DISSERTAÇÃO DE MESTRADO

DESCRITOR LOCAL BASEADO NO ALGORITMO SIFT PARA RASTREAMENTO E SEGMENTAÇÃO DE OBJETOS EM VÍDEO VIA GRAFOS DE REGIÕES

Gustavo Maia Queiroz de Mendonça

Brasília, Fevereiro de 2016

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO

DESCRITOR LOCAL BASEADO NO ALGORITMO SIFT PARA RASTREAMENTO E SEGMENTAÇÃO DE OBJETOS EM VÍDEO VIA GRAFOS DE REGIÕES

Gustavo Maia Queiroz de Mendonça

Relatório submetido ao Departamento de Engenharia
Elétrica como requisito parcial para obtenção
do grau de Mestre em Engenharia de Sistemas Eletrônicos e Automação

Banca Examinadora

Prof. Ricardo Lopes de Queiroz, Ph.D., CIC/UnB _____
Orientador

João Luiz Azevedo de Carvalho, Ph.D., ENE/UnB _____
Examinador interno

Camilo Chang Dorea, Ph.D., CIC/UnB _____
Examinador externo

Bruno Luiggi Macchiavello Espinoza, Dr., CIC/UnB _____
Examinador suplente

FICHA CATALOGRÁFICA

DE MENDONÇA, GUSTAVO MAIA QUEIROZ
DESCRITOR LOCAL BASEADO NO ALGORITMO SIFT PARA RASTREAMENTO E SEGMENTAÇÃO DE OBJETOS EM VÍDEO VIA GRAFOS DE REGIÕES [Distrito Federal] 2016.
xvi, 118 p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2016).
Dissertação de Mestrado - Universidade de Brasília, Faculdade de Tecnologia.
Departamento de Engenharia Elétrica

1. Descritor local	2. SIFT
3. Grafos	4. Segmentação de vídeo
I. ENE/FT/UnB	II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

DE MENDONÇA, G. Q. M. (2016). *DESCRITOR LOCAL BASEADO NO ALGORITMO SIFT PARA RASTREAMENTO E SEGMENTAÇÃO DE OBJETOS EM VÍDEO VIA GRAFOS DE REGIÕES*. Dissertação de Mestrado, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 118 p.

CESSÃO DE DIREITOS

AUTOR: Gustavo Maia Queiroz de Mendonça
TÍTULO: DESCRITOR LOCAL BASEADO NO ALGORITMO SIFT PARA RASTREAMENTO E SEGMENTAÇÃO DE OBJETOS EM VÍDEO VIA GRAFOS DE REGIÕES.
GRAU: Mestre em Engenharia de Sistemas Eletrônicos e Automação ANO: 2016

Gustavo Maia Queiroz de Mendonça
Depto. de Engenharia Elétrica (ENE) - FT
Universidade de Brasília (UnB)
Campus Darcy Ribeiro
CEP 70919-970 - Brasília - DF - Brasil

RESUMO

Na segmentação de objetos em vídeos por intermédio de um rastreamento quadro a quadro de regiões, a manutenção da coerência temporal depende diretamente da qualidade desse rastreamento ao longo dos quadros. Para esse fim, adaptou-se para o domínio dos superpixels processados como grafos de regiões, princípios de um extrator de características bastante difundido, o SIFT, que exibe grande eficiência na identificação/rastreamento de objetos em cenas. Um descritor é criado para cada região, a partir de histogramas de orientação do gradiente de setores ao redor do vértice, calculado de forma a garantir, como no SIFT, invariância à escala, rotação e iluminação. As contribuições do descritor proposto na segmentação de objetos em vídeo, feita a partir de corte em grafos, são testadas em três níveis: ajuste, ou compensação, de movimento do objeto em cena; reforço nos pesos de ligação entre arestas dos grafos, para os elementos considerados correspondentes entre os quadros e; determinação de grafos equivalentes com redução no número elementos guiada pela correspondência encontradas a partir algoritmo proposto.

ABSTRACT

In the segmentation of object in video through frame to frame region tracking, the temporal coherence maintenance depends directly on the quality of the regions tracking along the frames. To this aim, principles of a widespread feature extractor, the SIFT, were adapted for the superpixels domain rendered as region graphs, which exhibits high efficiency in identification/tracking of objects in scenes. A descriptor is created to each vertex of graph, from orientation histograms of the gradient of bins around the vertex, calculated to ensure, as the SIFT, a scale, rotation and lighting invariance. The contributions of the proposed descriptor in the segmentation of objects in video, performed by a graph cut, are tested on three levels: the adjustment or compensation of the movement of object in scenes; the strengthening of the connection weights between edges of the graphs for the elements considered matches between frames and; the determination of equivalent graphs with reduction in the number elements guided by matches found through the proposed algorithm.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS, ABREVIAÇÕES E ACRÔNIMOS

