desvantagens cita-se o fato de não considerar a localização espacial da distorção e não utilizar as outras componentes de cor.

Destaca-se que no caso de avaliação de vídeo, o VSNR é aplicado quadro a quadro e o índice global é calculado como a média das notas obtidas pelo VSNR para os quadros. Quanto maior for o índice, mais próxima a imagem degradada está da original.

## 2.12 TRABALHOS RELACIONADOS

Alguns trabalhos foram propostos para avaliar a qualidade de vídeo digital tentando melhorar as técnicas existentes, como por exemplo, o PSNR (EDEN, 2008), (FEI et al., 2007). Seguindo na linha de melhoria das técnicas, Eden (EDEN, 2008) apresentou uma nova abordagem para medir a qualidade baseada na percepção do observador humano. Esta abordagem tem como fundamento a métrica de PSNR (PSNR) melhorada com a utilização de um indicador de atividade espacial no quadro de vídeo. Tal indicador é calculado por meio de uma estrutura que detecta o gradiente das altas intensidades no quadro. Após obtê-lo, segue-se para o cálculo do valor quadrático médio que será utilizado como limiar ( $T$ ) na técnica modificada. O cálculo da métrica modificada é dado por:

$$pPSNR = 10 \times \log_{10} \left( \frac{255^2}{pMSE} \right) \quad (32)$$

$$pMSE = \max \left( 1, \frac{MSE}{\max(1, T^2)} \right), \quad (33)$$

Os autores (EDEN, 2008) concluíram que a modificação do PSNR adicionando a característica de atividade espacial prediz a qualidade da imagem percebida e melhora significativamente o PSNR, em termos de correlação com a avaliação subjetiva. A vantagem do trabalho consistiu em não utilizar uma referência para calcular a atividade espacial, pois ela é calculada sobre a imagem decodificada.

No trabalho de (FEI et al., 2007), também foi proposta uma melhoria no algoritmo de avaliação da qualidade de vídeo: o PSNR. O objetivo do trabalho foi utilizar as propriedades do sistema visual humano para auxiliar na medida da qualidade. O primeiro passo é dado pela aplicação do Filtro de Sobel sobre o quadro avaliado, com isso, obtêm-se o gradiente horizontal e vertical do quadro ( $g(m, n)$ ). O próximo passo consiste no cálculo da métrica MSE sobre os pixels de borda dos supostos objetos  $MSE_{edge}$ , obtidos pelo passo anterior. Por fim, o PSNR melhorado é calculado por:

$$EPSNR = 10 \times \log_{10} \frac{255^2}{EMSE}, \quad (34)$$

$$EMSE = \frac{1}{L} \sum_{k=1}^L MSE_{edge}(k), \quad (35)$$

onde  $L$  é a quantidade de pixels de borda,  $k$  é o quadro que está sendo avaliado. O valor 255 foi utilizado por se tratar de uma imagem de 8 bits. Por fim, os autores concluíram que a utilização

de propriedades do sistema visual humano trazem melhorias para a avaliação da qualidade. O método apresentou uma melhor correlação com a avaliação subjetiva quando comparado com o método convencional.

### 3 METODOLOGIA

Esta dissertação foi dividida em etapas, como segue: (1) segmentação, (2) avaliação da segmentação e cálculo da relevância, (3) geração de artefatos nos vídeos e (4) avaliação da qualidade de vídeo. Nesta seção serão discutidos todos os aspectos, desde a técnica de segmentação até a métrica de avaliação da qualidade de vídeo.

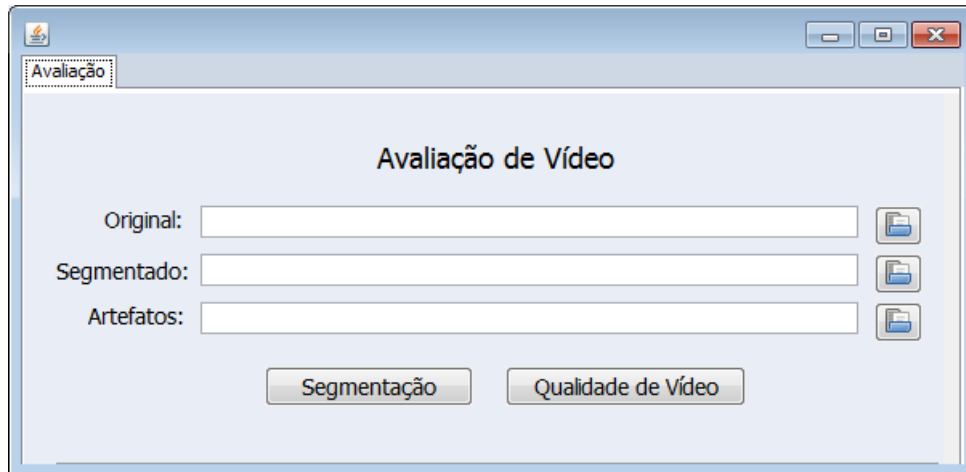
#### 3.1 MATERIAIS

Ao longo do desenvolvimento do trabalho foram utilizadas diversas ferramentas de software para a implementação, controle e visualização dos resultados. Alguns deles são: PYUV (PYUV, 2009), Netbeans (NETBEANS, 2009) e IVQUEST (MURTHY; KARAM, 2011). O primeiro passo foi a escolha da fonte de vídeo sem compressão (formato YUV), onde os valores de pixels são armazenados diretamente não sofrendo degradação. Tal formato também permite a facilidade na leitura e na manipulação dos pixels (AUWERA et al., 2008).

O software PYUV, instalado em plataforma Windows, reproduz os vídeos e é importante para visualizar e monitorar cada quadro obtido pela segmentação, permitindo a avaliação visual da qualidade da segmentação e da qualidade de vídeo. Apresenta uma interface amigável de configuração, onde é necessário informar o formato do vídeo e a taxa de quadros.

Para o desenvolvimento dos algoritmos de avaliação da qualidade de vídeo e de segmentação foi usada a linguagem de programação Java. Como IDE (Ambiente de Desenvolvimento Integrado) foi utilizada a ferramenta Netbeans com o JDK (*Java Development Kit*) 1.6. IDE é o ambiente composto por um conjunto de ferramentas (compilador, editor, etc.) que auxilia o desenvolvedor na criação de software, visando aumentar a produtividade com o ganho de agilidade no processo. O Netbeans trata-se de uma IDE gratuita e de código-aberto, que pode ser executada em Windows, Linux, Solaris e MacOS (NETBEANS, 2009). Destaca-se que esta ferramenta foi muito importante no desenvolvimento do trabalho por facilitar a criação das interfaces de interação com o usuário final e, através dela, o avaliador verificará os resultados obtidos pela segmentação, pela relevância dos objetos e pela avaliação da qualidade do vídeo.

A interface principal construída para a avaliação de vídeo é exibida na figura 22. Salienta-se que para análise dos resultados é criado, em tempo de execução, um arquivo de *log*, denominado “avaliacao” com extensão “.txt”, que conterà os valores obtidos pelas diversas características avaliadas e pela nota final atribuída pela métrica de avaliação.



**Figura 22: Tela Principal do Sistema de Avaliação.**

O software IVQUEST é um *framework* desenvolvido em Matlab cujo objetivo é avaliar a qualidade de vídeo e imagem. Foi proposto por (MURTHY, 2010) e, neste trabalho, foi usado tanto para avaliar os vídeos quanto para medir a correlação entre as métricas testadas com a avaliação subjetiva. As métricas SSIM, VQM e VSNR foram executadas e os resultados serão exibidos no capítulo 4. Para validar os testes, o software também foi útil para calcular os coeficientes de correlação (*Spearman*, *Pearson*) e prever, através de uma função logística, as notas obtidas pelas métricas testadas.

As simulações foram efetuadas em um notebook HP G42-250BR Intel Core I3 2.26 GHz com 6 GB de RAM DDDR3, rodando o sistema operacional Windows Seven. Este computador foi utilizado para o desenvolvimento do sistema de avaliação, testes, simulações e processamento dos resultados.

### 3.2 VÍDEO ORIGINAL UTILIZADO NOS TESTES

A escolha da fonte de vídeo está entre uma das mais importantes fases de desenvolvimento, pois o vídeo será utilizado tanto para a avaliação da segmentação quanto para a avaliação da qualidade de vídeo. O vídeo original “Akiyo” faz parte da base *Video Traces Research* (GROUP, 2009) e foi escolhido por se tratar de um vídeo encontrado em inúmeras pesquisas na área de avaliação e possuir o conteúdo estável, com pouca movimentação. Este vídeo está

no formato CIF (*Common Intermediate Format*) e tem 10 segundos de duração utilizando uma taxa de 30 quadros por segundo, possuindo 300 quadros no total. A figura 23 representa o primeiro quadro da sequência de vídeo Akiyo.



**Figura 23: Quadro 1 da sequência de vídeo Akiyo**

### 3.3 DESCRITORES DE REGIÕES

Esta seção é dedicada a discussão de alguns descritores de regiões que serão utilizados no cálculo das métricas para avaliação da qualidade. Os principais descritores de regiões utilizados como base para os cálculos são: área, perímetro e espessura.

#### 3.3.1 ÁREA

A área é calculada contando-se todos os *pixels* que formam o objeto. Para isso, é importante que a imagem tenha sido segmentada, de modo que os objetos estejam diferenciados (por tons ou cores distintas) (CONCI et al., 2008). O cálculo da área é realizado através de uma função  $I_x(i, j)$  descrita por um objeto mapeado de  $M \times N$  (WU et al., 2008):

$$I_x(i, j) = \begin{cases} 1, & \text{se } I(i, j) = n, \\ 0, & \text{outros casos.} \end{cases} \quad (36)$$

sendo  $n$  o número que corresponde ao objeto.



Logo, a área de pixels de um objeto  $x$  é dada por:

$$area_x = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_x(i, j), \quad (37)$$

sendo  $M$  a largura e  $N$  a altura do quadro avaliado.

A figura 24 mostra o algoritmo proposto para calcular a área de um objeto.

---

**Algoritmo** Area

---

**Entrada:** *imagemSegmentada*[largura][altura], *valorPixelObjeto*

**Saída:** *area*

```

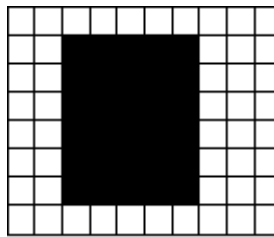
1: area ← 0
2: para  $i = 0$  até largura faça
3:   para  $j = 0$  até altura faça
4:     {Verificar se pixel pertence ao objeto.}
5:     se imagemSegmentada[ $i$ ][ $j$ ] = valorPixelObjeto então
6:       area ← area + 1
7:     fim se
8:   fim para
9: fim para
10: retorne area

```

---

**Figura 24:** Algoritmo usado para calcular a área de um objeto.

A figura 25 exibe um exemplo deste descritor, onde o objeto possui área de tamanho 30. Este valor foi obtido pela contagem de pixels pertencentes ao objeto de interesse.



**Figura 25:** Exemplo de área

### 3.3.2 PERÍMETRO

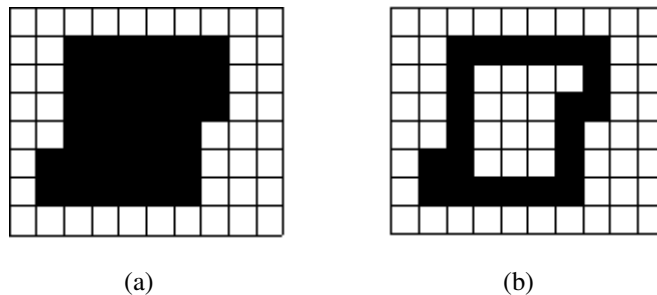
O perímetro de uma região é o tamanho (comprimento) da sua fronteira (GONZALEZ; WOODS, 2010) considerando o número de *pixels* conexos que constituem o contorno da região (CONCI et al., 2008).

Tal descritor é calculado pela contagem dos *pixels* pertencentes à borda do objeto

(PEDRINI; SCHWARTZ, 2008). Para se fazer a contagem, o primeiro passo é dado pela segmentação da imagem e diferenciação dos valores de *pixels* para os objetos. Posteriormente é verificado se o *pixel* da região do objeto apresenta como vizinho um *pixel* não pertencente a sua região. Em caso positivo, incrementa-se o valor 1 em um contador. Esse processo é realizado sobre todos os *pixels* da imagem e o resultado final é dado pela variável, na qual foi armazenada a quantidade de pixels com vizinhos pertencentes ao fundo.

É importante ressaltar que dependendo do tipo de vizinhança adotada para a verificação dos *pixels*, os resultados obtidos pelo cálculo do perímetro serão diferentes. Por exemplo, usando como referência a figura 12, o perímetro considerando uma vizinhança-4 foi de 20, diferentemente da vizinhança-8 onde o resultado seria 22 (ver figura 26). Destaca-se que existem inúmeras formas de se calcular o perímetro de uma região e que tais métodos tentam diminuir o efeito de “quantização” da imagem, que implica em erros na contagem. Existem diversos trabalhos que visam diminuir esses erros, tais como: uso de um fator de correção (WU et al., 2008), perímetro de Crofton (FRIEL, 2000), perímetro através do código de cadeia (JÄHNE, 2002), entre outros.

A figura 26 exhibe um exemplo deste descritor, onde o objeto possui perímetro de 22. Este valor foi obtido pela contagem de pixels com vizinhos pertencentes ao fundo, admitindo uma vizinhança-8.



**Figura 26: Exemplo de perímetro: (a) Imagem binária, (b) Perímetro = 22.**

O algoritmo usado para o cálculo do perímetro é ilustrado na figura 27. Destaca-se que foram considerados os limites do quadro, logo foi necessário incluir duas linhas e duas colunas com valores zeros e posteriormente incluir os valores referentes aos pixels do quadro (linhas 4-8). Outra informação relevante está presente na linha 12, onde este trabalho considerou a vizinhança-4 para o cálculo do perímetro. Ressalta-se, que o trabalho dos autores (CORREIA; PEREIRA, 2002) não informou qual o tipo de vizinhança utilizada para o cálculo do perímetro.

---

**Algoritmo** Perimetro
 

---

**Entrada:** *imagemSegmentada*[*largura*][*altura*], *valorPixelObjeto*
**Saída:** *perimetro*

```

1: Inicialize variável temp[largura + 2][altura + 2], vizinho, fundo
2: perimetro ← 0
3: fundo ← 0
4: para i = 0 até largura faça
5:   para j = 0 até altura faça
6:     temp[i + 1][j + 1] ← imagemSegmentada[i][j]
7:   fim para
8: fim para
9: para i = 1 até largura + 1 faça
10:  para j = 1 até altura + 1 faça
11:    se imagemSegmentada[i][j] <> valorPixelObjeto então
12:      Verificar se um dos vizinhos, considerando vizinhança-4, pertence ao fundo
13:      se vizinho = fundo então
14:        perimetro ← perimetro + 1
15:      fim se
16:    fim se
17:  fim para
18: fim para
19: retorne perimetro

```

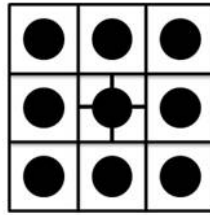
---

**Figura 27:** Algoritmo proposto para calcular o perímetro de um objeto.

### 3.3.3 ESPESSURA

A espessura de um determinado objeto é calculada através da aplicação dos conceitos da morfologia matemática utilizando a técnica de erosão. O valor da espessura é dado pelo número de iterações que a técnica de erosão necessita para desaparecer com o objeto. Os conceitos da morfologia matemática podem ser vistos no capítulo 2.

O primeiro passo para o cálculo é dado pela definição do elemento estruturante. O elemento estruturante utilizado neste trabalho está representado na figura 28 (3x3 cheio) e é conhecido como *sebox*. Esta escolha foi baseada na ampla utilização deste em diversos trabalhos na área de processamento digital de imagens. É importante ressaltar, que os autores (CORREIA; PEREIRA, 2002) também não apresentam qual o elemento estruturante utilizado em sua avaliação.



**Figura 28: Elemento Estruturante utilizado no trabalho.**

Fonte: Adaptado de (GONZALEZ; WOODS, 2010)

A figura 29 apresenta o algoritmo utilizado neste trabalho para calcular a espessura de um determinado objeto.