3D	Tridimensional
AC	Acurácia
AJ	Ajustado
AJQ	Segmentação quadro a quadro aplicada em um grafo do tipo ajustado
AJT	Segmentação em toda a sequência aplicada em um grafo do tipo ajustado
BK	Fundo
BKQ	Rastreamento quadro a quadro das regiões referentes ao fundo
BKT	Rastreamento em toda a sequência das regiões referentes ao fundo
CIELAB	Espaço de cores em oposição (L*a*b) da *Commission internationale de l'éclairag*
EQ	Equivalente
EQQ	Segmentação quadro a quadro aplicada em um grafo do tipo reforçado
EQT	Segmentação em toda a sequência aplicada em um grafo do tipo reforçado
GT	*Ground truth*
N	Número de elementos
NA	Não ajustado
NE	Número de elementos equivalentes
NAQ	Segmentação quadro a quadro aplicada em um grafo do tipo não ajustado
NAT	Segmentação em toda a sequência aplicada em um grafo do tipo não ajustado
NGL	Núcleo geniculado lateral
RE	Reforçado
REQ	Segmentação quadro a quadro aplicada em um grafo do tipo reforçado
RET	Segmentação em toda a sequência aplicada em um grafo do tipo reforçado
RGB	*Red, green and blue*
SE	Sobrestimação
SIFT	*Scale-invariant feature transform*
SLIC	*Simple linear iterative clustering*
OB	Objeto
OBQ	Rastreamento quadro a quadro das regiões referentes ao objeto
OBT	Rastreamento em toda a sequência das regiões referentes ao objeto

1 INTRODUÇÃO

1.1 CONTEXTO

A segmentação de objetos ou regiões de interesse em vídeos é um problema básico em visão computacional. Na segmentação de áreas de interesse ao longo de quadros de um vídeo, regiões que convirjam ao longo de vários quadros, se faz necessário um agrupamento/rotulação não supervisionado de pixels ou elementos. Em geral, esses agrupamentos utilizam relações de textura, cor e/ou movimento para serem construídos [1], essas relações se dão entre pixels próximos, vizinhos, ao longo do espaço e ao longo do tempo.

Quando se realiza um agrupamento em mais de um quadro simultaneamente, os pixels tomam formato de uma unidade espaço-temporal, unidades de volume conhecidas como voxels, que quando relacionados a um mesmo bloco de voxels, são chamados de supervoxels. Essa abordagem relaciona os pixels dentro de um quadro e de seus vizinhos, consequentemente A quantidade de dados gerados por um volume espaço-temporal de um vídeo, ao longo de um número pode demandar um grande esforço computacional, principalmente ao se analisar os elementos como volumes [2, 3].

Uma forma de se reduzir o esforço computacional produzido por uma análise de volumes ao longo de vários quadros, é agrupar previamente os pixels de um quadro em regiões, chamadas de superpixels. Esse procedimento elimina redundâncias em um quadro agrupando pixels semelhantes e grandes regiões de pixels, reduzindo o número total de elementos por quadro. Ao se analisar elementos vizinhos sejam eles vou apenas por suas relações de vizinhança, sejam eles pixels ou voxels, superpixels ou supervoxels, podemos representá-los como um problema de grafos, amplamente utilizado em agrupamento de regiões e segmentação de imagens.

A segmentação de imagem por meio de grafos demostram uma alta performance quando orientadas a um objeto, ou seja, quando um usuário define uma certa quantidade de elementos pertencentes ao fundo ou ao objeto [4]. Entretanto, essa abordagem supervisionada se torna ineficiente para vários quadros, fazendo-se necessário a intervenção de um algoritmo que oriente a segmentação do objeto automaticamente, um rastreamento. Em geral, os algoritmos de rastreamento mais difundidos se restringem ao domínio dos pixels, os menores elementos que representam uma imagem, apresentando poucas representações para superpixels.

1.2 APRESENTAÇÃO DO PROBLEMA E JUSTIFICATIVA

A segmentação hierárquica utilizando grafos [5, 6] vem sendo aplicada para redução do esforço computacional ao se analisar um vídeo por meio de voxels. Uma abordagem de correspondências entre superpixels é vastamente empregada dada [1, 7, 8, 4, 9], entretanto questiona-se a manutenção da coerência temporal e espacial, dada a instabilidade de uma segmentação quadro a quadro [5], ou seja, pode haver distorções entre regiões correspondentes de um quadro e seu vizinho, bem como regiões com erros de correspondência.

Para uma boa manutenção da coerência temporal, se faz necessária uma técnica de rastreamento que aumente o desempenho do casamento de superpixels tendo como base o confronto entre as características de aparência (cor, textura) e posição dessas regiões ao longo do tempo. Essas propriedades são limitadas, isto é, duas regiões pertencentes a dois quadros consecutivos apresentando níveis de cor ou posição muito próximas (senão iguais) não representam necessariamente um mesmo objeto. O movimento de regiões entre quadros deve ser estimado para uma melhor utilização de propriedades que projetam pouca informação. A definição de características ditas discriminantes para essas regiões (superpixels) também estabelece uma boa relação entre quadros, consequentemente, um bom rastreamento.

A Transformação de Características Invariante à Escala (SIFT) [10] é um algoritmo bastante difundido em visão computacional. Inspirado em algoritmos que tentam imitar o funcionamento do sistema visual humano, o SIFT tem se mostrado eficiente na captura de pontos relevantes de uma imagem para confronto de características, seja para rastreamento ou reconhecimento de objetos em imagens ou vídeo. A aplicação do algoritmo SIFT em processamento de vídeos não se restringe apenas ao rastreamento de objetos [11], sendo aplicada, por exemplo, em estabilização de vídeos [12].

O algoritmo SIFT trabalha com a seleção de pontos especiais para a extração de características, a fim de aumento na precisão no casamento de pontos e diminuição de esforço computacional. A utilização do histograma associado às regiões e de um descritor produzido pela SIFT, ou semelhante, exibe uma melhora de desempenho na segmentação vídeos [5, 13, 14]. Entretanto, o cálculo dos descritores é realizado nas imagens construídas por pixels, ou seja, não se aproveita a simplificação das imagens enquanto representadas por regiões e nem a redução do esforço computacional associado a essa simplificação.

1.3 MÉTODOS PROPOSTOS

Este trabalho tem como objetivo a construção de um descritor local que opere em uma imagem construída por superpixels, criando vínculos entre superpixels de quadros consecutivos, as quais podem definir, com razoável grau de coerência, regiões correspondentes. Pare esse fim, adaptou-se o algoritmo SIFT para extração de características em regiões de pixels. Primeiramente, a técnica *watershed* é aplicada em um espaço de escala, em conjunto com o algoritmo de agrupamento SLIC [15], efetuando uma sobre-segmentação das imagens, os superpixels.

Em seguida, as definições de gradiente são adaptadas de forma a atender regiões de diferentes tamanhos, posições (por vezes conflitantes) e conformações. O descritor de gradientes é construído a partir das vizinhanças de uma região em análise. Diferentemente da SIFT, processo é realizado para todos os elementos de uma imagem, porém, em uma escala determinada.

Para testar a contribuição do descritor proposto segmentações utilizando um corte de grafos de regiões foram aplicadas, visando o isolamento do objeto ao longo de 9 quadros. Os pesos de ligação dos grafos foram determinados de 4 maneiras, de forma medir as possíveis contribuições do rastreamento e casamento de regiões proposta. A segmentação de objetos em cenas é tem como base a segmentação manual (*ground truth* - GT) do objeto no primeiro quadro, por meio dessa referência deseja-se avaliar a contribuição do descritor proposto no rastreado e segmentação dos objetos a partir do primeiro até o último quadro das sequências testadas.

1.4 APRESENTAÇÃO DO MANUSCRITO

Incluindo este capítulo de Introdução, o trabalho se desenvolve em total de oito capítulos. O Capítulo 2 apresenta definições sobre o funcionamento do sistema visual humano, os processos biológicos que são base para o algoritmo SIFT e, de forma resumida, apresenta a implementação desse algoritmo que inspira o presenta trabalho. Para uma melhor compreensão dos métodos de segmentação propostos, bem como o campo da teoria dos grafos adotada, o Capítulo 3 traz definições desses conceitos. O processo de segmentação por escalas e a extração de características de regiões que são detalhadamente explicados nos Capítulos 4 e 5, respectivamente. Os resultados são apresentados no Capítulo 6 e as conclusões acerca desses resultados, inclusas propostas de trabalhos futuros, no Capítulo 7.

2 SISTEMA VISUAL HUMANO E O ALGORITMO SIFT

2.1 INTRODUÇÃO

São abordados neste capítulo princípios de anatomia e processos fisiológicos humanos que estão envolvidos no sistema visual. O início do capítulo envolve aspectos da anatomia do olho humano, como as estruturas que o compõem se organizam para a formação de imagens na retina e como as células receptoras se distribuem ao longo dessa rede receptora. Essas primeiras definições dão base para o entendimento dos campos receptivos, que funcionam de acordo com os arranjos formados pelas células receptoras na retina, extraindo informações importantes para os processos da percepção, que inspiram o algoritmo SIFT (transformação de características invariante à escala). Esse algoritmo, que reflete princípios biológicos em um descritor de gradientes, se mostra um referencial para algoritmos de reconhecimento e confronto de características e é brevemente descrito ao fim deste capítulo.