---

**Algoritmo** Espessura

---

**Entrada:** *imagemSegmentada*[largura][altura], *elementoEstruturante*[3][3]

**Saída:** *espessura*

```

1: Inicialize temp[largura][altura], modificacao, espessura
2: temp ← imagemSegmentada
3: modificacao ← VERDADEIRO
4: espessura ← 0
5: enquanto modificacao faça
6:   para i = 1 até largura – 1 faça
7:     para j = 1 até altura – 1 faça
8:       Comparar todas as posições do elementoEstruturante com a imagemSegmentada
          iniciando do pixel central (i, j) verificando a vizinhança-8
9:       se elementoEstruturante = imagemSegmentada então
10:        temp[i][j] ← 255
11:        modificacao ← VERDADEIRO
12:       senão
13:        temp[i][j] ← 0
14:       fim se
15:     fim para
16:   fim para
17: se modificacao então
18:   imagemSegmentada ← temp
19:   espessura ← espessura + 1
20:   Zerar matriz temp
21: fim se
22: fim enquanto
23: retorne espessura

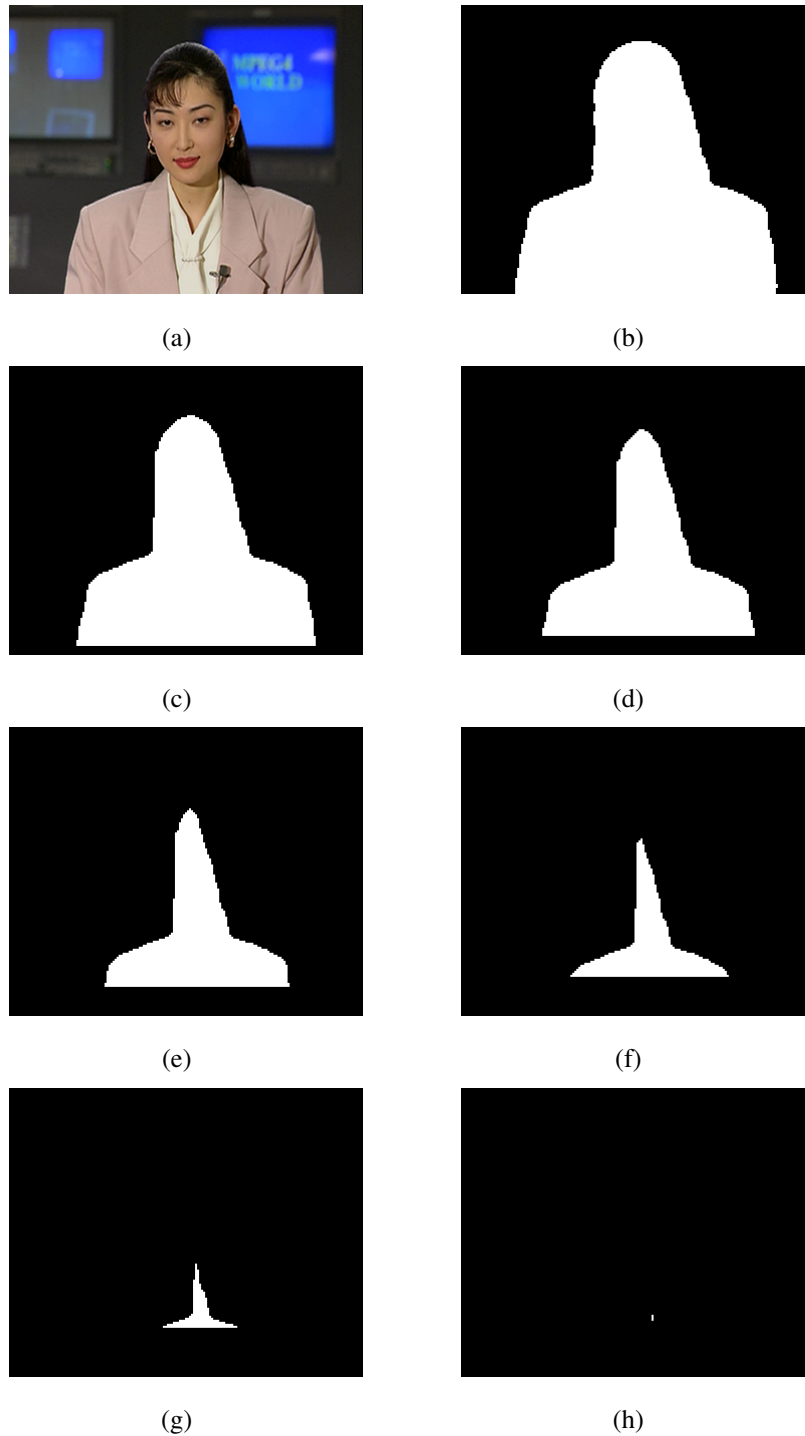
```

---

**Figura 29: Algoritmo proposto para calcular a espessura de um objeto.**

A figura 30 exhibe os resultados obtidos por este algoritmo quando aplicado sobre o

quadro 5 do vídeo “Akiyo” segmentado. Baseado na análise da imagem é possível concluir que a espessura do objeto de interesse é igual a 57, pois este é o número de iterações que foram aplicadas até o desaparecimento completo do objeto.



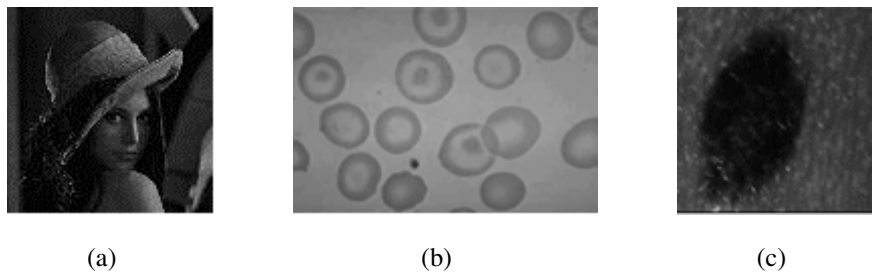
**Figura 30: Cálculo da Espessura: (a) Original, (b) Imagem Segmentada, (c) 10 iterações, (d) 20 iterações, (e) 30 iterações, (f) 40 iterações, (g) 50 iterações e (h) 57 iterações.**

### 3.4 SEGMENTAÇÃO

Neste trabalho são abordadas duas técnicas de segmentação. A primeira tem como objetivo propor o uso da meta-heurística ACO (Otimização por Colônia de Formigas) aplicado no problema de segmentação. A segunda trata de um novo modelo para detectar regiões de atenção visual em vídeo digital proposto por (BEZERRA, 2011).

#### 3.4.1 OTIMIZAÇÃO POR COLÔNIA DE FORMIGAS NA SEGMENTAÇÃO DE IMAGENS

As imagens utilizadas nos testes preliminares da técnica de segmentação foram: Lena, Células, Mancha. A imagem Lena, exibida na figura 31(a), cujo tamanho é de 102 pixel de largura por 102 de altura, totaliza 10404 pixels na imagem. A imagem Células é exibida na figura 31(b) com 128 de largura por 87 de altura, contendo 11136 pixels. A imagem Mancha (RAMOS; MUGE, 2000), apresentada na figura 31(c) tem como dimensão 133 de largura por 134 de altura, apresentando 17822 pixels.



**Figura 31: Imagens utilizadas no experimento: (a) Lena, (b) Células, (c) Mancha.**

**Fonte: (PEREIRA et al., 2010c)**

As três imagens estão em escala de cinza e seus valores foram normalizados. É importante salientar que a imagem Células (figura 31(b)) foi utilizada como base para ajustar os diversos parâmetros do algoritmo e medir a influência destes nos resultados obtidos pela técnica proposta. Após a definição da melhor configuração encontrada, os parâmetros serão aplicados para as demais imagens testadas.

Os seguintes parâmetros foram constantes para todas as execuções:

1. Quantidade de centros: 2;
2. Número de iterações: 100;

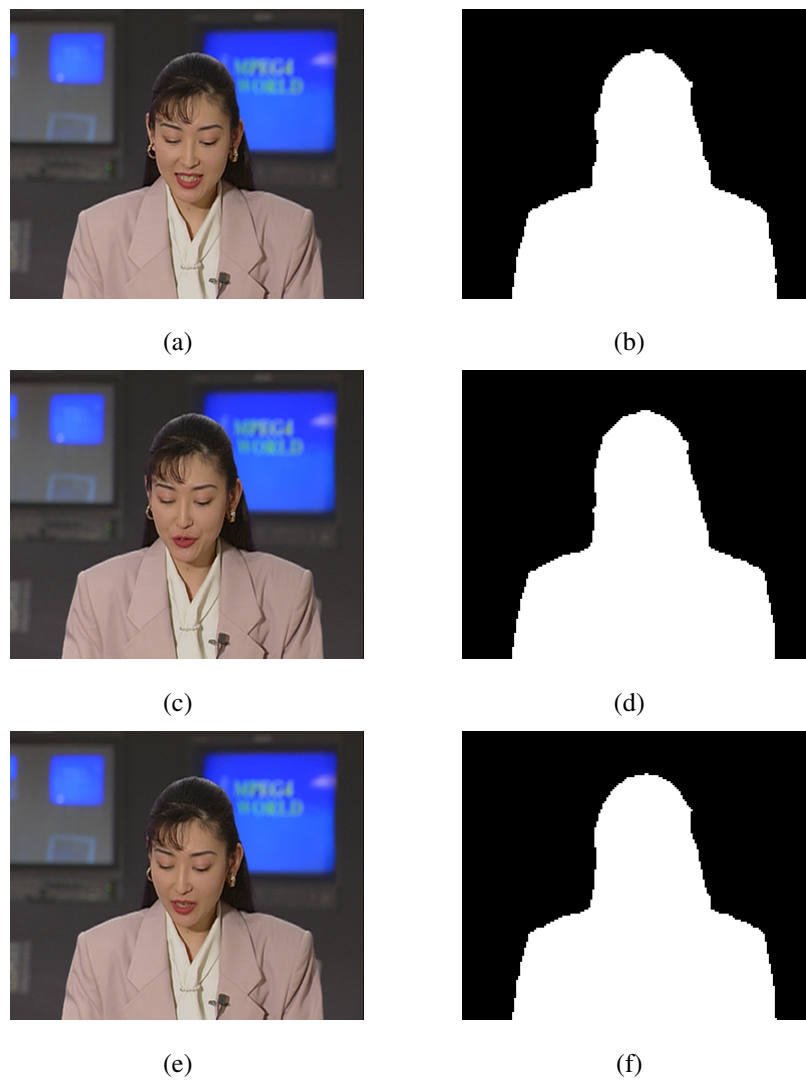
3. Repetições: 10;
4. Quantidade de formigas: 10;
5. Influência do fator de evaporação ( $\rho$ ): 0,5;
6. Influência da heurística: 5;
7. Cada cluster deve possuir no mínimo um objeto.

Para ajustar a configuração, os seguintes parâmetros foram testados:

1. Probabilidade de intensificação ( $q_0$ ): 0,4 e 0,8;
2. Inicialização dos centros: aleatório e os mais distantes possíveis um do outro;
3. Técnica: ACO puro e ACO com busca local.

### 3.4.2 SEGMENTAÇÃO UTILIZANDO MDRAV

Os resultados da segmentação pela técnica MDRAV são exibidos na figura 32. Este vídeo segmentado será utilizado como parâmetro de entrada para os métodos de avaliação.

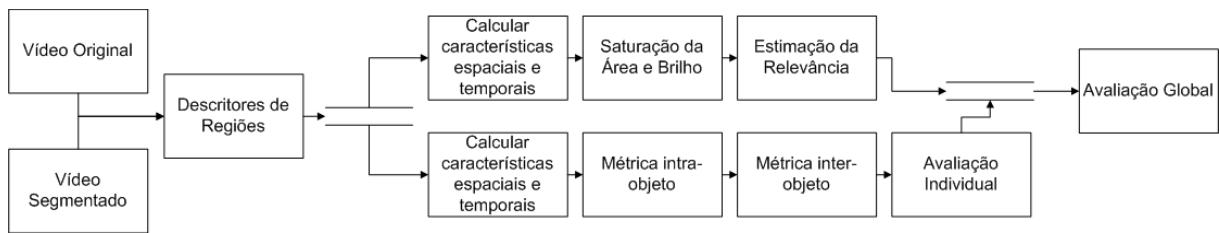


**Figura 32: Resultado da Segmentação: (a) Quadro 100 original, (b) Quadro 100 segmentado, (c) Quadro 200 original, (d) Quadro 200 segmentado, (e) Quadro 300 original, (f) Quadro 300 segmentado.**

### 3.5 AVALIAÇÃO DA QUALIDADE DE SEGMENTAÇÃO

Nesta parte do trabalho são descritos os passos para avaliar a qualidade de segmentação. O software desenvolvido manipula tanto valores para as três componentes de cor (RGB) quanto de luminância (Y). Por exemplo, as métricas temporais são calculadas utilizando somente a componente de luminância enquanto a métrica de contraste utilizará as três componentes de cor. O sistema de avaliação está ilustrado na figura 33 e as etapas desenvolvidas serão vistas a seguir.





**Figura 33: Diagrama do sistema de avaliação da qualidade de segmentação.**

Nesta dissertação foram implementadas as métricas propostas por (CORREIA; PEREIRA, 2002) para avaliar a qualidade de segmentação. Os autores propõem métricas de avaliação independente onde não existe uma segmentação de referência para comparação.

A avaliação independente pode ser dividida em dois principais tipos: avaliação individual dos objetos e avaliação geral. Na avaliação individual, os objetos segmentados são avaliados de forma independente baseados na qualidade de segmentação. Sua principal aplicação é quando os objetos podem ser individualmente manipulados, como por exemplo, na criação de uma nova cena utilizando o objeto segmentado (GASPAR, 2009).

Em comparação, a avaliação da qualidade geral avalia por completo o quadro de vídeo, considerando os objetos como componentes de uma partição em um quadro. Esta métrica utiliza conceitos do SVH para verificar os valores e os pesos que são dados para cada objeto, já que eles, na maioria das vezes, não recebem a mesma atenção do observador humano. Ressalta-se que a implementação da métrica individual é um pré-requisito desta avaliação. Ou seja, tal métrica é uma evolução da avaliação individual, pois insere pesos baseados na relevância dos objetos na cena. Sua principal função é determinar se o algoritmo de segmentação é apropriado para um tipo de aplicação, como por exemplo, câmeras de vigilância.

Destaca-se que as métricas propostas em (CORREIA; PEREIRA, 2002) são baseadas tanto no conhecimento sobre a avaliação da qualidade de segmentação quanto nos aspectos relevantes da avaliação da qualidade de vídeo.

### 3.5.1 AVALIAÇÃO INDIVIDUAL

A avaliação individual da qualidade de segmentação é fundamentada em dois tipos de métricas: intra-objeto e inter-objeto. A métrica intra-objeto é baseada na homogeneidade esperada pela característica de cada objeto. Já a métrica inter-objeto é fundamentada pelas diferenças observadas, em função do tempo, nas regiões de vizinhança do objeto.

### 3.5.1.1 MÉTRICA INTRA-OBJETO

A primeira métrica, intra-objeto, é calculada pela média das características tanto espaciais quanto temporais do objeto. A característica espacial utilizada na avaliação individual é chamada regularidade de forma. A regularidade de forma pode ser computada pelas características geométricas, entre elas, o alongamento (*elong*), a compactação (*compact*) e também uma combinação entre o alongamento e a circularidade (*circ\_compact*) do objeto. Os cálculos destas características são dados por:

$$compact(E) = \max\left(\frac{perimeter(E)^2}{75 \times area(E)}, 1\right), \quad (38)$$

$$circ\_elong(E) = \max\left(circ(E), \max\left(\frac{elong(E)}{5}\right), 1\right). \quad (39)$$

A circularidade (*circ*) e o alongamento (*elong*) são obtidos por:

$$circ(E) = \frac{4 \times \Pi \times area(E)}{perimeter^2(E)}, \quad (40)$$

$$elong(E) = \frac{area(E)}{(2 \times thickness(E))^2}. \quad (41)$$

Os valores constantes foram normalizados de acordo com um exaustivo conjunto de testes efetuados por (CORREIA; PEREIRA, 2002). Os cálculos de área (*area*), perímetro (*perimeter*) e espessura (*thickness*) estão representados nas figuras 24, 27 e 29, respectivamente. *E* representa o objeto que está sendo avaliado.

A característica temporal foi avaliada pela estabilidade e pela uniformidade de movimento. A estabilidade temporal tem como objetivo testar a evolução suave das características dos objetos em relação ao tempo. Funções usadas na medida da estabilidade são: a variância de textura, o alongamento, a criticidade, a posição, entre outros. Neste trabalho são usadas as funções:

$$size_{diff} = \|area(E_t) - area(E_{t-1})\|, \quad (42)$$

$$elong_{diff} = \|elong(E_t) - elong(E_{t-1})\|, \quad (43)$$

$$crit_{diff} = \|crit(E_t) - crit(E_{t-1})\|. \quad (44)$$

A uniformidade de movimento pode ser avaliada por diversas características, tais como, a variância no vetor de movimento ou pela criticidade. A criticidade foi adotada como característica para medir a uniformidade de movimento.

A criticidade foi proposta por Wolf e Webster (WOLF; WEBSTER, 1997) e é calcu-

lada utilizando dois parâmetros de entrada obtidos pela estimação da informação temporal (TI) e da informação espacial (SI). Os autores comentam a grande dificuldade na codificação de vídeo para que este consiga manter um nível constante de qualidade, principalmente quando existem muitos detalhes espaciais e uma alta variação de movimento. O objetivo do trabalho foi descrever um modelo que faça a medição da criticidade de uma cena utilizando conceitos subjetivos da qualidade de vídeo. Como vantagem, os autores citam que as características utilizadas para avaliar são simples de calcular, com isso os resultados são obtidos em tempo real.

A característica de informação espacial (SI) é dada por:

$$SI(t_n) = rms_{space} [Sobel(F(t_n))], \quad (45)$$

Enquanto a informação temporal (TI) é calculada por:

$$TI(t_n) = rms_{space} [F(t_n) - F(t_{n-1})], \quad (46)$$

onde  $F(t_n)$  é a luminância do quadro do vídeo no instante  $n$  e  $rms$  é a média quadrática. A média quadrática é obtida pela aplicação da raiz quadrada sobre os valores elevados ao quadrado. Este operador é usado para fornecer uma medida da energia total. O filtro de Sobel utilizado para calcular a informação espacial, é implementado pela convolução de duas matrizes 3x3 sobre o quadro do vídeo. Após esta etapa é aplicada a raiz quadrada da soma dos quadrados dos resultados dessas convoluções (ITU-T, 2008).

Para  $y = Sobel(x)$ , dado que  $x(i, j)$  representa a imagem de entrada na linha  $i$ -ésima e coluna  $j$ -ésima,  $Gv(i, j)$  é obtido pela primeira convolução expressa como:

$$Gv = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \times \begin{bmatrix} i-1, j-1 & i, j-1 & i+1, j-1 \\ i-1, j & i, j & i+1, j \\ i-1, j+1 & i, j+1 & i+1, j+1 \end{bmatrix} \quad (47)$$

O resultado da segunda convolução,  $Gh(i, j)$ , é obtido por:

$$Gh = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \times \begin{bmatrix} i-1, j-1 & i, j-1 & i+1, j-1 \\ i-1, j & i, j & i+1, j \\ i-1, j+1 & i, j+1 & i+1, j+1 \end{bmatrix} \quad (48)$$

Logo, o pixel resultante utilizando o filtro de Sobel sobre a imagem, na linha  $i$ -ésima e

na coluna  $j$ -ésima é calculado por:

$$y(i, j) = \sqrt{[Gv(i, j)]^2 + [Gh(i, j)]^2} \quad (49)$$

Os cálculos do filtro Sobel são feitos para todo  $2 \leq i \leq N - 1$  e  $2 \leq j \leq M - 1$ , onde  $N$  é o número de linhas e  $M$  é o número de colunas. O algoritmo para calcular o filtro de Sobel é apresentado na figura 34. É importante observar que quando os valores obtidos pelas convoluções são superiores a 255 (a maior intensidade de pixel admitida), o valor do pixel é truncado (linha 27). No trabalho de (CORREIA; PEREIRA, 2002), os autores não divulgam como tratam este caso em específico. Como consequência, o valor do pixel pode ultrapassar o maior valor suportado para uma imagem representada em 8 bits.