2.2 VISÃO HUMANA

Para compreender as demandas e aplicações de grafos em processamento de imagens, é essencial uma breve explanação sobre alguns aspectos da visão e percepção humana, visto que a resolução de problemas que contemplem a captura e interpretação de cenas, é o principal objetivo em visão computacional. Mesmo quando registros não são realizados via ondas eletromagnéticas, como na ultrassonografia, ou na faixa não visível do espectro eletromagnético, como no caso de aquisições térmicas, imagens médicas em geral (ressonância magnética, radiografia), busca-se a interpretação desses dados em mapas de magnitude que traduzam essas informações para o espectro visível.

2.2.1 Estrutura dos olhos e formação da imagem na retina

De maneira superficial, apresentando a clássica analogia do olho humano a uma câmera fotográfica, ambos são dotados de mecanismos para controle de entrada de luz, lentes para ajuste do foco e receptores para transdução da luz. A luz inicia seu percurso pelo olho humano através da córnea, que está diretamente conectada à esclera (branco do olho), esse conjunto forma a proteção inicial de todo o aparato visual. O olho humano tem um formato aproximadamente esférico com cerca de 20 mm de diâmetro [16].

Por meio de dois músculos, a pupila controla a entrada de luz que atinge a retina. Tanto por uma questão de proteção, em situações de grande exposição luminosa, quanto para uma melhora no contraste da imagem, em momentos de escassez de luz. A resposta da pupila a estímulos externos que atingem a retina é simétrica, controlada pelo sistema nervoso autônomo [16].

O cristalino é uma estrutura translúcida, constituída basicamente de água e proteínas, tendo suas dimensões controladas por meio do músculo ciliado. Apesar da analogia proposta, e da intuição induzirem o pensamento de que o sistema de lentes do olho humano é composto apenas pelo cristalino, esse tem um papel importante na formação de uma imagem, mas uma contribuição pequena na refração da luz até a chegada à retina. Grande parte da refração sofrida pela luz é gerada pela curvatura da córnea, que se soma à do cristalino e à refração oferecida pelos fluidos que constituem os humores aquoso e vítreo.

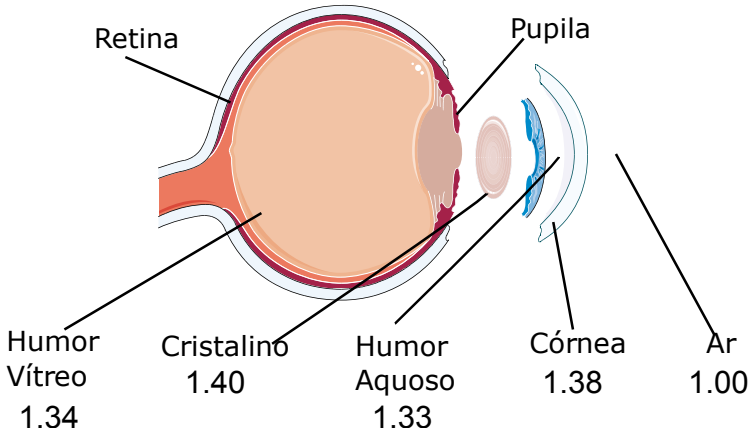


Figura 2.1: Estruturas do olho que operam no controle da entrada de luz e sua refração para formação da imagem no olho (inspirado em [17]).

A contração do músculo ciliado alivia a tensão nos ligamentos de suspensão do cristalino, o qual entra em conformação mais arredondada para ajuste de foco de imagens na retina, originadas de objetos mais próximos aos olhos (cerca de 10 cm). Em estado relaxado, essa musculatura aumenta a tensão nos ligamentos de suspensão e deixa o cristalino mais achatado, em uma conformação ideal para objetos que, devido à distância, têm incidência de raios aproximadamente paralelos na córnea. Um objeto muito próximo aos olhos exige um ajuste no cristalino para compensar a incidência difusa dos raios que chegam à córnea.

A luz que atravessa o cristalino em direção à córnea, passa pelo fluido translúcido e impuro do humor vítreo. Tal impureza pode ser notada por meio das partículas suspensas que são exibidas no campo visual, chamadas de mocas volantes. Salienta-se tal peculiaridade do olho humano para destacar a capacidade do cérebro em eliminar ou ignorar certos tipos de interferências causadas pelas limitações, ou propriedades intrínsecas, como no caso dos cílios, nariz e, por ventura, armações de óculos.

2.2.2 Cones, bastonetes e a transdução do sinal luminoso

Ao ser projetada na retina, a imagem é enviada ao cérebro na forma de sinais nervosos que são inciados nos fotorreceptores, que se exibem em dois tipos, cones e bastonetes. Ambos receptores são ativados por meio de proteínas específicas que se decompõe na presença de luz, em um processo em cascata que chega a amplificar em milhões de vezes um estímulo. Os bastonetes, por exemplo, podem atingir metade de sua saturação na presença de 30 fótons de luz. A relação entre o potencial desencadeado no receptor e a intensidade de luz absorvida pode ser aproximada por uma relação logarítmica, permitindo aos olhos operarem em uma vasta faixa de intensidades [17], comparativamente a uma relação linear.

Além de seus formatos que dão referência aos seus nomes no português, o que diferencia cones e bastonetes são as proteínas envolvidas no processo de absorção de luz. Ativados pela proteína rodopsina, os bastonetes respondem a parte do espectro visível, são responsáveis pela percepção em baixo nível da visão humana, com grande contribuição na detecção de movimentos na periferia do campo visual e durante a visão noturna.

Os cones se apresentam em três tipos, representados por três variantes de proteínas com diferentes respostas a diferentes faixas do espectro de luz. Antes da constatação da existência desses três tipos de receptores, sensíveis principalmente às cores vermelha, verde e azul, tal fato foi previso 200 anos antes pelo físico Thomas Young, que demonstrou que todas as cores do arco-íris, incluindo o branco, poderiam ser obtidas a partir de proporção exata entre o vermelho, verde e azul. Young induziu então que a retina do olho humano deveria perceber cores obedecendo uma codificação composta por essas três cores [16].

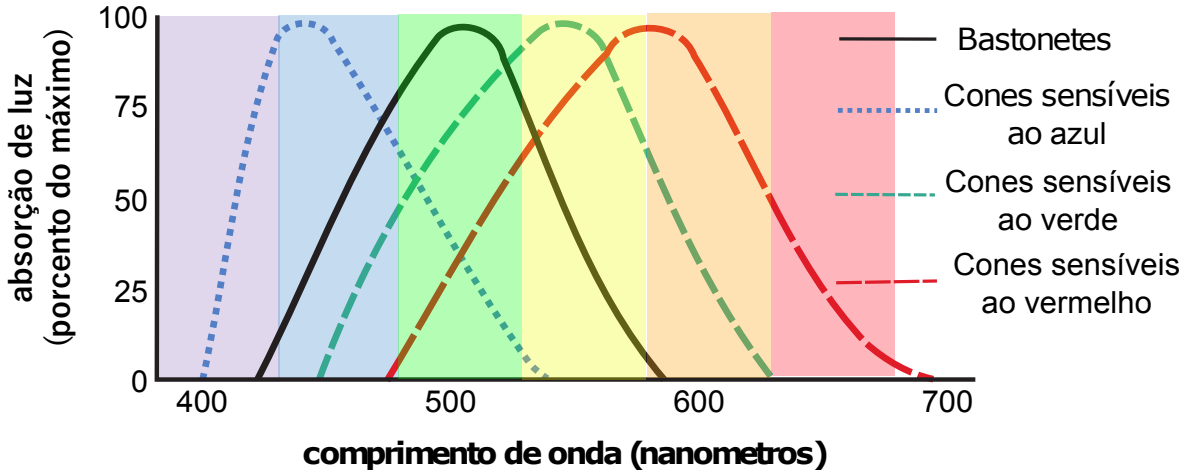


Figura 2.2: Resposta em codificação de cores por comprimento de onda das células receptoras da retina: cones e bastonetes (adaptado de [17]).

2.2.3 Distribuição dos fotorreceptores na retina

A quantidade de bastonetes na retina é da ordem de grandeza de 100 milhões de receptores, enquanto o número de cones é consideravelmente menor, em torno de 3 milhões de receptores (Figura 2.3) [17]. Grande parte dos cones se concentram na região da fóvea, região na qual não são encontrados bastonetes, a densidade de cones nesta região chega aos 150 mil elementos por mm^2, podendo ser comparada a um sensor quadrado com dimensões 1,5×1,5 mm [18]. Saindo da fóvea em direção à região mais periférica da retina, essa relação se inverte e a densidade de bastonetes se torna largamente maior que a de cones [16].