---

**Algoritmo** Filtro de Sobel
 

---

**Entrada:** *imagemSegmentada*[largura][altura]

**Saída:** *imagemFiltrada*

```

1: Inicialize temp, vetorGv[9], vetorGh[9], imagemFiltrada[largura][altura]
2: para i = 1 até largura - 1 faça
3:   para j = 1 até altura - 1 faça
4:     {Primeira Convolução}
5:     vetorGv[0] ← imagemSegmentada[i - 1][j - 1] × -1
6:     vetorGv[1] ← imagemSegmentada[i - 1][j] × -2
7:     vetorGv[2] ← imagemSegmentada[i - 1][j + 1] × -1
8:     vetorGv[3] ← imagemSegmentada[i][j - 1] × 0
9:     vetorGv[4] ← imagemSegmentada[i][j] × 0
10:    vetorGv[5] ← imagemSegmentada[i][j + 1] × 0
11:    vetorGv[6] ← imagemSegmentada[i + 1][j - 1] × 1
12:    vetorGv[7] ← imagemSegmentada[i + 1][j] × 2
13:    vetorGv[8] ← imagemSegmentada[i + 1][j + 1] × 1
14:    {Segunda Convolução}
15:    vetorGh[0] ← imagemSegmentada[i - 1][j - 1] × -1
16:    vetorGh[1] ← imagemSegmentada[i - 1][j] × 0
17:    vetorGh[2] ← imagemSegmentada[i - 1][j + 1] × 1
18:    vetorGh[3] ← imagemSegmentada[i][j - 1] × -2
19:    vetorGh[4] ← imagemSegmentada[i][j] × 0
20:    vetorGh[5] ← imagemSegmentada[i][j + 1] × 2
21:    vetorGh[6] ← imagemSegmentada[i + 1][j - 1] × -1
22:    vetorGh[7] ← imagemSegmentada[i + 1][j] × 0
23:    vetorGh[8] ← imagemSegmentada[i + 1][j + 1] × 1
24:    Somar todos os valores de vetorGv
25:    Somar todos os valores de vetorGh
26:    temp = sqrt(somaGv2 + somaGh2)
27:    se temp > 255 então
28:      temp ← 255
29:    fim se
30:    imagemFiltrada[i][j] ← temp
31:   fim para
32: fim para
33: retorne imagemFiltrada

```

---

**Figura 34: Algoritmo proposto para aplicação do Filtro de Sobel.**

Após calcular as informações temporais e espaciais é efetuado o cálculo da criticidade, definido como:

$$crit = 4.68 - 0.54 \times p1 - 0.46 \times p2, \quad (50)$$

sendo

$$p1 = \log_{10} \{ \text{mean}_{time} [SI(t_n) \times TI(t_n)] \}, \quad (51)$$

$$p2 = \log_{10} \{ \max_{time} [ \text{abs}(SI(t_n) - SI(t_{n-1})) ] \}. \quad (52)$$

O valor de  $p1$  é uma medida do valor médio (sobre o tempo) do produto instantâneo de  $SI$  e  $TI$ , enquanto o parâmetro  $p2$  é a medida da variação máxima de  $SI$  quadro a quadro. Os valores constantes foram obtidos através de uma bateria de testes efetuada por (WOLF; WEBSTER, 1997).

O cálculo da criticidade retorna um valor que representa a dificuldade da cena em ser codificada, ou seja, quanto maior o valor de criticidade mais complexa é a cena.

### 3.5.1.2 MÉTRICA INTER-OBJETO

A métrica inter-objeto avalia se os objetos foram corretamente separados utilizando as informações dos pixels de fronteira. No cálculo são empregadas características como, por exemplo, a diferença entre os pixels internos comparados com a parte externa do objeto e a comparação entre objetos vizinhos utilizando-se o vetor de movimento. Neste trabalho é utilizado o contraste local dos vizinhos.

A métrica de contraste local foi usada para avaliar se existe uma expressiva diferença entre o contraste local da parte interna quando comparada com a parte externa. Essa diferença é feita ao longo da região de borda do objeto e seu cálculo é dado por:

$$contrast = \frac{1}{4 \times 255 \times N_b} \times \sum_{i,j} (2 \times \max(DY_{ij}) + \max(DU_{ij}) + \max(DV_{ij})), \quad (53)$$

onde  $N_b$  é a quantidade de pixels de borda,  $DY$ ,  $DU$  e  $DV$  são as diferenças entre os pixels de borda calculadas com vizinhança-4, utilizando as componentes Y, U e V, respectivamente.

### 3.5.1.3 CONSTRUÇÃO DA MÉTRICA DE AVALIAÇÃO INDIVIDUAL

A composição da métrica para calcular a qualidade de segmentação individual é dada por:

$$Inst\_seg\_qual\_std\_stable_i = intra + inter, \quad (54)$$

onde

$$intra = 0.30 \times (shape\_regi) + 0.33 \times temp\_stab_i \quad (55)$$

$$shape\_regi = 0.5 \times circ\_elong_i + 0.5 \times compact_i, \quad (56)$$

$$temp\_stab_i = 0.33 \times size\_diff + 0.33 \times elong\_diff + 0.33 \times critdiff, \quad (57)$$

$$inter = 0.37 \times contrast. \quad (58)$$

Os pesos utilizados nas métricas foram os mesmos dentro de cada classe, devido ao fato deles serem considerados entre si como complementares, ou seja, não há uma preferência por uma determinada característica. Outro ponto que merece destaque, refere-se as pesos relativos das diversas classes de características que compõem a métrica, onde estes valores foram selecionados para refletir a maior influência do aspecto espacial (77%) em relação ao aspecto temporal (33%). Além disso, ressalta-se que estes pesos foram estipulados por um conjunto de testes realizado por (CORREIA; PEREIRA, 2003).

Neste ponto do trabalho é importante abrir um parênteses para os tipos de conteúdo de vídeo: estável ou estático e com movimentação. No conteúdo estável enquadram-se vídeos utilizados em aplicações onde os objetos apresentam formas regulares e são estáveis temporalmente, como por exemplo, o vídeo “Akiyo”. Em comparação, os vídeos com movimentação são importantes para as aplicações onde a regularidade de forma não é tão importante quando comparado com a uniformidade de movimento. Assim, se a regularidade de forma já é esperada no vídeo, então é possível assumir que esta métrica tem menor importância. Um exemplo deste conteúdo pode ser visto no vídeo Stefan (CORREIA; PEREIRA, 2002).

Por fim, a equação 54 representa a métrica utilizada para avaliar um instante de tempo ( $t$ ) cujo conteúdo é estável. Logo, para obter a qualidade de segmentação de um objeto  $i$  ( $Seg\_qual\_std\_stable$ ) faz-se uma média temporal de todos os valores obtidos nos instantes avaliados.

### 3.5.2 AVALIAÇÃO GLOBAL

A avaliação global tem como objetivo unir a avaliação individual da qualidade de segmentação de cada objeto, com o valor correlacionado da relevância. A avaliação global da segmentação é computada por:

$$Seg\_qual = \sum_i (Seg\_qual\_std\_stable(E_i) \times Relevance\_context(E_i)), \quad (59)$$

sendo  $Relevance\_context$  a relevância do objeto que está sendo avaliado e  $Seg\_qual\_std\_stable$  a nota dada pela qualidade de segmentação. Destaca-se que a soma é executada sobre todos os objetos segmentados. No trabalho de (CORREIA; PEREIRA, 2002), além da expressão (59) os autores utilizam um fator de ponderação para verificar se todos os objetos foram detectados e se estes são corretos. Neste trabalho não se utiliza tal ponderador, pois o vídeo “Akiyo” apresenta apenas dois objetos (fundo e apresentadora) e ambos foram segmentados.

### 3.6 ESTIMAÇÃO DA RELEVÂNCIA DO OBJETO

Esta seção é dedicada a discussão de como é calculada a relevância de um determinado objeto em uma cena de vídeo. Neste trabalho foi implementada a métrica proposta por (CORREIA; PEREIRA, 2000).

Os autores (CORREIA; PEREIRA, 2000), além de considerar as características do SVH, utilizam algumas métricas elementares para estimar a relevância de um objeto. Tais métricas são fundamentadas em características espaciais e temporais. Cita-se as seguintes características: atividade de movimento, complexidade do objeto, forma do objeto, brilho e intensidade do vermelho.

O movimento é uma das mais importantes características, visto que recebe um foco maior de atenção do SVH. Existem dois tipos de métricas que mensuram o movimento: a soma média absoluta entre os componentes do vetor de movimento do objeto e a característica TI. Neste trabalho é utilizada esta última.

A TI é baseada na característica da diferença do movimento,  $M_n(i, j)$ , entre os valores do pixel na mesma posição mas em instantes distintos ( $t$  e  $t - 1$ ) (ITU-T, 2008). Este cálculo é feito utilizando a luminância do pixel.  $M_n(i, j)$  é definida como uma função temporal calculada por:

$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j) \quad (60)$$

onde  $F_n(i, j)$  é o pixel da  $i$ -ésima linha e  $j$ -ésima coluna no quadro  $n$ .

A medida de informação temporal é calculada através do desvio-padrão aplicado sobre todo  $M_n$  que está contido no conjunto ( $std_{space}$ ). Após, é selecionado o maior valor obtido para todo o tempo avaliado ( $max_{time}$ ). Por exemplo, o vídeo avaliado é composto por 300 quadros ( $max_{time}$ ), onde é aplicado a equação (60) para cada instante sobre todos  $i$  e  $j$ . Em seguida, com os valores calcula-se o desvio-padrão de todos os quadros e o maior valor obtido é assumido como TI.

$$TI = max_{time} \{std_{space} [M_n(i, j)]\} \quad (61)$$

Destaca-se que, quanto maior a quantidade de movimento entre os quadros adjacentes maior serão os valores de TI.

A segunda característica é a complexidade do objeto, determinada pela SI e pela criti-



cidade. A criticidade foi discutida na seção 3.5.1.1 e seu cálculo é dado pela equação (50).

A SI aplica sobre cada quadro do vídeo no instante  $n$  o filtro de Sobel [ $Sobel(F_n)$ ]. Após esta etapa executa-se o cálculo do desvio-padrão sobre os pixels ( $std_{space}$ ) em cada quadro filtrado. Esta operação é repetida para cada quadro da sequência e os resultados formam uma série cronológica das informações espaciais da cena. O valor máximo na série cronológica ( $max_{time}$ ) é escolhido para representar o índice SI (ITU-T, 2008). Este processo é expresso pela equação:

$$SI = max_{time} \{std_{space}[Sobel(F_n)]\}. \quad (62)$$

O filtro de Sobel foi discutido anteriormente e seu algoritmo apresentado na figura 34.

A terceira característica é dada pela forma do objeto, cujo objetivo é determinar a preferência do SVH quando deparado com um determinado tipo de forma e orientação. Esta característica foi computada pelas funções alongamento e compactação ( $elong\_compact$ ) e a circularidade ( $circ$ ).

$$circ = \frac{(4 \times \pi \times area)}{(perimeter)^2}, \quad (63)$$

$$elong\_compact = \frac{elong}{10} + \frac{compactness}{255}, \quad (64)$$

onde

$$elong = \frac{area}{(2 \times thickness)^2}, \quad (65)$$

$$compactness = \frac{perimeter^2}{area}. \quad (66)$$

A área, o perímetro e a espessura já foram mencionados anteriormente e os algoritmos que representam tais cálculos podem ser vistos nas figuras 24, 27, 29, respectivamente. Os valores fixos (10, 255) tem como objetivo ponderar as características avaliadas e são usados para normalizar, juntamente com a operação de truncamento, os valores no limite [0,1]. Estes pesos foram obtidos através de um exaustivo conjunto de testes efetuados por (CORREIA; PEREIRA, 2006).

A quarta característica é obtida pelo tamanho do objeto, o qual é computado pela área do objeto. Sabe-se que se um objeto ocupa mais de 25% de todo quadro, este merece um fator de saturação. O algoritmo para calcular a área é visto na figura 24. A saturação foi calculada

por:

$$size = \begin{cases} 4 \times \frac{area(k)}{area(I)}, & \text{para } 4 \times area(k) < area(I), \\ 1, & \text{para } 4 \times area(k) \geq area(I), \end{cases} \quad (67)$$

onde  $k$  representa a área do objeto que está sendo avaliado e  $I$  é o tamanho do quadro (largura x altura).

A quinta e última característica, está relacionada a luminância. Destaca-se que o luminância e as cores dos objetos chamam mais atenção visual, fazendo com que os observadores humanos tenham um foco especial nessas regiões. A característica é calculada pela função:

$$BR = \frac{3 \times avgY(I) + avgV(I)}{4 \times 255}, \quad (68)$$

sendo  $avgY$  e  $avgV$  os valores médios da diferença da componente Y (luminância) e da componente V do quadro  $I$ . O espaço de cor utilizado foi o YUV. A multiplicação por 255 no denominador foi utilizada para que a característica gere valores no limite [0,1].

### 3.6.1 COMPOSIÇÃO DA MÉTRICA PARA ESTIMAÇÃO DA RELEVÂNCIA

A métrica proposta por (CORREIA; PEREIRA, 2000) para estimação da relevância dos objetos em um vídeo é dada pela equação (69). É importante destacar algumas modificações em relação ao que foi proposto neste trabalho. Por exemplo, os autores citados acima utilizaram o vetor de movimento juntamente com a TI. Esta característica não foi computada neste trabalho, pois o vídeo testado apresenta conteúdo do tipo estável, onde a maior relevância é dada para a forma do objeto e não pela sua movimentação.

$$RelIndividual = \frac{MA + COMP + SHAPE + 0.5 \times size + BR}{9.25}, \quad (69)$$

onde

$$MA = 3.5 \times TI, \quad (70)$$

$$COMP = 1.5 \times SI + 1.5 \times crit, \quad (71)$$

$$SHAPE = 0.5 \times circ + 0.75 \times elong\_compact. \quad (72)$$

A equação (69) é usada para o cálculo da relevância individual do objeto na cena. Os valores constantes foram definidos pelo impacto de cada métrica em relação a atenção visual humana. Além disso, tais pesos foram obtidos por uma bateria de testes efetuada por (CORREIA; PEREIRA, 2006). Destaca-se que os pesos maiores foram indicados para a característica

de movimento (TI) e métricas de complexidade (COMP) .

O objetivo desta avaliação é dado pela importância do contexto onde o objeto está localizado, ou seja, quando se quer medir a qualidade global de uma segmentação, este é o cálculo de relevância adequado. Sabe-se que a soma dos valores na relevância individual não é necessariamente, igual a 1. Logo, para obter a relevância contextual utiliza-se a seguinte equação:

$$RelCont_{kt} = \frac{RI_{kt}}{\sum_{j=1}^{num\_obj} RI_{jt}}, \quad (73)$$

onde  $RelCont_{kt}$  é a métrica de relevância contextual para o objeto  $k$ , no instante  $t$ ,  $RI_{kt}$  representa a relevância individual do objeto  $k$ , e  $num\_obj$  é quantidade de objetos avaliados.

A relevância global é calculada pela média de todos os instantes avaliados.

### 3.7 GERAÇÃO DE VÍDEOS COM ARTEFATOS

Os vídeos com degradações são resultados do gerador de artefatos proposto por (ALBINI, 2009). Foram gerados vídeos com cinco diferentes degradações: ruído branco gaussiano, ruído sal-e-pimenta, borramento, blocagem e travamento.

#### 3.7.1 RUÍDO BRANCO GAUSSIANO

A equação (74) mostra a função utilizada para gerar o artefato ruído branco no vídeo.

$$p(i, j) = \begin{cases} Y[i, j] = O[i, j] + R & \text{se } X < P \\ Y[i, j] = O[i, j] & \text{caso contrário,} \end{cases} \quad (74)$$

onde  $p$  é o pixel que sofrerá ou não a adição do artefato,  $Y$  é o quadro modificado,  $O$  representa o quadro original,  $i$  e  $j$  são as coordenadas do pixel no quadro,  $P$  representa a probabilidade de um determinado pixel ser selecionado,  $X$  é a variável (distribuição uniforme obtida de forma aleatória) que determina a escolha de um pixel e  $R$  é o ruído branco gerado por uma variável aleatória que apresenta distribuição gaussiana, média zero e desvio-padrão de 40. Tais valores constantes foram estipulados em (ALBINI, 2009).

Foram testados 10 vídeos com nível crescente de degradação: 2%, 4%, 6%, 8%, 10%, 15%, 20%, 30%, 50% e 100%. Cada porcentagem representa a quantidade de pixels alterados por este artefato. É possível visualizar os resultados obtidos pelo gerador na figura 35.



**Figura 35: Quadros com ruído branco gaussiano: (a) Original, (b) 2% de ruído branco, (c) 10% de ruído branco, (d) 15% de ruído branco, (e) 20% de ruído branco, (f) 30% de ruído branco, (g) 50% de ruído branco, (h) 100% de ruído branco.**

### 3.7.2 RUÍDO SAL-E-PIMENTA

O ruído sal-e-pimenta é obtido através da função:

$$p(i, j) = \begin{cases} Y[i, j] = 255 \times (X \text{MOD} 2) & \text{se } R < P \\ Y[i, j] = O[i, j] & \text{caso contrário,} \end{cases} \quad (75)$$

sendo  $p$  o pixel que sofrerá ou não a adição do artefato,  $i$  e  $j$  são as coordenadas do pixel no quadro modificado  $Y$ ,  $X$  e  $R$  são variáveis aleatórias que apresentam distribuição uniforme,  $P$  representa a probabilidade de cada pixel ser selecionado constituinte do quadro, de forma que o valor do pixel seja alterado,  $O$  é o quadro original,  $MOD$  é o operador utilizado para retornar o resto inteiro da divisão, ou seja, caso o valor seja ímpar a função retornará 1, senão retornará 0.

Foram avaliados 5 diferentes sequências com níveis de degradação ( $P$ ) de: 1%, 2%, 3%, 4% e 5%. Os resultados são exibidos na figura 36.



**Figura 36: Quadros com ruído sal-e-pimenta: (a) Original, (b) 1% de probabilidade, (c) 2% de probabilidade, (d) 3% de probabilidade, (e) 4% de probabilidade e (f) 5% de probabilidade.**

### 3.7.3 BORRAMENTO

O borramento apresentou duas abordagens de cálculo: máscara linear vertical e matricial. O resultado do gerador é obtido através da utilização do filtro da média, onde uma máscara (matriz) define quais os pixels envolvidos no cálculo. Destaca-se que o valor da média utiliza as informações de luminância, o qual é atribuído ao pixel central. Os passos para o cálculo do filtro são: (a) definição do tamanho da máscara: 3 linear; (b) seleção dos pixels vizinhos (1,3) e central (2); (c) cálculo da média com os valores e (d) atribuição deste valor no pixel central; (e) percorrer toda a imagem e substituir os valores centrais. Um exemplo de máscara linear vertical é representado na equação (76).

$$M = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \quad (76)$$

O resultado é dado pela média aritmética desses pixels. Sendo assim, o borramento com máscara linear será calculado conforme:

$$Y[i, j] = \frac{1}{n} \left( \sum_{k=i-\frac{(n-1)}{2}}^{k=i+\frac{(n-1)}{2}} O[k, j] \right), \quad (77)$$

onde  $Y$  é o quadro modificado,  $n$  é o tamanho da máscara,  $i$  e  $j$  são as coordenadas do pixel no quadro e  $O$  é o quadro original.

Outra abordagem é dada pela utilização de uma máscara matricial, (equação (78)). O cálculo segue a mesma sequência de passos comentada anteriormente e a geração do artefato é feita através da equação (79).

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (78)$$

$$Y[i, j] = \frac{1}{n^2} \left( \sum_{k=i-\frac{(n-1)}{2}}^{k=i+\frac{(n-1)}{2}} \sum_{l=j-\frac{(n-1)}{2}}^{l=j+\frac{(n-1)}{2}} O[k, l] \right), \quad (79)$$

onde  $Y$  é o quadro gerado com o borramento,  $n$  é o tamanho da máscara matricial ( $n = 5$  representa uma matriz  $5 \times 5$ ),  $O$  é o quadro original e  $i$  e  $j$  representam as coordenadas do pixel no quadro.

Foram testados 6 vídeos, sendo 3 vídeos borrados com uma máscara linear (tamanhos de 3, 5 e 7 pixels) e os outros 3 através com máscaras matriciais (tamanhos de  $3 \times 3$ ,  $5 \times 5$  e  $7 \times 7$  pixels). Os resultados obtidos são apresentados na figura 37.



**Figura 37: Quadros com borramento: (a) Original, (b) com máscara linear 3, (c) com máscara linear 5, (d) com máscara linear 7, (e) com máscara matricial 3x3, (f) com máscara matricial 5x5 e (g) com máscara matricial 7x7**

#### 3.7.4 BLOCAGEM

O algoritmo de blocagem utilizou como parâmetro de entrada um gerador de números aleatórios com uma distribuição uniforme. O gerador tem como função selecionar os blocos



que serão alterados, selecionar a duração do artefato e delimitar os quadros de início e fim da presença da degradação.

O algoritmo inicia percorrendo os quadros do vídeo até o momento em que um pixel é escolhido. Neste instante, o pixel tem seu valor copiado aos seus vizinhos adjacentes, considerando uma vizinhança-8. Em (ALBINI, 2009), o bloco foi definido com tamanho 8x8 devido ao tamanho dos macro-blocos encontrados nos codificadores H.264 e MPEG-2 (STOLFI, 2008).

A figura 38 ilustra o resultado obtido pela aparição deste artefato no quadro 166 da sequência “Akiyo”. Neste trabalho, foram testados 2 vídeos com: 1% e 5%.



**Figura 38: Quadro com blocagem.**

### 3.7.5 TRAVAMENTO

O algoritmo de travamento utilizou dois geradores de variáveis aleatórias com distribuição uniforme. O primeiro é responsável pela seleção do quadro inicial que sofrerá o travamento. O segundo calcula a duração deste artefato no intervalo de [0-30] segundos.

O algoritmo é dividido em duas etapas: (1) cópia do quadro inicial do travamento e (2) atribuição deste para os quadros subsequentes até a duração ser atingida.

Para este artefato, não foi necessário exibir os resultados obtidos pelo gerador, visto que o efeito da degradação é a repetição dos quadros, o que impossibilita sua representação em figuras.

### 3.8 AVALIAÇÃO DA QUALIDADE DE VÍDEO

Nesta etapa serão discutidos os algoritmos implementados para avaliar a qualidade de vídeo. As duas métricas implementadas foram: MSE e PSNR. As demais métricas foram calculadas utilizando o software IVQUEST (MURTHY; KARAM, 2011). O algoritmo do MSE é demonstrado na figura 39.

---

#### Algoritmo MSE

---

**Entrada:**  $imagemOriginal[largura][altura]$ ,  $imagemArtefato[largura][altura]$

**Saída:**  $mse$

```

1: Inicialize variável  $mse \leftarrow 0, temp \leftarrow 0$ 
2: para  $i = 0$  até  $largura$  faça
3:   para  $j = 0$  até  $altura$  faça
4:     {Comparar o valor da luminância do pixel original com o pixel artefato}
5:      $temp = imagemOriginal[i][j] - imagemArtefato[i][j]$ 
6:      $temp \leftarrow pow(temp, 2)$ 
7:      $mse \leftarrow mse + temp$ 
8:   fim para
9: fim para
10:  $mse \leftarrow mse / (largura + altura)$ 
11: retorne  $mse$ 

```

---

**Figura 39: Algoritmo para avaliar a qualidade de vídeo: MSE.**

Sabe-se que a métrica PSNR tem como parâmetro de entrada a nota dada pela métrica MSE. Neste trabalho, o cálculo do PSNR será feito sobre a nota final do MSE, ou seja, a média temporal do MSE será utilizada para calcular o PSNR do vídeo. O algoritmo do cálculo de PSNR é apresentado na figura 40. É importante ressaltar que, segundo a linha 2, quando o quadro original for igual ao quadro com artefatos este receberá o valor 100, como nota de PSNR, este artifício é usado apenas para facilitar a plotagem dos resultados nos gráficos.

---

#### Algoritmo PSNR

---

**Entrada:**  $mse$

**Saída:**  $psnr$

```

1: Inicialize variável  $psnr \leftarrow 0$ 
2: se  $mse = 0$  então
3:    $psnr \leftarrow 100$ 
4: senão
5:    $psnr \leftarrow 10 * log_{10}(pow(255, 2) / mse)$ 
6: fim se
7: retorne  $psnr$ 

```

---

**Figura 40: Algoritmo proposto para avaliar a qualidade de vídeo: PSNR.**

Tais métricas não representam, em alguns casos, a nota dada pelos avaliadores humanos. Para tentar contornar este problema a seguir é proposta uma modificação do PSNR, a fim de aumentar sua correlação com a avaliação subjetiva.

### 3.9 MÉTRICA PROPOSTA

A equação (80) apresenta a função proposta para avaliar a qualidade de vídeo. A métrica será chamada de *PSNR\_Mod* por se tratar de uma modificação da métrica tradicional. Para isto, nota-se que os objetos são avaliados individualmente e, a nota do PSNR convencional será ponderada pela relevância dos objetos. Esta modificação tem como objetivo evitar o problema evidenciado pela figura 20, cuja responsabilidade é atribuída pela avaliação de forma igual para todos os pixels. Espera-se que, com esta modificação, a métrica apresente uma maior correlação com a avaliação subjetiva.

$$PSNR\_Mod = \frac{\sum_{x=1}^{qtd\_obj} PSNR_x \times RelContext_x}{qtd\_obj}, \quad (80)$$

sendo  $RelContext_x$  a relevância contextual do objeto  $x$  na cena e  $qtd\_obj$  é a quantidade de objetos avaliados. O  $PSNR_x$  é calculado por

$$PSNR_x(t) = 10 \log_{10} \frac{L_{max}^2}{MSE_x(t)}, \quad (81)$$

onde  $L_{max}$  tem o valor de 255 (para amostragem de um pixel com 8 bits) e  $MSE_x(t)$  é obtido por

$$MSE_x(t) = [se \quad Seg(i, j, t) = pixelObjeto(i, j, t) \quad \text{então}] \frac{1}{area} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [Ref(x, y, t) - Art(x, y, t)]^2, \quad (82)$$

as variáveis  $Ref(x, y, t)$  e  $Art(x, y, t)$  representam, respectivamente, o valor do pixel do vídeo original e o valor do pixel no vídeo com artefatos no instante  $t$ .  $Seg$  é o valor do pixel na imagem segmentada nas coordenadas  $i, j$  do objeto  $x$ .  $pixelObjeto$  é o valor que representa o objeto, como por exemplo, 0 para fundo e 1 para objeto.

A figura 41 ilustra o algoritmo implementado para o cálculo do *PSNR\_Mod* para o instante  $t$ .

**Algoritmo** PSNR\_Mod**Entrada:**  $quadroSeg[largura][altura]$ ,  $quadroRef[largura][altura]$ ,  $quadroArt[largura][altura]$ **Saída:**  $psnrMod$ 

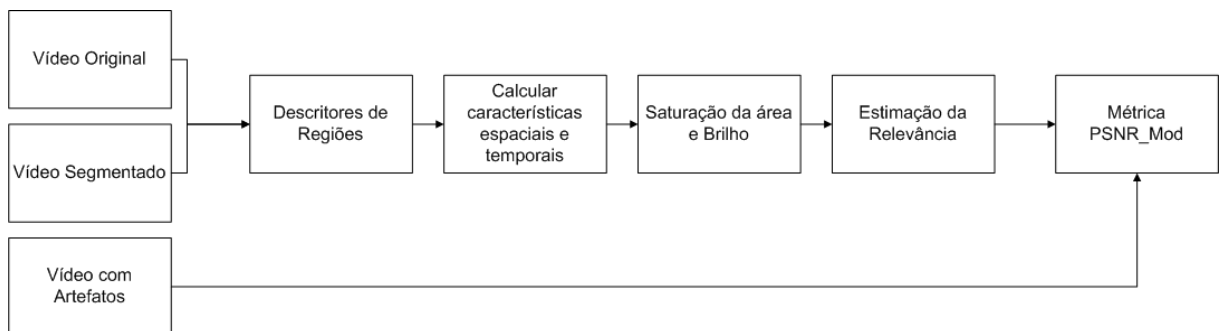
```

1: Ler  $pixelObjeto$ 
2: Calcular  $area$  utilizando  $quadroSegmentado$  e  $pixelObjeto$ 
3: Calcular  $RelContext$  do objeto representado pelo  $pixelObjeto$ 
4: Inicialize  $mseMod \leftarrow 0$ ,  $psnrMod \leftarrow 0$ 
5: para  $i = 0$  até  $largura$  faça
6:   para  $j = 0$  até  $altura$  faça
7:     se  $quadroSegmentado[largura][altura] = pixelObjeto$  então
8:        $mseMod \leftarrow pow((quadroRef[i][j] - quadroArt[i][j]), 2) + mseMod$ 
9:     fim se
10:  fim para
11: fim para
12:  $mseMod \leftarrow mseMod / area$ 
13:  $psnrMod \leftarrow (10 * log_{10}(pow(255, 2) / mseMod)) * RelContext$ 
14: retorne  $psnrMod$ 

```

**Figura 41: Algoritmo proposto para avaliação ponderada da qualidade de vídeo.**

O sistema de avaliação da qualidade de vídeo é ilustrado na figura 42. Os vídeos são as entradas para o algoritmo de avaliação, onde o vídeo original e o segmentado são utilizados para estimar a relevância dos objetos nos quadros. Os cálculos da relevância estão detalhados na seção 3.6. Em seguida, com base na estimativa da relevância e utilizando o vídeo com degradações, calcula-se a métrica objetiva da qualidade de segmentação.

**Figura 42: Diagrama do sistema de avaliação da qualidade de vídeo.**

## 4 RESULTADOS

Este capítulo descreve inicialmente os resultados obtidos pelas técnicas de segmentação que posteriormente foram avaliados visando medir sua qualidade. Após, realiza-se a avaliação do vídeo com artefato, comparando-se as métricas convencionais com métrica proposta por este trabalho.

O objetivo principal deste capítulo é apresentar e discutir os resultados de ambas as avaliações. Esta seção exibirá os resultados obtidos utilizando o vídeo “Akiyo”, que possui conteúdo estável e sua resolução atende os pré-requisitos do gerador de artefatos. Visando auxiliar o entendimento das próximas seções, é importante destacar que para a avaliação da segmentação foram gerados resultados para os dois tipos: avaliação independente e global. Para o caso da avaliação de vídeo, foram obtidos resultados com as métricas convencionais (PSNR, SSIM, VQM e VSNR) juntamente com o PSNR modificado.

### 4.1 SEGMENTAÇÃO

Nesta seção são demonstrados os resultados obtidos pelas técnicas de segmentação descritas na seção 3.4.

#### 4.1.1 SEGMENTAÇÃO UTILIZANDO ACO

Variando os parâmetros do ACO para a imagem utilizada como base (Células), nesta seção são demonstrados os resultados da segmentação usando a técnica ACO. Com a análise dos resultados, espera-se que seja possível encontrar a melhor configuração para a aplicação nas demais imagens testadas. Após, caso os resultados sejam satisfatórios, serão feitos testes com a melhor configuração sobre o vídeos original “Akiyo”.

Os experimentos avaliam a melhor, a pior e a média das soluções. Conforme as tabelas 1 e 2, é possível observar os resultados obtidos em termos da FO quando alterados os parâmetros centros e probabilidade de intensificação. Nota-se que, com a execução com centros distantes

a solução apresentou melhora quando comparado com o parâmetro de centros aleatórios. As melhores notas obtidas com centros distantes foram de 551,33 para  $q_0$  0,4 e 527,52 quando  $q_0$  0,8 em comparação com 553,47 para  $q_0$  0,4 e 541,46 para  $q_0$  0,8 obtidos pelos centros aleatórios.

<b>Probabilidade de intensificação</b>	<b>Melhor solução</b>	<b>Pior Solução</b>	<b>Média Soluções</b>
$q_0$ 0,4	551,33	560,41	554,75
$q_0$ 0,8	527,52	530,41	528,57

**Tabela 1: FO para o ACO puro com centros distantes.**

**Fonte: (PEREIRA et al., 2010a)**

<b>Probabilidade de intensificação</b>	<b>Melhor solução</b>	<b>Pior Solução</b>	<b>Média Soluções</b>
$q_0$ 0,4	553,47	558,02	555,61
$q_0$ 0,8	541,46	552,67	548,46

**Tabela 2: FO para o ACO puro com centros aleatórios.**

**Fonte: (PEREIRA et al., 2010a)**

Após segue-se para a execução da técnica proposta (ACO com busca local) combinado com a utilização dos centros distantes. Os resultados são exibidos na tabela 3

<b>Probabilidade de intensificação</b>	<b>Melhor solução</b>	<b>Pior Solução</b>	<b>Média Soluções</b>
$q_0$ 0,4	508,56	543,80	537,06
$q_0$ 0,8	500,63	517,10	511,14

**Tabela 3: FO para o ACO com Busca Local e centros distantes.**

**Fonte: (PEREIRA et al., 2010a)**

Comparando os dados obtidos pelos testes, conclui-se que a melhor configuração é dada quando aplicada a busca local proposta juntamente aos centros distantes. Destaca-se a melhora significativa na qualidade da solução, onde a melhor solução sem busca local obteve 527,52 comparados a 500,63 quando utilizados a técnica de busca local e uma probabilidade de intensificação ( $q_0$ ) de 0,8.

Também é importante salientar que o valor de  $q_0$  0,8 obteve melhores resultados em todos os casos em relação aos valores obtidos pelo  $q_0$  0,4. Ressalta-se que quanto maior for à

probabilidade de intensificação, menor será a probabilidade de este encontrar soluções diferentes para o problema do agrupamento, visto que, com valores altos o algoritmo tende a seguir as melhores soluções já encontradas. Logo, definiu-se como melhor configuração:

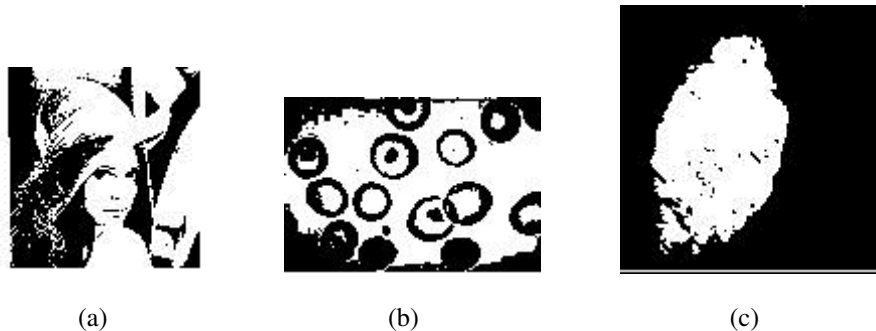
1. Uso dos centros distantes;
2. Utilização do fator  $q_0$  em 0,8;
3. Combinação entre ACO e Busca Local.

O próximo passo foi à aplicação desta configuração sobre as outras imagens testadas. A tabela 4 exibe os resultados obtidos pelas simulações e a figura 43 mostra os resultados obtidos pela técnica de segmentação proposta.

Imagem	Melhor solução	Pior Solução	Média Soluções
Lena	886,51	901,04	891,36
Mancha	787,99	788,17	788,11

**Tabela 4: FO para o ACO com Busca Local e centros distantes.**

Fonte: (PEREIRA et al., 2010a)



**Figura 43: Imagens segmentadas: (a) Lena (FO: 886,51), (b) Células (FO: 500,63), (c) Mancha (FO: 787,99).**

Fonte: (PEREIRA et al., 2010c)

É importante salientar que, como pode ser visto na figura 43, o método proposto para segmentação de imagens obteve bons resultados preliminares tanto na melhora das soluções (FO) quanto na imagem segmentada final. Porém, apesar de apresentar resultados promissores, a execução em vídeos foi abortada devido à demanda maior de recursos computacionais para o processamento dos quadros, quando comparada a outras técnicas propostas na literatura. Por

exemplo, 1 quadro levou aproximadamente 5 horas para ser segmentado, logo, como o trabalho avaliará os vídeos com 300 quadros, isto levaria 1500 horas (62,5 dias) para segmentar um vídeo com uma configuração. Ressalta-se que seriam necessários testes com diferentes configurações até a definição do melhor ajuste de parâmetros. Modificações desta técnica podem ser vistas em: (PEREIRA et al., 2010c) onde é definido um limiar automático utilizando o ACO para segmentar imagem através da técnica de Limiarização e em (PEREIRA et al., 2010b) a FO é modificada, ao invés da distância euclidiana, a técnica usa o índice Davies-Bouldin. Tal índice visa minimizar a similaridade média entre cada agrupamento como função da relação entre a soma da dispersão intra-cluster e a distância entre os centros dos clusters.