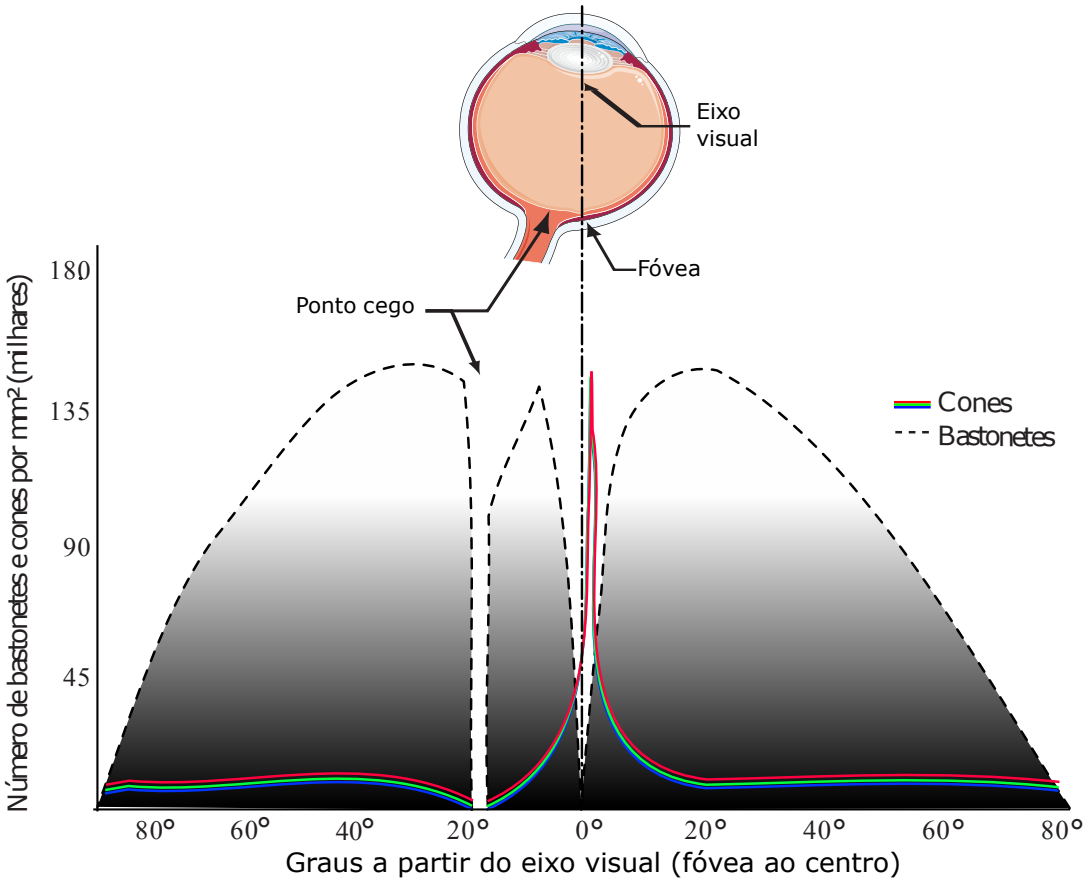


Figura 2.3: Gráfico da concentração de cones e bastonetes versus a ângulo de afastamento em relação à fóvea (adaptado de [18])

A percepção de uma visão periférica menos definida do que a central, relacionada a fóvea, é amplificada pela conexão de vários receptores a um mesmo neurônio ganglionar, o qual transmite o sinal para o córtex visual. Os 103 milhões de receptores são distribuídos por cerca de 1,6 milhão de células ganglionares, perfazendo uma média de 60 bastonetes por célula e 2 cones por célula. Na região da fóvea, cada cone está conectado diretamente a uma única célula ganglionar, cones que ao longo da região periférica vão se tornando maiores e mais escassos, enquanto em regiões mais periféricas, cerca de 200 bastonetes convergem a uma mesma fibra nervosa [17].

A Figura 2.4 ilustra a varição de concentração de tipos de receptores ao logo da retina e as conexões desses com as células ganglionares. A maior concentração de bastonetes por ganglionares amplifica o sinal luminoso em relação os recebidos pelos cones. Esse fato, aliado à característica intrínseca dos bastonetes de maior sensibilidade à luz, de 30 a 300 vezes mais sensíveis que cones, tornam os bastonetes importantíssimos durante estímulos sob baixa luminosidade e nas periferias do campo visual [17].