#### 4.1.2 SEGMENTAÇÃO POR MDRAV

Analisando de forma subjetiva os resultados obtidos pela técnica (figura 32), percebe-se que a segmentação atingiu um bom resultado. A divisão entre fundo e primeiro plano ficou evidente, como é ilustrado na figura 44. Neste vídeo foram detectados dois objetos de interesse, o fundo (preto) e a apresentadora (branco).

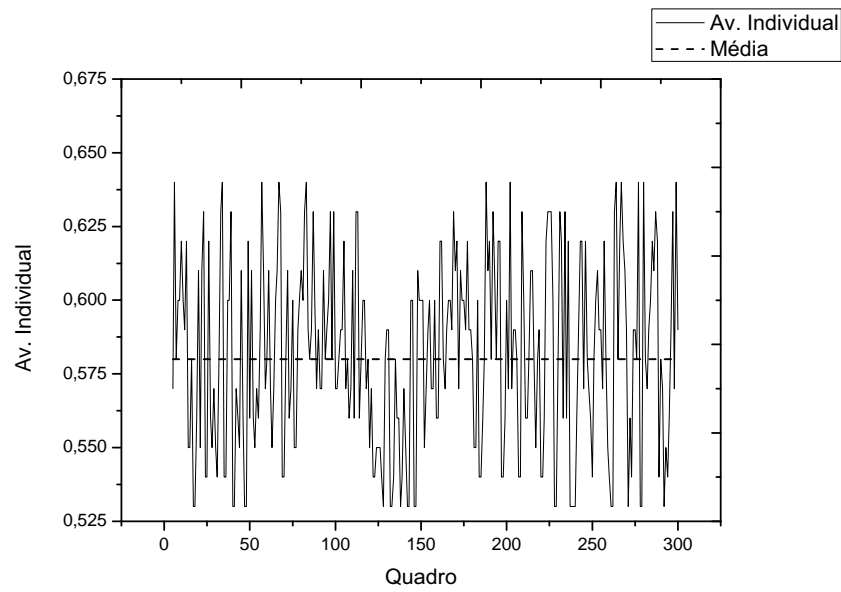


**Figura 44: Separação dos objetos na cena: (a) Quadro original, (b) Objetos separados.**

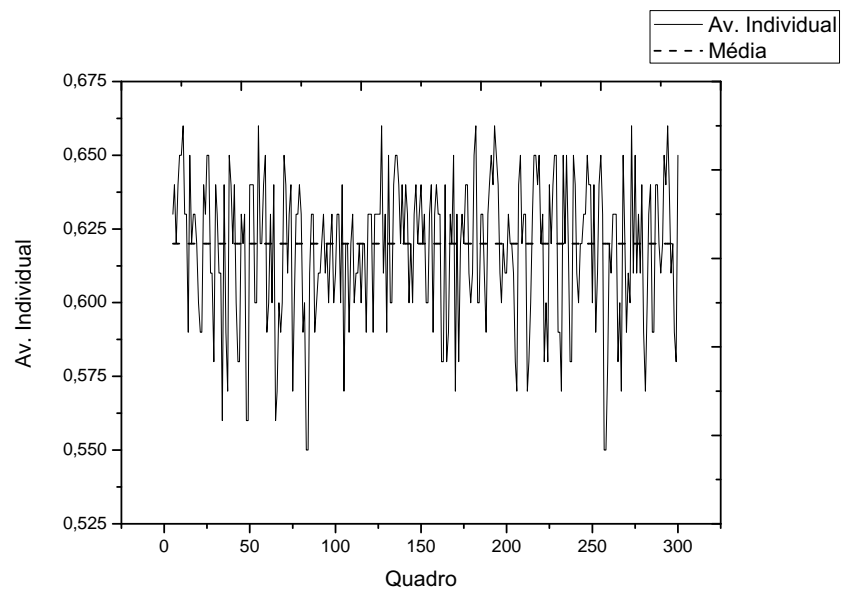
#### 4.2 AVALIAÇÃO DA QUALIDADE DE SEGMENTAÇÃO

Nesta seção serão avaliados, através de uma métrica objetiva, os resultados obtidos pela técnica MDRAV. A avaliação objetiva independente da qualidade de segmentação dos objetos é calculada pela equação (54), utilizando como parâmetros o vídeo original “Akiyo” e o vídeo segmentado. Os resultados, para todos os instantes, dessa avaliação são exibidos na figura 45.





(a)



(b)

**Figura 45: Avaliação Independente: (a) Objeto 0 e (b) Objeto 1.**

A tabela 5 exibe os resultados médios obtidos por esta segmentação e pela pior segmentação avaliada pelo trabalho de (CORREIA; PEREIRA, 2002). Esta tabela só é representativa e aqui não estamos comparando os resultados, pois isso seria impossível pelo fato deles avaliarem o vídeo “Akiyo” com formato QCIF (176 x 144) no total de 30 quadros, diferentemente da utilizada por este trabalho que possui 300 quadros na resolução 352 x 288 (CIF).

Segmentação	(BEZERRA, 2011)	(CORREIA; PEREIRA, 2002)
Objeto 0	0,58	0,65
Objeto 1	0,62	0,56

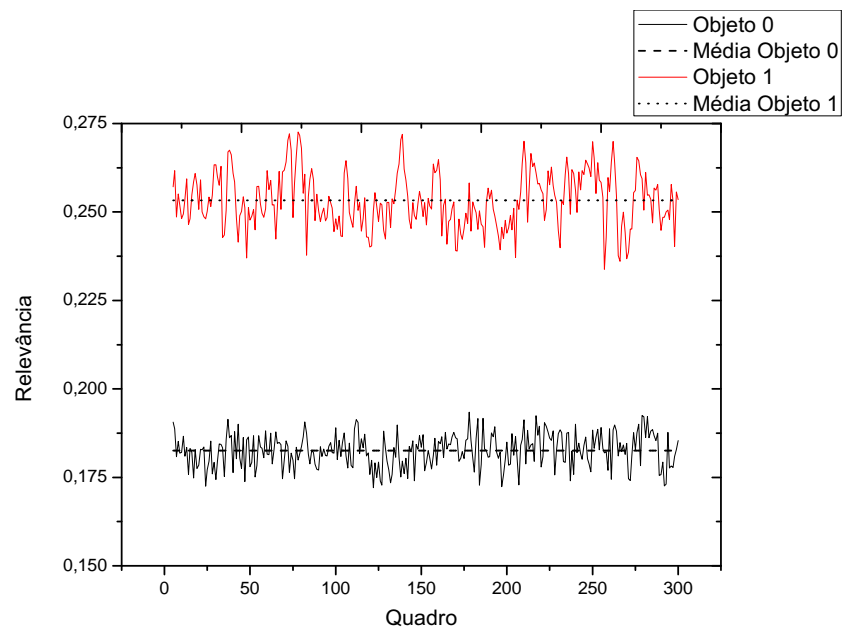
**Tabela 5: Avaliação Individual da Qualidade de Segmentação.**

Referenciando-se as curvas exibidas na figura 45 verifica-se que os resultados obtidos pela técnica de segmentação com o objeto 0 obteve média de 0,58, enquanto a nota obtida com o objeto 1 teve média de 0,62. A principal observação refere-se à característica inter-quadro, obtida através da equação 53, que avalia o contraste. A técnica de segmentação obteve notas baixas neste quesito, principalmente pelo fato de regiões do fundo serem adicionadas ao longo do objeto 1. A adição destas áreas pode ser observada reparando-se na borda do objeto 1 segmentado, como ilustrado na figura 46.



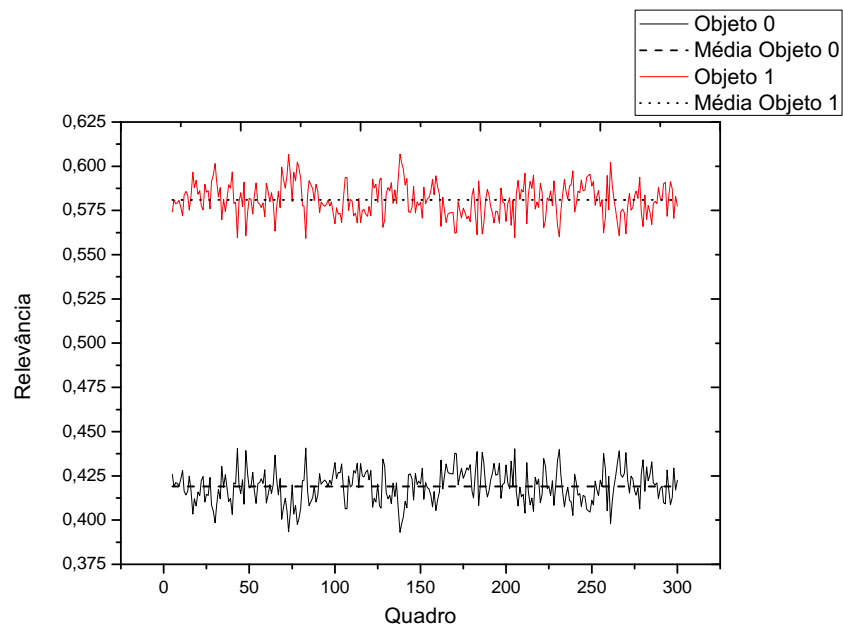
**Figura 46: Representação da adição de áreas do fundo no objeto 1.**

Em seguida foi realizada a avaliação global do vídeo segmentado. A avaliação global é baseada na avaliação individual pesando cada objeto através da sua relevância no quadro. A relevância individual dos objetos no quadro, calculada pela equação (69), têm seus resultados exibidos na figura 47.



**Figura 47: Relevância individual dos objetos.**

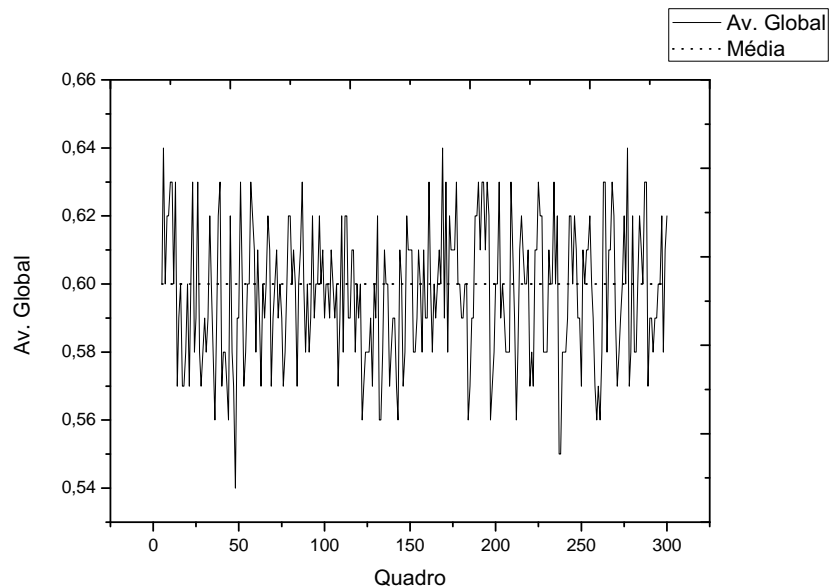
Após avaliar individualmente a relevância dos objetos na cena, procedeu-se à avaliação da relevância contextual. A figura 48 exibe os resultados da relevância contextual para os dois objetos de interesse.



**Figura 48: Relevância contextual dos objetos.**

Como esperado, o objeto 1 obteve notas maiores quando comparados com o objeto 0, pelo fato do observador humano prestar mais atenção na face humana quando comparada com um fundo estático. Os fatores que mais pesaram no resultado foram a informação espacial (*SI*)

e a criticidade. Com o resultado da relevância contextual, a avaliação global é calculada pela equação (59) e seu resultado apresentado na figura 49. A média obtida pela avaliação global foi de 0,60, em contraste com a avaliação individual onde a média do objeto 0 foi 0,58 e do objeto 1 foi 0,62. Nota-se que, para este caso, se a simulação fizesse uma média aritmética com os valores obtidos pela avaliação individual se obteria a mesma nota na avaliação global. Este fato pode ocorrer devido à utilização de algumas características tanto para avaliar a segmentação quanto para medir a relevância dos objetos no quadro.



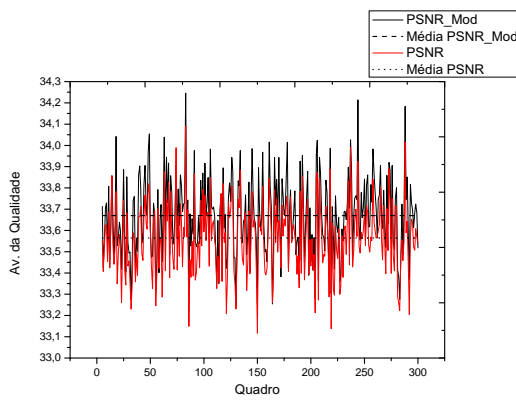
**Figura 49: Avaliação Global da Segmentação.**

### 4.3 AVALIAÇÃO DA QUALIDADE DE VÍDEO

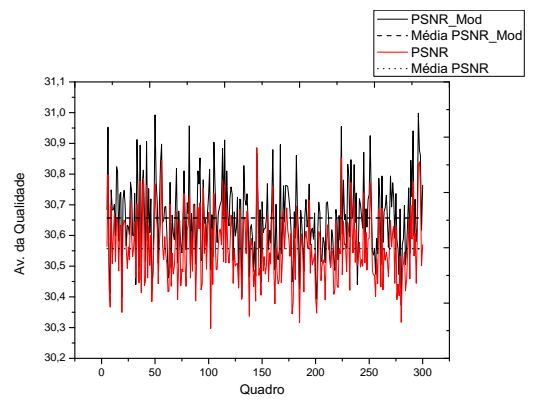
Esta seção apresenta os resultados obtidos pela avaliação da qualidade de vídeo afetado por degradações. Os vídeos degradados são avaliados com: PSNR, SSIM, VSNR, VQM e PSNR modificado.

#### 4.3.1 RUÍDO BRANCO GAUSSIANO

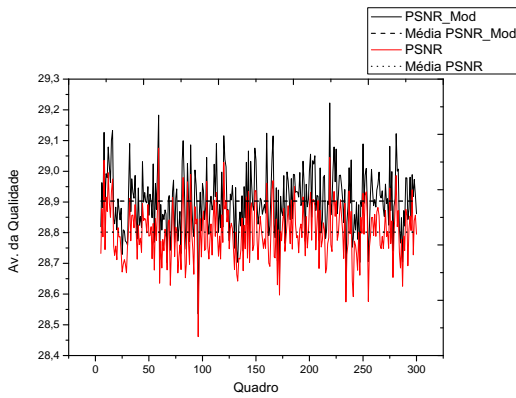
Para o ruído branco gaussiano foram avaliados 10 vídeos com degradações nos seguintes perceptuais: 2%, 4%, 6%, 8%, 10%, 15%, 20%, 30%, 50%, 100%. Tais valores representam a porcentagem de taxa de ocorrência em um quadro do vídeo. As figuras 50 e 51 exibem os resultados obtidos pelo PSNR e PSNR modificado.



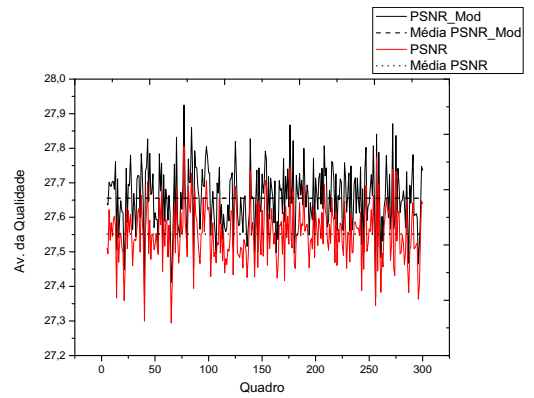
(a)



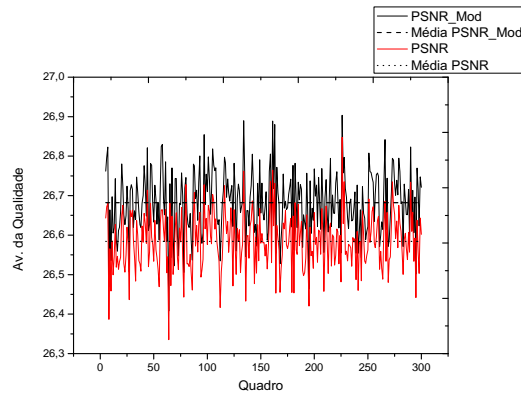
(b)



(c)

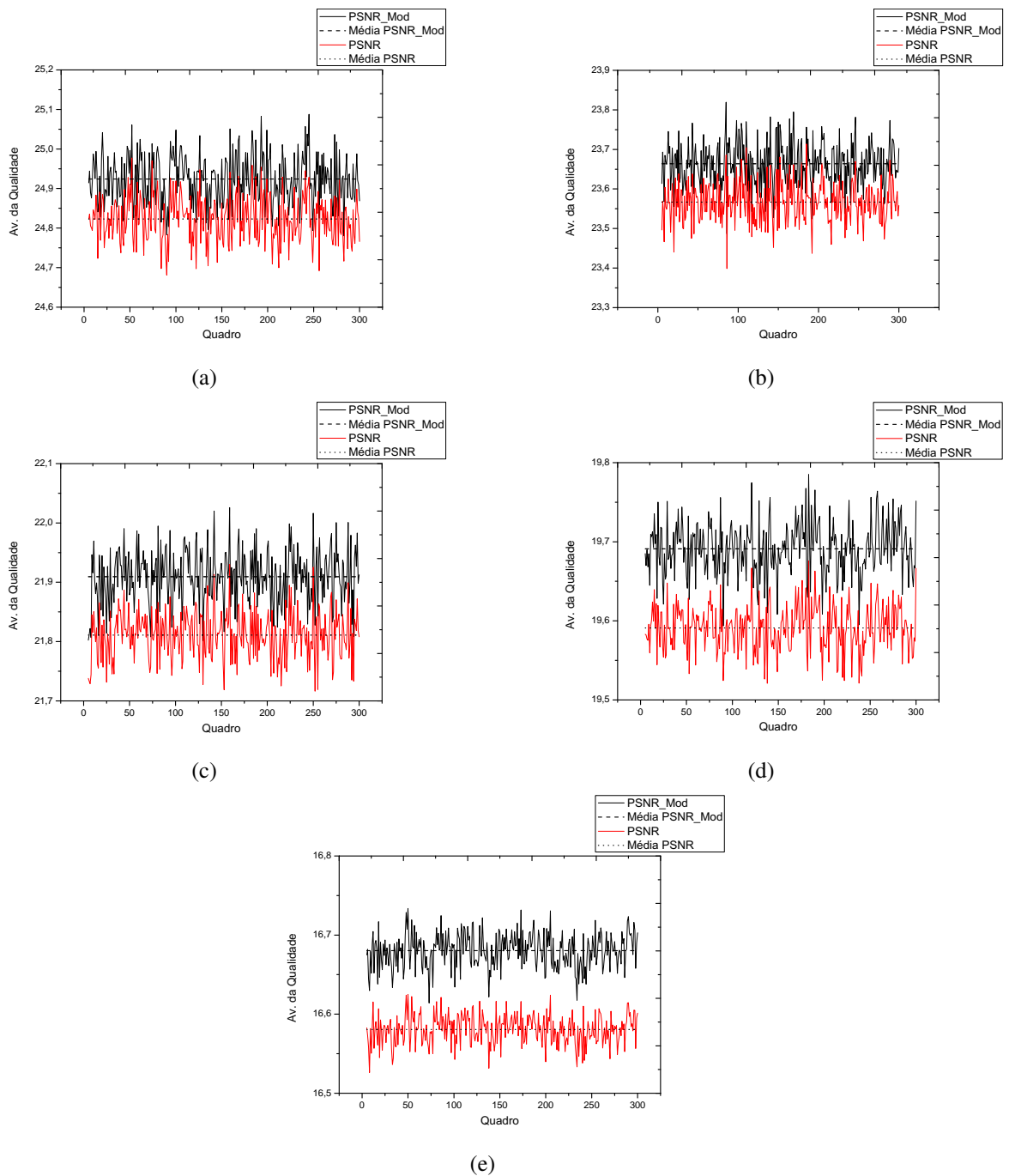


(d)



(e)

**Figura 50: Degradação com ruído com ruído branco gaussiano: (a) 2% de degradação, (b) 4% de degradação, (c) 6% de degradação, (d) 8% de degradação, (e) 10% de degradação.**



**Figura 51: Degradação com ruído branco gaussiano: (a) 15% de degradação, (b) 20% de degradação, (c) 30% de degradação, (d) 50% de degradação, (e) 100% de degradação.**

Como as métricas não estão na mesma escala é necessário analisar a tabela 6 para tirar conclusões dos resultados. Todas as métricas foram sensíveis ao aumento do nível de degradação, ou seja, quanto maior o nível pior a nota atribuída pela métrica. Em relação à métrica proposta, esta apresentou uma variação maior quando comparada ao PSNR (ver figuras 50 e 51). No entanto as médias foram próximas, em torno de 0,13 a menos em relação à métrica tradicional.

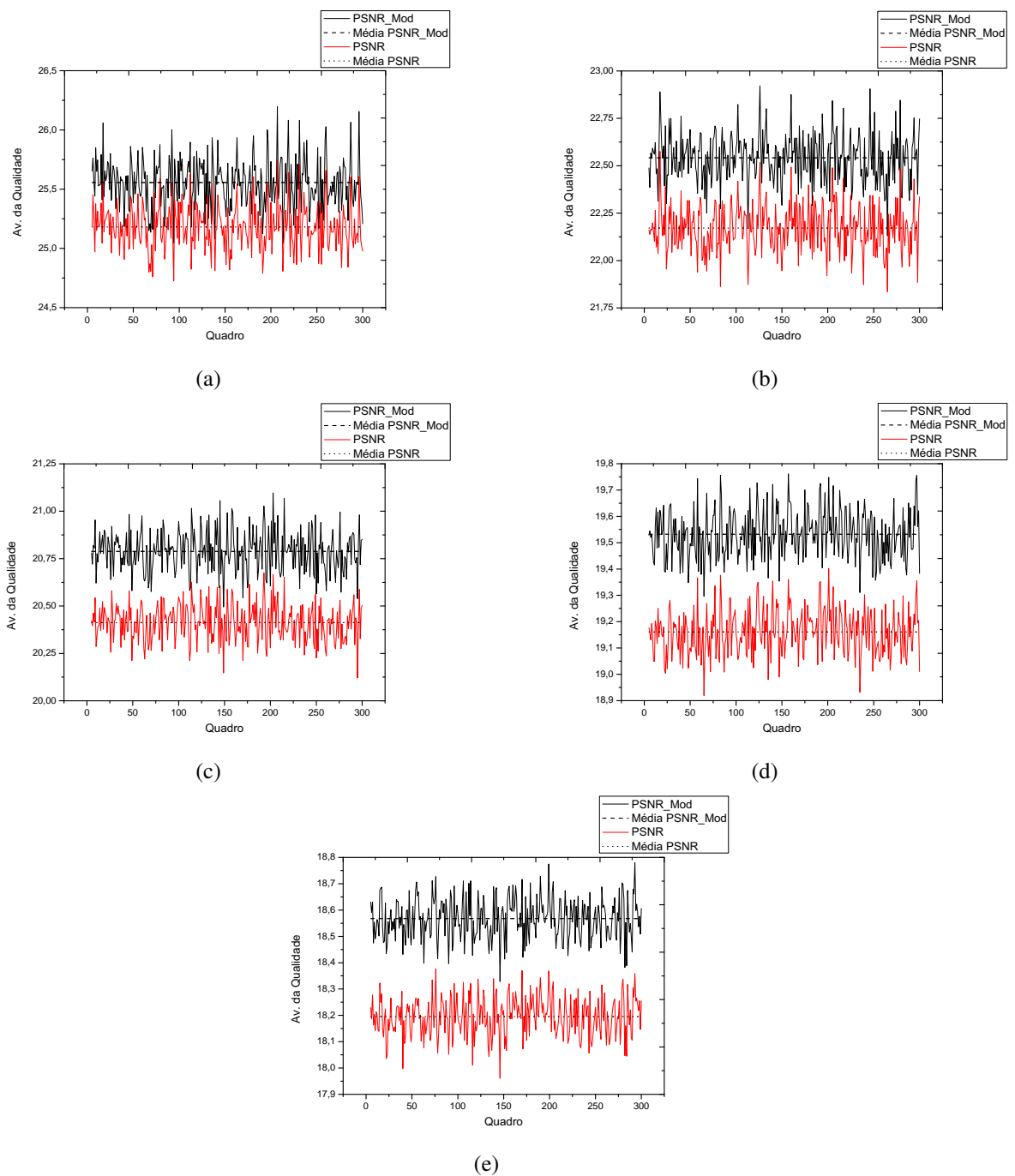
Artefato	Métricas Objetivas				
	PSNR	SSIM	VQM	VSNR	PSNR_Mod
RB2	33,56	0,86	0,28	32,27	33,67
RB4	30,53	0,75	0,54	28,60	30,65
RB6	28,78	0,67	0,76	26,50	28,90
RB8	27,54	0,60	0,94	25,06	27,65
RB10	26,56	0,54	1,03	23,95	26,68
RB15	24,79	0,44	1,11	21,97	24,92
RB20	23,56	0,37	1,16	20,57	23,66
RB30	21,79	0,28	1,22	18,66	21,90
RB50	19,57	0,20	1,28	16,29	19,69
RB100	16,56	0,12	1,35	13,19	16,68

**Tabela 6: Resultado da avaliação para o artefato ruído branco gaussiano.**

Segundo (SILVA, 2009), o ruído branco gaussiano em pequenas proporções é aceitável, entretanto, conforme a sua incidência aumenta, maior é a irritabilidade percebida pelo observador humano. Este fato pode ser confirmado analisando os resultados obtidos onde os vídeos com menor incidência deste artefato apresentaram notas maiores quando comparadas com a maior incidência.

#### 4.3.2 ARTEFATO SAL-E-PIMENTA

Utilizando-se o artefato sal-e-pimenta, foram feitos testes com 5 diferentes níveis de degradação: 1%, 2%, 3%, 4% e 5%. A porcentagem representa a taxa de ocorrência deste artefato em cada quadro do vídeo. A figura 52 ilustra os resultados obtidos pelo PSNR e PSNR modificado, que estão na mesma escala de valores.



**Figura 52: Degradação com o artefato sal-e-pimenta: (a) 1% de degradação, (b) 2% de degradação, (c) 3% de degradação, (d) 4% de degradação, (e) 5% de degradação.**

A tabela 7 destaca as diferenças obtidas pelas métricas avaliadas. Observa-se que para todos os casos a avaliação modificada apresentou notas superiores quando comparada ao PSNR. Analisando as curvas obtidas pela avaliação, ilustradas na figura 52, verifica-se que a diferença média entre os valores foi de 0,39 a mais para a métrica proposta.

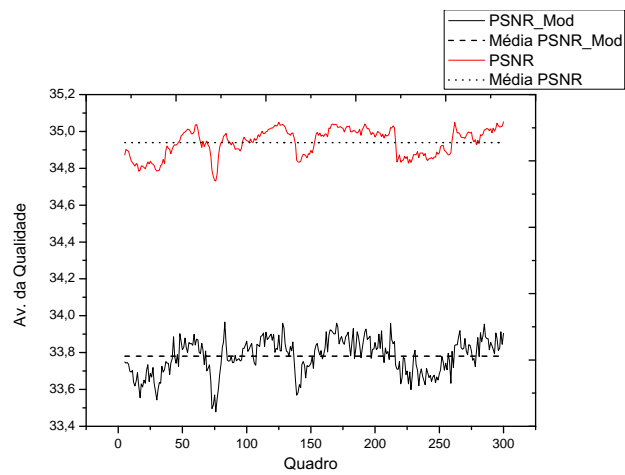


Artefato	Métricas Objetivas				
	PSNR	SSIM	VQM	VSNR	PSNR_Mod
SEP1	25,16	0,73	1,10	16,72	25,55
SEP2	22,15	0,55	1,23	13,63	22,54
SEP3	20,39	0,43	1,30	11,84	20,78
SEP4	19,14	0,34	1,34	10,59	19,53
SEP5	18,17	0,28	1,36	9,62	18,56

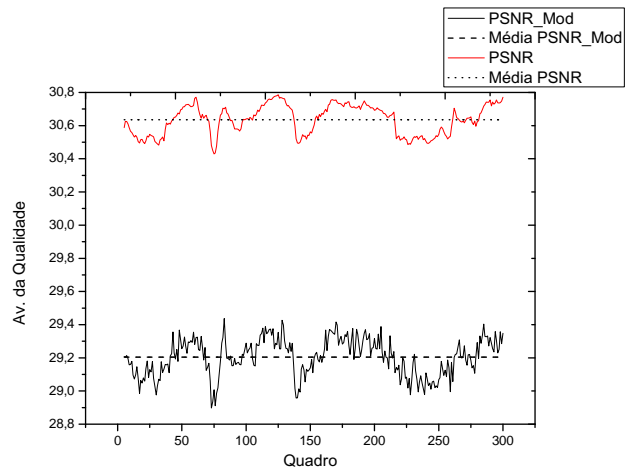
**Tabela 7: Resultado da avaliação para o artefato sal-e-pimenta.**

#### 4.3.3 BORRAMENTO

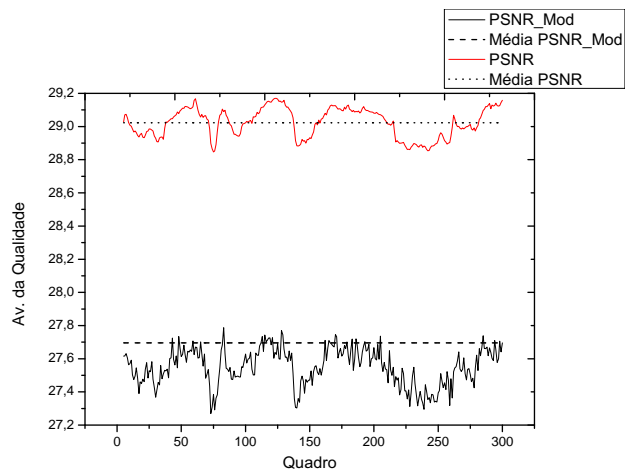
Os resultados obtidos pela avaliação do vídeo degradado com o borramento, utilizando máscaras de tamanho 3 pixels (*Bor3*), 5 pixels (*Bor5*), 7 pixels (*Bor7*), 3x3 pixels (*Bor3x3*), 5x5 pixels (*Bor5x5*) e 7x7 pixels (*Bor7x7*) para as avaliações quadro a quadro para o PSNR e PSNR modificado são exibidos nas figuras 53 e 54. Para todos os casos o PSNR convencional obteve notas maiores do que a métrica proposta. A tabela 8 ilustra as notas médias obtidas pelas avaliações para todas as configurações testadas.



(a)

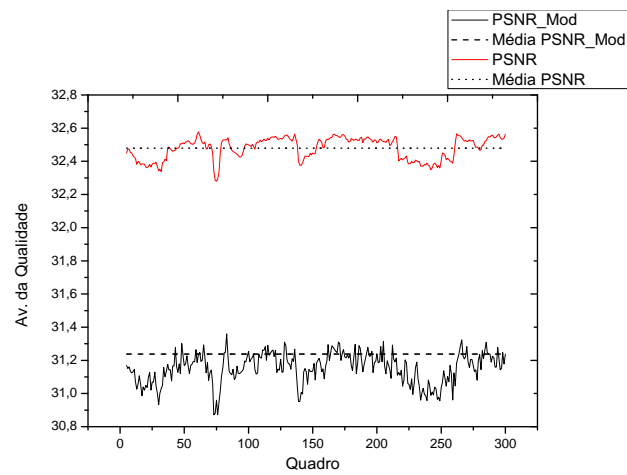


(b)

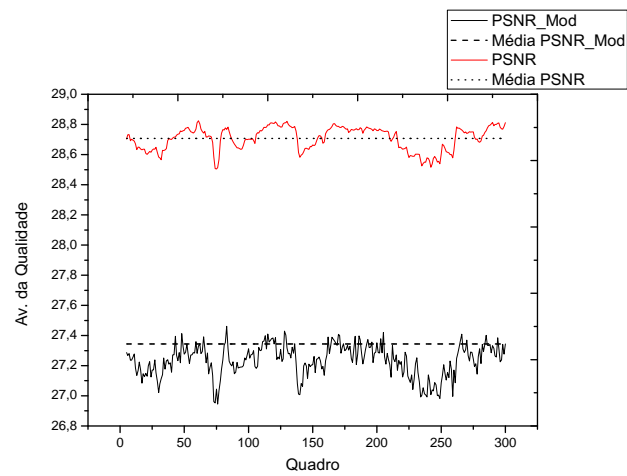


(c)

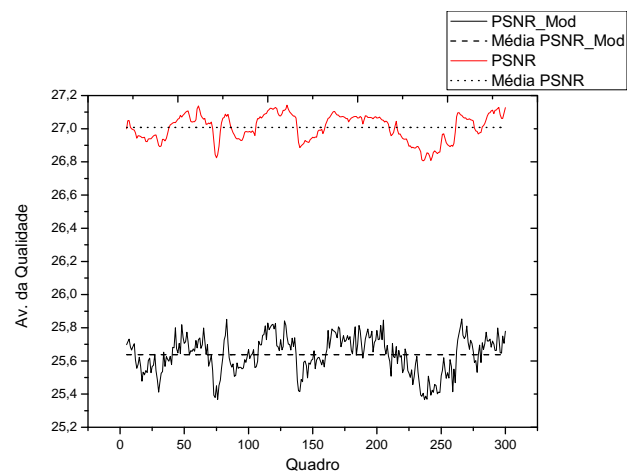
**Figura 53: Degradação com emprego do artefato borramento linear: (a) com máscara linear 3, (b) com máscara linear 5 e (c) com máscara linear 7.**



(a)



(b)



(c)

**Figura 54: Degradação com emprego do artefato borramento linear: (a) com máscara 3x3, (b) com máscara 5x5 e (c) com máscara 7x7.**

Artefato	Métricas Objetivas				
	PSNR	SSIM	VQM	VSNR	PSNR_Mod
Bor3	34,92	0,97	0,09	32,37	33,78
Bor5	30,61	0,92	0,34	23,61	29,20
Bor7	29,00	0,88	0,57	20,03	27,55
Bor3x3	32,46	0,96	0,14	29,86	31,15
Bor5x5	28,69	0,89	0,43	21,40	27,23
Bor7x7	26,99	0,83	0,75	16,94	25,63

**Tabela 8: Resultado da avaliação para o artefato borrramento com máscara linear.**

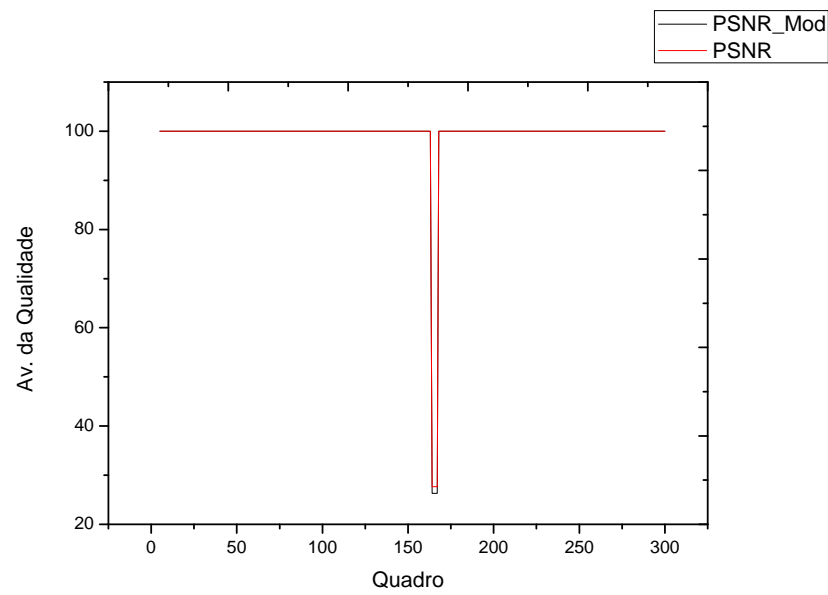
#### 4.3.4 BLOCAGEM

Os resultados obtidos pelas técnicas de avaliação para o artefato blocagem são ilustrados na tabela 9. Foram feitos testes com 2 diferentes níveis de degradação: 1% e 5%. A porcentagem representa a taxa de probabilidade de um pixel ser transformado em um bloco.