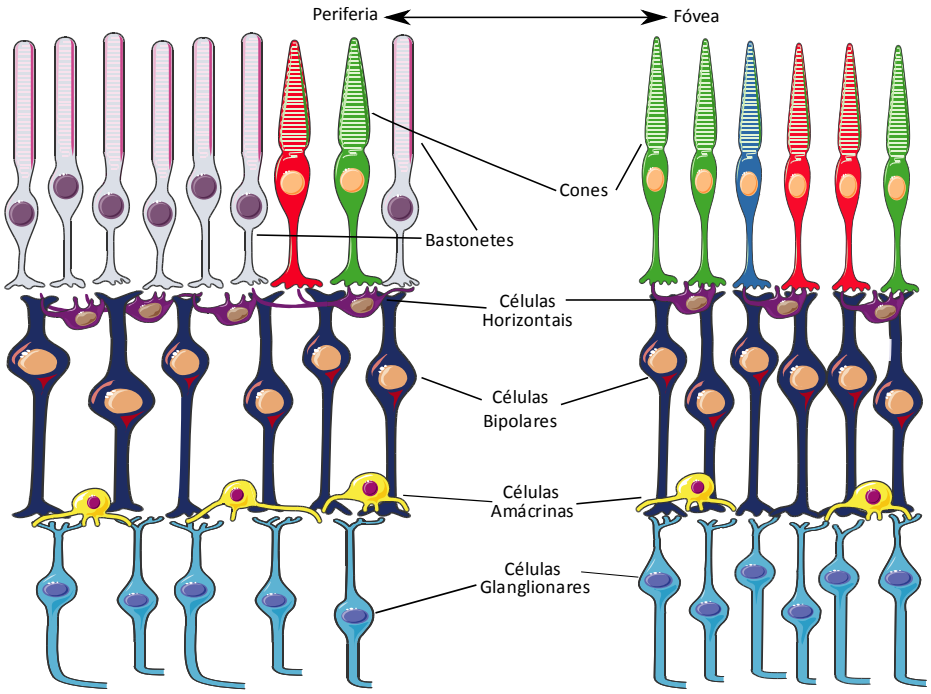


Figura 2.4: Ilustração da distribuição de receptores pela retina e suas conexões com células nervosas que transmitem e processam previamente o sinal recebido (inspirado em [17] e [16]).

O sinal oriundo dos receptores é transmitido até o córtex visual por meio das células ganglionares, entretanto, a ligação entre essas duas células é intermediada por neurônios bipolares. Essas células bipolares estão sob influência de inibitória das células horizontais, mecanismo que aprimora a percepção do contraste. Se apresentando em cerca de 30 tipos e expressando em meia dúzia de funções, as células amácrinas são interneurônios que ajudam na análise do sinal antes de sua chegada ao córtex visual.

2.3 CAMPOS RECEPTIVOS

Adaptado a partir de estudos para descrição de uma região da pele a qual quando submetida a um estímulo poderia induzir um reflexo, o termo campo receptivo foi utilizado para definir disposições de regiões no sistema nervoso, assim como na retina, regiões caracterizadas por uma resposta específica dada uma certa organização neurônios [19].

2.3.1 Modulação centro-periferia

Toda a área de fotorreceptores que circunda uma célula bipolar, e contribui para a despolarização de sua membrana, é considerado um campo receptivo, despolarização a qual atua na produção de impulsos nervosos. Células bipolares estão conectados diretamente a um conjunto de fotorreceptores, ou a uma unidade quando se trata das proximidades da fóvea, essas conexões diretas formam o centro do campo receptivo. A periferia do campo receptivo é determinada por aqueles receptores conectados a uma célula bipolar por meio de células horizontais, em uma conexão de inibitória.

As células bipolares são separadas em duas categorias, ON ou OFF, que representam a forma na qual as células respondem na presença ou ausência de luz. Pode-se ilustrar uma conexão centro-ON em uma configuração simples, com um bastonete no centro do campo receptivo e dois na periferia (Figura 2.5(a)).

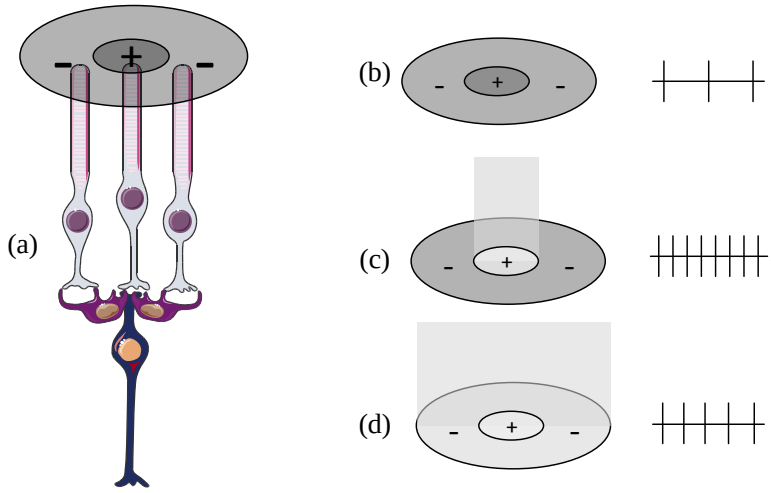


Figura 2.5: Organização dos campos receptivos: (a) configuração simples para uma célula bipolar centro-ON, com um bastonete no centro do campo receptivo e dois na periferia; (b) campo receptivo em estado de repouso; (c) resposta para um estímulo luminoso na região central de um campo receptivo centro-ON; e (d) em toda área do campo receptivo, percebe-se uma diminuição na frequência de disparos devido a inibição provocada pela presença de luz na periferia do campo (inspirado em [16]).