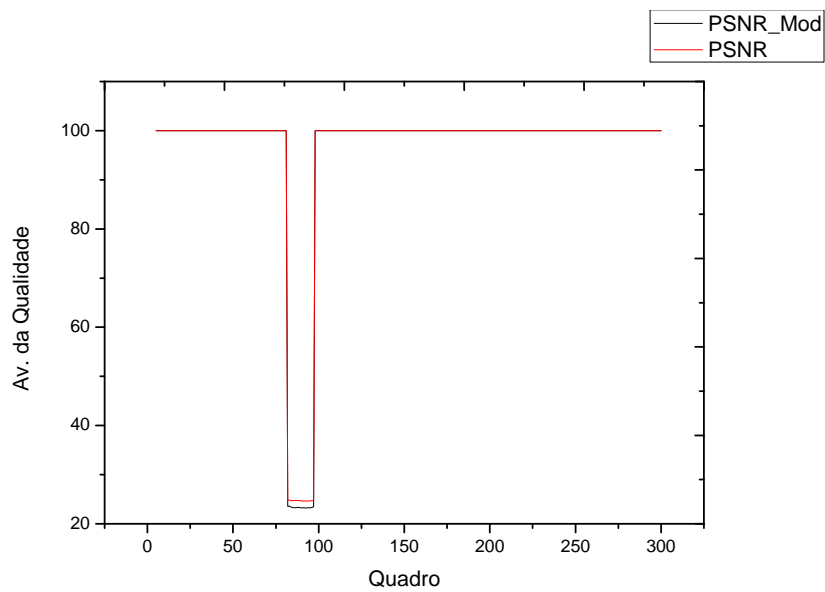
Artefato	Métricas Objetivas				
	PSNR	SSIM	VQM	VSNR	PSNR_Mod
Bloc1	41,06	0,99	0,002	73,71	48,45
Bloc5	40,89	0,98	0,01	71,19	39,36

**Tabela 9: Resultado da avaliação para o artefato blocagem**

A figura 55 ilustra os testes do PSNR e PSNR modificado para todos os instantes avaliados. É possível observar que, com 1% de probabilidade, o artefato apareceu entre os quadros 150 e 175, e com 5% este efeito esteve presente entre os quadros 75 e 100. Destaca-se que para ambos os testes, o PSNR modificado retornou notas superiores às atribuídas pelo PSNR. Também fica visível, a variação dos algoritmos quando a degradação aumenta, ou seja, todos eles se comportaram de forma esperada.



(a)



(b)

**Figura 55: Degradação com emprego do artefato blocagem: (a) com 1%, (b) com 5%.**

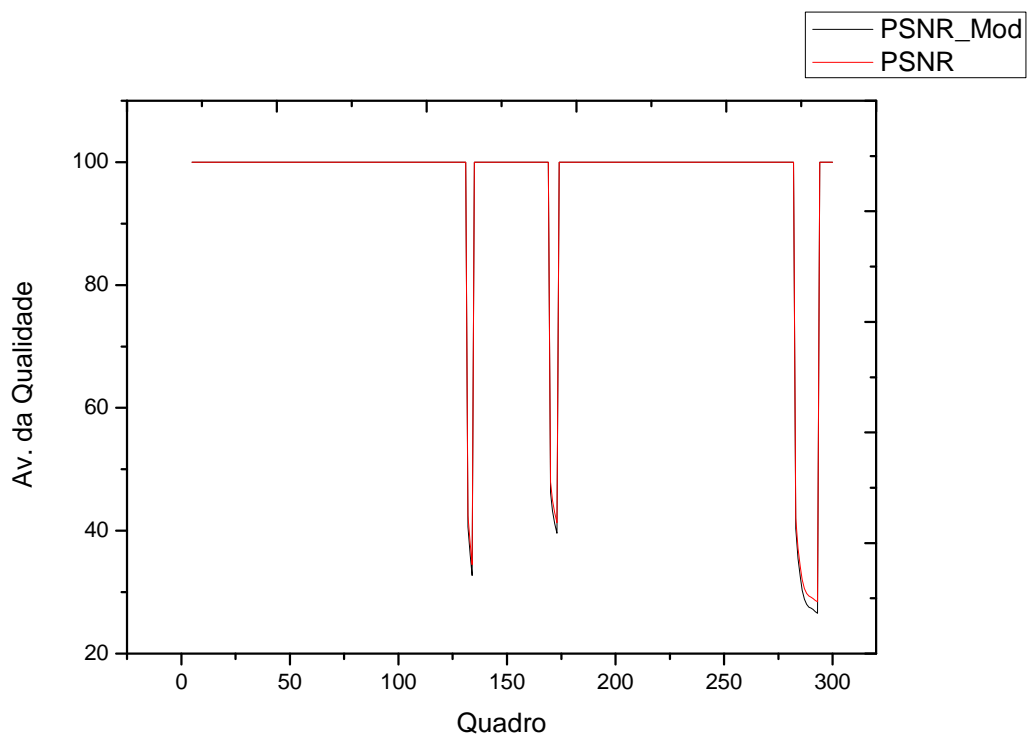
#### 4.3.5 TRAVAMENTO

Os resultados obtidos para o artefato travamento pelas métricas objetivas são ilustrados na tabela 10.

Artefato	Métricas Objetivas				
	PSNR	SSIM	VQM	VSNR	PSNR_Mod
Trav	44,48	0,99	0,01	71,74	46,05

**Tabela 10: Resultado da avaliação para o artefato travamento**

A figura 56 ilustra os testes executados quadro a quadro pelo PSNR e PSNR modificado. Neste caso, o vídeo apresentou o travamento entre os quadros 125-150, 150-200 e 275-300.



(a)

**Figura 56: Degradação com emprego do artefato travamento.**

#### 4.4 DISCUSSÃO DOS RESULTADOS

Nesta seção algumas discussões serão feitas por se encaixarem em todas as degradações testadas.

Primeiramente, o artefato borramento foi o que apresentou uma maior variação nas curvas e onde o PSNR modificado atribuiu notas inferiores aos vídeos quando comparado ao PSNR, em torno de 1,35 a mais. A métrica proposta demonstrou uma maior sensibilidade na detecção deste artefato, enquanto o PSNR apresentou um comportamento mais suave. Para os

outros artefatos avaliados foram obtidas notas superiores às atribuídas pela métrica convencional (PSNR) inferindo que, possivelmente, tais artefatos estiveram mais presentes na região do fundo do que no primeiro plano (apresentadora).

Tais afirmações são feitas baseadas nos resultados obtidos pelos cálculos da relevância, onde a apresentadora obteve notas maiores quando comparados com o fundo. Logo, o primeiro plano obteve um maior peso no cálculo do PSNR modificado. Além disso, por se tratarem de artefatos que ocorrem de forma probabilística, ambas as avaliações apresentaram um comportamento de grande variação quadro a quadro.

Outro fato interessante é ilustrado nas gerações das curvas, onde mesmo aumentando o grau de degradação o comportamento temporal era próximo aos outros com o mesmo artefato. Segundo (ALBINI, 2009), isto acontece devido ao fato que em todos os casos foram feitas as mesmas interferências entre os pixels e também pelo fato do vídeo original ser o mesmo.

Para esclarecer as diferenças obtidas entre as técnicas é necessário validar a métrica proposta, pois as avaliações não estão na mesma escala, dificultando a comparação.

#### 4.5 VALIDAÇÃO DA MÉTRICA PROPOSTA

Após a aplicação da métrica objetiva deseja-se comparar seus resultados com os obtidos pela avaliação subjetiva. Em (SILVA, 2009) é feita a avaliação subjetiva dos vídeos obtidos pelo gerador de artefatos proposto por (ALBINI, 2009). O autor (SILVA, 2009), forneceu os valores obtidos pela avaliação subjetiva, para que as curvas fossem novamente plotadas com a adição da variável PSNR modificado, visando facilitar a visualização e comparação entre as métricas. A avaliação subjetiva utilizada foi a SDSCE (*Single Double Stimulus for Continuous Evaluation*), onde há um duplo estímulo simultâneo para avaliação contínua (ITU-R, 2002). Nesse processo é exibido em um monitor o vídeo de referência (sem artefatos) posicionado à esquerda e o vídeo contendo artefatos à direita. Após a execução de ambos os vídeos simultaneamente, os avaliadores atribuem notas em uma escala de 1 a 5, onde o 1 representava uma imagem excelente e 5 uma qualidade ruim.

Assim sendo, a validação da métrica foi baseada nos resultados obtidos sobre os diferentes artefatos avaliados, comparando-os com as métricas propostas na literatura. O teste de desempenho foi realizado utilizando duas métricas: CCS (Coeficiente de Correlação de Spearman) e CCP (Coeficiente de Correlação de Pearson). CCS avalia a unimodalidade e monotonicidade entre as avaliações objetivas e as avaliações subjetivas. Tal coeficiente baseia-se no *rank* dos valores medidos e é comumente utilizado na análise de dados não paramétricos e de

correlação. O CCP estima a exatidão da previsão através da análise dos dados que possuem uma distribuição normal (TRIOLA, 2005). Para calcular o CCS é utilizada a equação (83).

$$r = 1 - \frac{6\sum_i d_i^2}{(n^3 - n)}, \quad (83)$$

onde  $n$  é o número de pares  $(x_i, y_i)$ ,  $d_i$  é a posição de  $x_i$  menos a posição de  $y_i$ . Caso as posições de  $x$  e  $y$  sejam as mesmas, então  $d_i$  será igual a zero e  $r$  será igual a 1.

O cálculo do CCP é exibido na equação (84).

$$r = \frac{C_{x,y}}{S_x S_y}, \quad r \in [-1, 1] \quad (84)$$

onde  $x$  é a nota obtida pela avaliação objetiva,  $y$  é o valor atribuído pelo método subjetivo,  $C_{x,y}$  é a variância conjunta das variáveis  $x$  e  $y$ ,  $S_x$  e  $S_y$  são os desvios padrões das variáveis  $x$  e  $y$ , respectivamente.

Salienta-se que, antes de calcular as métricas, é necessário realizar a regressão não linear para mapear os valores das avaliações objetivas com o MOS. Neste trabalho é utilizada a função logística para calcular a regressão. Tal função está representada na equação (85) (SESHADRINATHAN et al., 2010), (VQEG, 2000).

$$Q'_j = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-\left(\frac{Q_j - \beta_3}{|\beta_4|}\right)}}, \quad (85)$$

onde  $Q_j$  é a nota obtida pela métrica objetiva  $j$  é o vídeo testado, as variáveis  $\beta$  representam os quatro parâmetros utilizados para ajustar a predição da avaliação da qualidade com as notas medidas de maneira subjetiva (VQEG, 2000). Os autores (SESHADRINATHAN et al., 2010), utilizam a função “nlinfit” para encontrar os parâmetros ótimos  $\beta$ , com o objetivo de minimizar os erros quadráticos entre as notas subjetivas e os valores alcançados pelas avaliações objetivas.

Por fim, o CCS e o CCP são calculados utilizando os valores obtidos pela predição da métrica ( $Q'_j$ ) com as notas subjetivas ( $MOS_j$ ). Os resultados obtidos são exibidos na tabela 11.

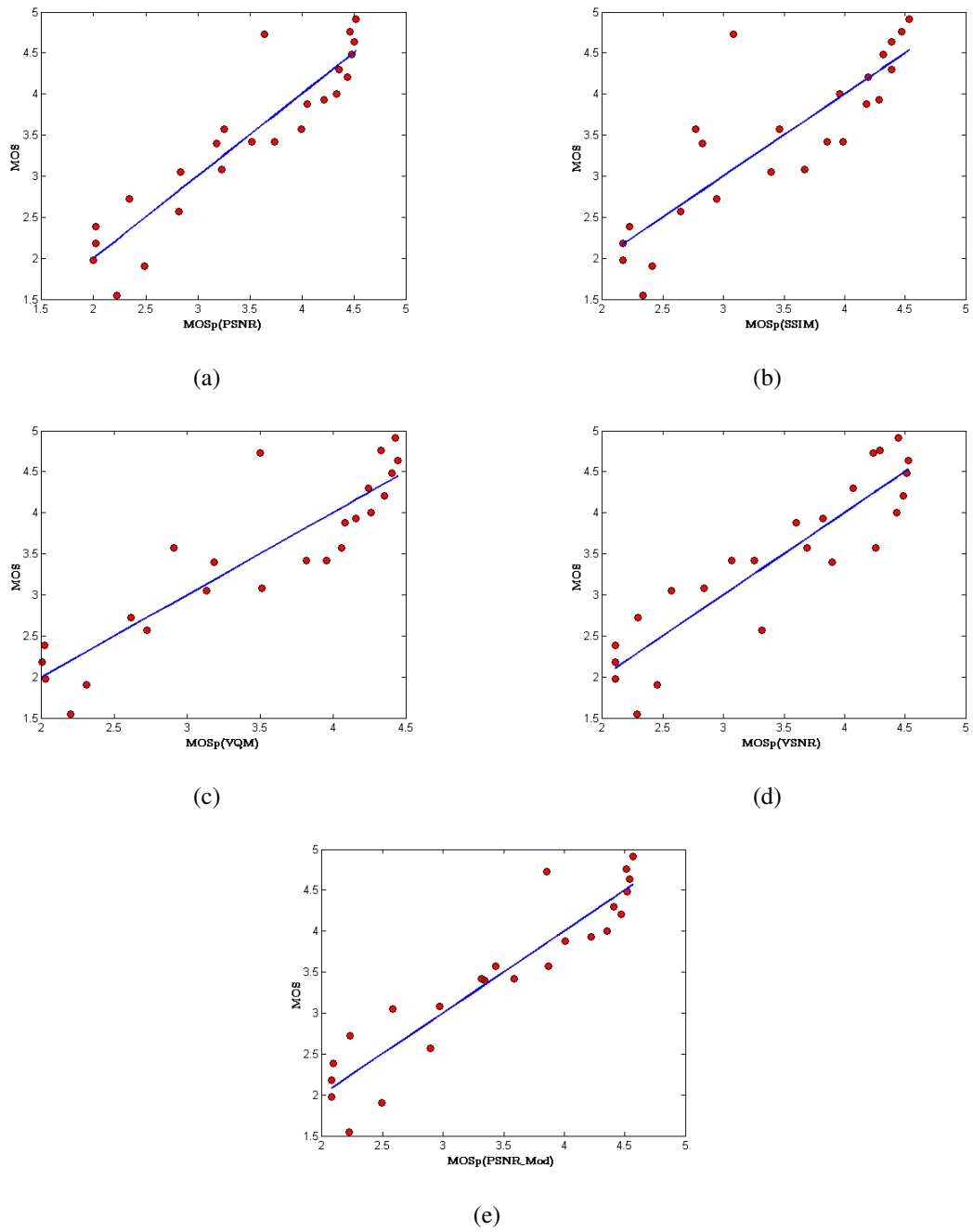


Métrica	Coeficiente de Correlação	
	Spearman (CCS)	Pearson (CCP)
PSNR	0,9289	0,9219
SSIM	0,8663	0,8509
VQM	0,8885	0,8988
VSNR	0,8937	0,9078
PSNR_Mod	<b>0,9372</b>	<b>0,9328</b>

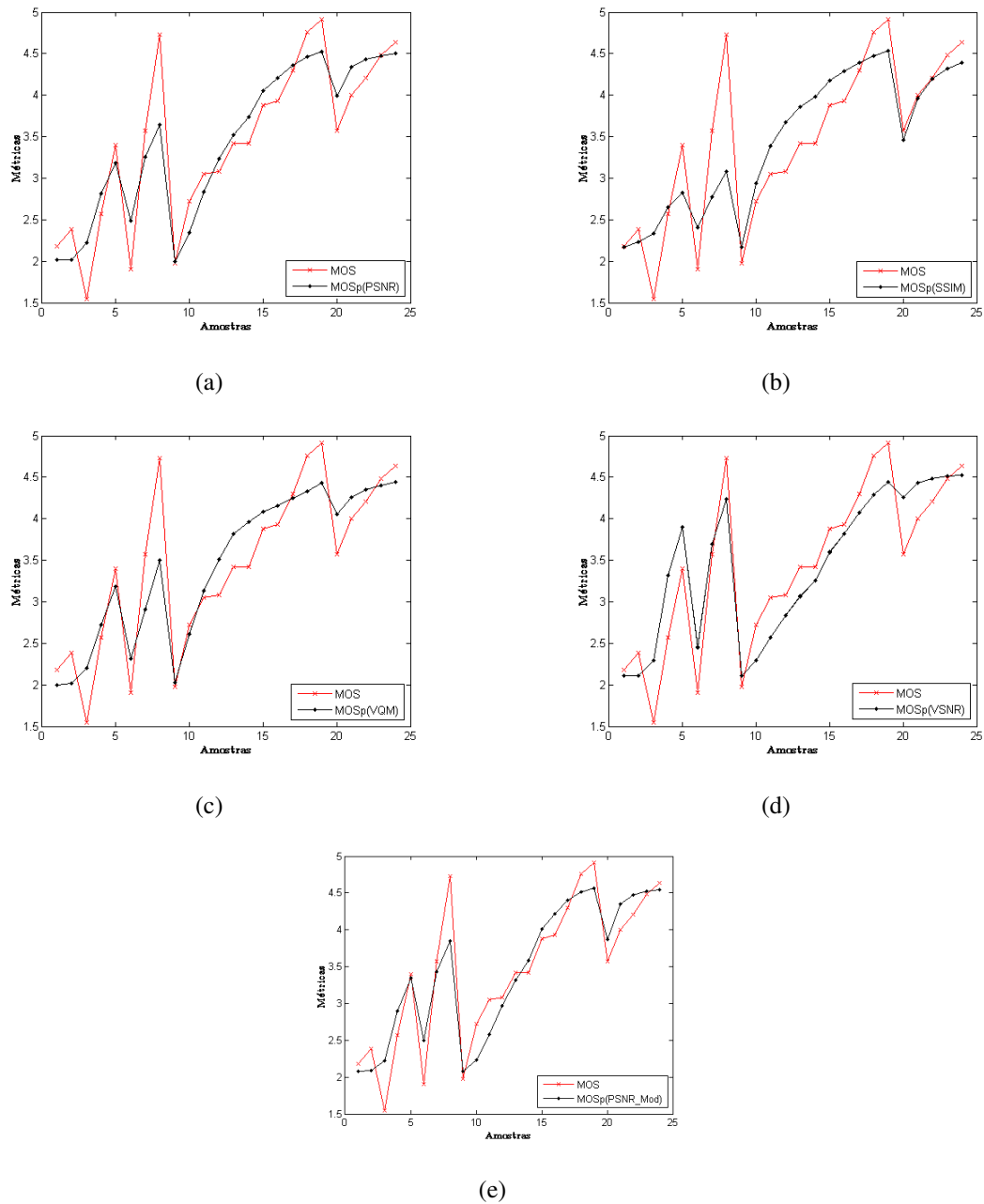
**Tabela 11: Comparação do desempenho das métricas objetivas.**

Na tabela 11, os valores em negrito representam a maior correlação obtida para cada coeficiente. Analisando os resultados, os valores obtidos indicam uma forte correlação, ou seja, existe uma dependência entre a nota dada pelos observadores quando comparadas com a nota dada de forma objetiva. É possível notar que, em ambos os casos, a métrica proposta apresentou a maior correlação com a avaliação subjetiva.

Com os resultados obtidos pelas predições, foram geradas curvas exibindo a relação entre as notas objetivas com a medida subjetiva para todos os algoritmos testados. Também é representada a apropriada curva de tendência da função logística. Neste trabalho, fica evidente a significativa correlação do PSNR modificado quando comparado com os outros algoritmos.



**Figura 57: Correlação entre Av. Objetiva x MOS: (a) PSNR, (b) SSIM, (c) VQM, (d) VSNR, (e) PSNR\_Mod.**



**Figura 58: Relação entre MOS e a Predição da Avaliação Objetiva: (a) PSNR, (b) SSIM, (c) VQM, (d) VSNR, (e) PSNR\_Mod.**

Por fim, é importante destacar que este nível de correlação alta normalmente não acontece entre as métricas, visto que as métricas objetivas muitas vezes não representam a nota dada de forma subjetiva (SESHADRINATHAN et al., 2010). Os vídeos empregados, particularmente aqueles com alta variação de conteúdo podem influenciar fortemente tal resultado. Observa-se que o vídeo utilizado neste trabalho possui o conteúdo estável.

## 5 CONCLUSÃO

A avaliação objetiva da qualidade de vídeo não representa, em alguns casos, uma boa indicação da qualidade observada visto que, geram valores discrepantes quando comparados com as avaliações feitas pelos observadores humanos através da métrica subjetiva. Portanto, o presente trabalho apresentou uma métrica para avaliação da qualidade de vídeo digital através da modificação do PSNR. O primeiro passo do processo consistiu na aplicação de técnicas de segmentação de imagem/vídeo com o intuito de agrupar os pixels da imagem/vídeo baseado na similaridade (valores de luminância). Duas técnicas foram testadas, onde a primeira utilizou um sistema de busca chamada meta-heurística ACO, e a outra, usou o modelo MDRAV para segmentar vídeo. Em relação aos resultados obtidos, a técnica ACO mostrou-se promissora, porém demanda maiores recursos computacionais. Logo, a sequência do trabalho foi avaliar a qualidade de segmentação alcançada pela técnica MDRAV.

Admitindo que os resultados individuais da segmentação foram razoáveis, seguiu-se para o cálculo da estimativa da relevância dos objetos. Como era esperado, o objeto que estava no primeiro plano apresentou uma maior estimativa quando comparado com o objeto do plano de fundo, indicando um maior foco visual de atenção do observador humano. Posteriormente, foi efetuada a avaliação global da segmentação (casamento entre a avaliação individual e a relevância contextual), onde a técnica obteve a nota de 0,60 em uma escala de  $[0,1]$ , sendo 0 o pior caso.

A última fase do trabalho foi avaliar a qualidade de vídeo digital com testes sobre vídeos degradados por diferentes artefatos. Com isso, através de uma função logística para efetuar a predição dos valores obtidos pelas métricas e utilizando dois coeficientes de correlação, é possível concluir que o PSNR modificado atingiu seu objetivo de aumentar a correlação com a avaliação subjetiva. Ressalta-se que este comportamento foi testado para vídeos com conteúdo estável, porém fazem-se necessários futuros estudos para verificar os casos onde, como por exemplo, só uma região do vídeo apresenta artefatos (fundo). Com este exemplo, será possível verificar a eficiência da métrica, pois como o foco de atenção do observador será o primeiro plano e não o fundo, a métrica proposta dará um valor alto (indicando qualidade) enquanto o

método tradicional, que avalia todos os pixels da mesma forma, dará uma nota baixa (indicando uma baixa qualidade).

Como contribuições do trabalho, cita-se: 1) o desenvolvimento de duas ferramentas para avaliação de vídeo (segmentação e qualidade); 2) proposta de uma técnica para segmentar imagens; 3) replicação de um trabalho discutido na literatura; 4) testes e simulações.

É importante salientar que existe a necessidade de estudos com artefatos reais, obtidos através da transmissão e codificação, para verificar se a métrica proposta se comporta bem para os casos práticos ou reais. Sendo assim, espera-se que a métrica proposta obtenha valores significativos, quando comparados com a PSNR, de correlação com a métrica subjetiva. Uma sugestão é melhorar o gerador de artefatos fazendo com que as degradações sejam geradas de forma independente, ou seja, para cada vídeo gerar 3 vídeos com os determinados artefatos. O primeiro apresentaria o artefato no vídeo como um todo, o segundo somente na região de fundo e o terceiro somente na região de plano de frente. Como a métrica proposta avalia separadamente cada objeto pesando sua relevância na cena, espera-se que a nota dada pela métrica se aproxime da real qualidade observada.

Como trabalhos futuros sugere-se a implementação do algoritmo de avaliação da qualidade de segmentação para vídeos com conteúdo de alta movimentação. Como melhoria, pretende-se avaliar outras técnicas de segmentação, e se estas obtiverem melhores resultados, serão publicados trabalhos comparando-as com os resultados obtidos por esta dissertação.

## REFERÊNCIAS

- ALBINI, F. L. P. **Geração de Artefatos em Vídeo Digital**. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná - Campus Curitiba, CPGEI, 2009.
- AUWERA, G. Van der; DAVID, P.; REISSLEIN, M. Traffic characteristics of h.264/avc variable bit rate video. **Communications Magazine, IEEE**, v. 46, n. 11, p. 164 –174, november 2008. ISSN 0163-6804.
- BEZERRA, S. A. C. **Modelo para Detecção de Regiões de Atenção Visual em Vídeo Digital**. Curitiba, 2011.
- BRAMER, M. **Principles of data mining**. Springer, 2007. (Undergraduate topics in computer science). ISBN 9781846287657. Disponível em: <<http://books.google.com/books?id=xVW7NsIHNHsC>>.
- CHANDLER, D. M.; HEMAMI, S. S. **VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images**. 2007.
- CONCI, A.; AZEVEDO, E.; LETA, F. R. **Computação Gráfica**. Rio de Janeiro: Elsevier Editora Ltda, 2008.
- CORREIA, P.; PEREIRA, F. Estimation of video object’s relevance. **Proceedings of the 10th European Signal Processing Conference (EUSIPCO ’00)**, v. 2, p. 925–928, 2000.
- CORREIA, P.; PEREIRA, F. Objective evaluation of video segmentation quality. **IEEE Transactions on Image Processing**, v. 12, n. no.2, p. 186–200, 2003.
- CORREIA, P.; PEREIRA, F. Video object relevance metrics for overall segmentation quality evaluation. **EURASIP J. Appl. Signal Process.**, Hindawi Publishing Corp., New York, NY, United States, v. 2006, p. 214–214, January 2006. ISSN 1110-8657. Disponível em: <<http://dx.doi.org/10.1155/ASP/2006/82195>>.
- CORREIA, P. L.; PEREIRA, F. Stand-alone objective segmentation quality evaluation. **EURASIP J. Appl. Signal Process.**, Hindawi Publishing Corp., New York, NY, United States, v. 2002, p. 389–400, January 2002. ISSN 1110-8657. Disponível em: <<http://portal.acm.org/citation.cfm?id=1283100.1283141>>.
- DORIGO, M. **Optimization, Learning and Natural Algorithms**. Tese (Doutorado) — Politecnico di Milano, 1992.
- DORIGO, M. Learning by probabilistic boolean networks. **Proceedings of the IEEE International Conference on Neural Networks**, p. 887–891, 1994.
- DORIGO, M.; MANIEZZO, V.; COLORNI, A. The ant system: Optimization by a colony of cooperating agents. **IEEE Transactions on Systems, Man, and Cybernetics-Part B**, v. 26, n. 1, p. 1–13, 1996.

- DURANT, L. **Comparing Objective and Subjective Picture Quality Measurements**. 1998. Acesso em: 01 de junho de 2011.
- EDEN, A. No-reference image quality analysis for compressed video sequences. **Broadcasting, IEEE Transactions on**, v. 54, n. 3, p. 691 – 697, sept. 2008. ISSN 0018-9316.
- ERDEM, C. E.; SANKUR, B.; TEKALP, A. M. Performance measures for video object segmentation and tracking. **IEEE Transactions on Image Processing**, vol.13, n. no.7, p. 937–951, 2004.
- FACON, J. **Morfologia Matemática, Teoria e Exemplos**. Curitiba: Editora Universitária Champagnat, Pontifca Universidade Católica do Paraná, 1996.
- FEI, Y.; LIANFEN, H.; YAN, Y. An improved psnr algorithm for objective video quality evaluation. **Proc Chinese Control Conference CCC 2007**, IEEE, p. 376–380, 2007.
- FRIEL, J. J. Image measurements. **Practical Guide to Image Analysis**, p. 101–128, 2000.
- GASPAR, J. F. C. **Objective Evaluation of Video Segmentation Quality**. Dissertação (Mestrado) — Instituto Superior Técnico - Universidade Técnica de Lisboa, 2009.
- GONZALEZ, R. C.; WOODS, R. E. **Processamento Digital de Imagens**. 3. ed. São Paulo: Pearson Prentice Hall, 2010.
- GROUP, V. T. R. **YUV 4:2:0 Video Sequences**. 2009. Disponível em: <http://trace.eas.asu.edu/yuv/qcif.html>. Acesso em: 01 de abril de 2009.
- GUNAWAN, I. P.; GHANBARI, M. Reduced-reference picture quality estimation by using local harmonic amplitude information. **Proc. London Communications Symposium 2003, University College London, UK**, v. 8-9, p. 137–140, 2003.
- GUNAWAN, I. P.; GHANBARI, M. An efficient reduced-reference video quality metric. **Proceedings of the 26th Picture Coding Symposium (PCS 2007), Lisboa, Portugal**, p. 07–09, 2007.
- ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol.20, n. no.11, p. 1254–1259, 1998.
- ITU-R. **Recommendation BT.500: Methodology for the Subjective Assessment of the Quality of Television Pictures**. 2002.
- ITU-T. **Recommendation P.910: Subjective video quality assessment methods for multimedia applications**. 2008.
- JÄHNE, B. **Digital Image Processing**. Berlin: Springer, 2002.
- KAO, Y.; CHENG, K. An aco-based clustering algorithm. **Lecture Notes in Computer Science - Springer**, v. 4150, p. 340–347, 2006.
- KOUMARAS, H.; PLIAKAS, T.; KOURTIS, A. A novel model method for pre-encoding video quality prediction. **Mobile and Wireless Communications Summit**, v. 16th IST, p. 1–4, 2007.

- KUROSE, J. F.; ROSS, K. W. **Redes de Computadores e a Internet**. 3. ed. São Paulo: Addison Wesley Bra, 2006.
- LUCCA, E. V. D. **Avaliação e Comparação de Algoritmos de Segmentação de Radar de Abertura Sintética**. Dissertação (Mestrado) — Mestrado em Sensoriamento Remoto - Instituto Nacional de Pesquisas Espaciais., São José dos Campos, 1998.
- MURTHY, A. V. **A MATLAB Based Framework For Image and Video Quality Evaluation**. Dissertação (Mestrado) — Arizona State University, maio 2010.
- MURTHY, A. V.; KARAM, L. J. **IVQUEST- Image and Video Quality Evaluation Software**. 2011. Disponível em: <http://ivulab.asu.edu/Quality/IVQUEST>. Acesso em: 04 de julho de 2011.
- NETBEANS. **NetBeans IDE 7.0.1 Develop desktop, mobile and web applications with Java, PHP, C/C++ and more**. 2009. Disponível em: <http://netbeans.org/>. Acesso em: 02 abril de 2009.
- OSBERGER, W.; BERGMANN, N.; MAEDER, A. Automatic identification of perceptually important regions in an image. **Proceedings of the 14th International Conference on Pattern Recognition**, vol.1, p. 701–704, 1998.
- PEDRINI, H.; SCHWARTZ, W. R. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações**. São Paulo: Thomson Learning, 2008.
- PEREIRA, D. A. B.; VENDRAMIN, A. C. B. K.; POHL, A. . Otimização por colônia de formigas e busca local na segmentação de imagens. **V SULCOMP - Congresso Sul Brasileiro de Computação, 2010, Criciúma. Anais do V Sulcomp**, V, 2010.
- PEREIRA, D. A. B.; VENDRAMIN, A. C. B. K.; POHL, A. Algoritmo de colônia de formigas aplicado na segmentação de imagens. **WCI 2010 - III Workshop on Computational Intelligence - Joint Conference, 2010, Sao Bernardo do Campo. Joint Conference 2010 - Workshops**, p. 434–439, 2010.
- PEREIRA, D. A. B.; VENDRAMIN, A. C. B. K.; POHL, A. de A. P. Colônia de formigas e busca local para definição automática de limiar na segmentação de imagens. **WebMedia - XVI Simpósio Brasileiro de Sistemas Multimídia e Web, 2010, Belo Horizonte. Anais de Artigos Resumidos e Artigos de Workshops do XVI Simpósio Brasileiro de Sistemas Multimídia e Web**, II, p. 37–40., 2010.
- PINSON, M.; WOLF, S. A new standardized method for objectively measuring video quality. **IEEE Transactions on Broadcasting**, v. 50, n. 3, p. 312–322, Setembro 2004.
- PYUV. **PYUV: raw video sequence player**. 2009. Disponível em: [http://dsplab.diei.unipg.it/pyuv\\_raw\\_video\\_sequence\\_player](http://dsplab.diei.unipg.it/pyuv_raw_video_sequence_player). Acesso em: 02 abril de 2009.
- RAMOS, V.; MUGE, F. Image colour segmentation by genetic algorithms. **RecPad'2000 - Proc. of the 11th Portuguese Conference on Pattern Recognition**, p. 11–12, 2000.
- RICHARDSON, I. E. **H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia**. 1. ed. Wiley, 2003. Hardcover. Disponível em: <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0470848375>.



SESHADRINATHAN, K. et al. A subjective study to evaluate video quality assessment algorithms. In: **Human Vision and Electronic Imaging**. [S.l.: s.n.], 2010. p. 75270.

SILVA, A. F. E. **Avaliação de Sistemas de Redução de Taxa de Bits de Vídeo com Enfoque na Percepção do Usuário**. Dissertação (Mestrado) — DECOM - FEEC - UNICAMP, Fevereiro 1998.

SILVA, E. S. R. da. **ESPECIFICAÇÃO, PROJETO E DESENVOLVIMENTO DE FERRAMENTAS DE AUXÍLIO À AVALIAÇÃO SUBJETIVA DE VÍDEO**. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná - Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, 2009.

SOILLE, P. **Morphological Image Analysis: Principles and Applications**. Berlin: Springer, 1999.

STOLFI, G. **Compressão de Imagens em Movimento, Padrões MPEG-2 e H.264**. Julho 2008. Disponível em: <<http://www.lcs.poli.usp.br/gstolfi/>>.

TRIOLA, M. F. **Introdução à estatística**. 9. ed. Rio de Janeiro: LTC, 2005. 650 p.

VQEG. **Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment**. April 2000. Disponível em: <http://www.vqeg.org/>.

VQEG. **Final Report from the Video Quality Experts Group on the validation of objective models of Video Quality Assessment Phase II**. Março 2003. Acessado em: 01 de junho de 2010. Disponível em: <<http://www.vqeg.org>>.

WANG, Y. **Survey of Objective Video Quality Measurements**. USA, 2006.

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

WANG, Z.; BOVIK, A. C. **Modern Image Quality Assessment**. USA: Morgan & Claypool, 2006.

WANG, Z. et al. Image quality assessment: From error measurement to structural similarity. **IEEE Transactions on Image Processing**, v. 13, n. 4, p. 600–612, 2004.

WOLF, S.; WEBSTER, A. Subjective and objective measures of scene criticality. **ITU Meeting on Subjective and Objective Audiovisual Quality Assessment Methods - JRG010**, p. 1, October 1997.

WU, H. R.; RAO, K. R. **Digital Video Image Quality and Perceptual Coding**. New York: CRC Press, 2006.

WU, Q.; MERCHANT, F. A.; CASTLEMAN, K. R. **Microscope Image Processing**. England: Academic Press, 2008.

ZHANG, Y. J. A survey on evaluation methods for image segmentation. **Pattern Recognition**, v. 29, p. 1335–1346, 1996.