Em uma configuração centro-ON, há uma constante produção de impulsos no repouso (Figura 2.5(b)). A presença de luz no centro do campo receptivo promove a despolarização da célula bipolar (c) e um aumento na frequência de impulsos; caso essa luz atinga a periferia, a célula bipolar sofre uma inibição que decresce o número de impulsos (d).

Como mencionado, os campos receptivos se estendem a todo o sistema nervoso. Em uma conexão em maior escala, em um segundo nível, as células ganglionares também formam campos receptivos, em que a conexão direta com uma célula bipolar, define seu tipo, ON ou OFF. O

mesmo esquema de campos receptivos das células bipolares (Figura 2.5(b), (c) e (d)) pode ser estendido aos campos receptivos das células ganglionares.

As células ganglionares são classificadas em duas categorias principais, tipo-M e tipo-P. As células ganglionares tipo-M têm campos receptivos de grande área e são sensíveis a estímulos de baixo contraste, conduzindo-os de maneira mais rápida pelo nervo óptico. Os campos receptivos explicam ilusões de óptica referentes à percepção visual do contraste, como bordas que aparentam sofrer um realce devido a transição abrupta de iluminação de uma área para outra (Figura 2.6(a)). A densidade de iluminação ao redor de um ponto cria a ilusão de existência de pontos negros entre os vértices dos quadrados exibidos na Figura 2.6(b), quando se foca o olhar em um dos vértices, esse não apresenta o ponto, mas os demais sim, consequência da distribuição não-uniforme de fotorreceptores na retina.

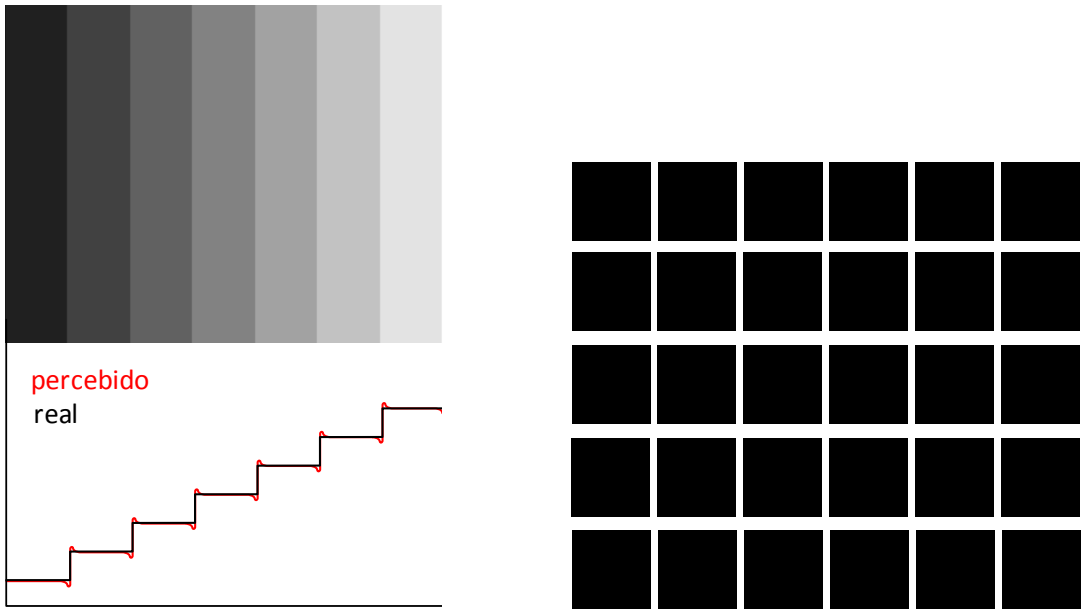


Figura 2.6: Ilusões de óptica de contraste: (a) para o cérebro e olho humano, um realce de bordas aparenta existir entre as barras, que é uma diferença entre a intensidade real e a percebida ao longo das transições entre as barras, cada uma com um nível de cinza constante, essa ilusão de realce é criada pelos campos receptivos e sua capacidade de detectar transições; (b) pontos escuros, que não existem na imagem, são percebidos entre os vértices dos quadrados, pela forma que as vizinhanças dos campos receptivos recebem luz para essa disposição de regiões claras e escuras, sendo que nas vizinhanças entre os vértices o número de regiões claras é maior. O efeito é maior na periferia da visão, onde os campos receptivos são mais extensos.

As células tipo-P têm resposta mais lenta rápida e com pulsos mais duradouros que as tipo-M. A sensibilidade quanto à diferença do comprimento de onda de um estímulo luminoso, também diferencia as células tipo-P das tipo-M. Na retina, os campos receptivos das células tipo-P são encontrados nas oposições de cores verde/vermelho e azul/amarelo. Os estímulos e as respostas são análogas às descritas para os campos receptivos de iluminação (Figura 2.7).

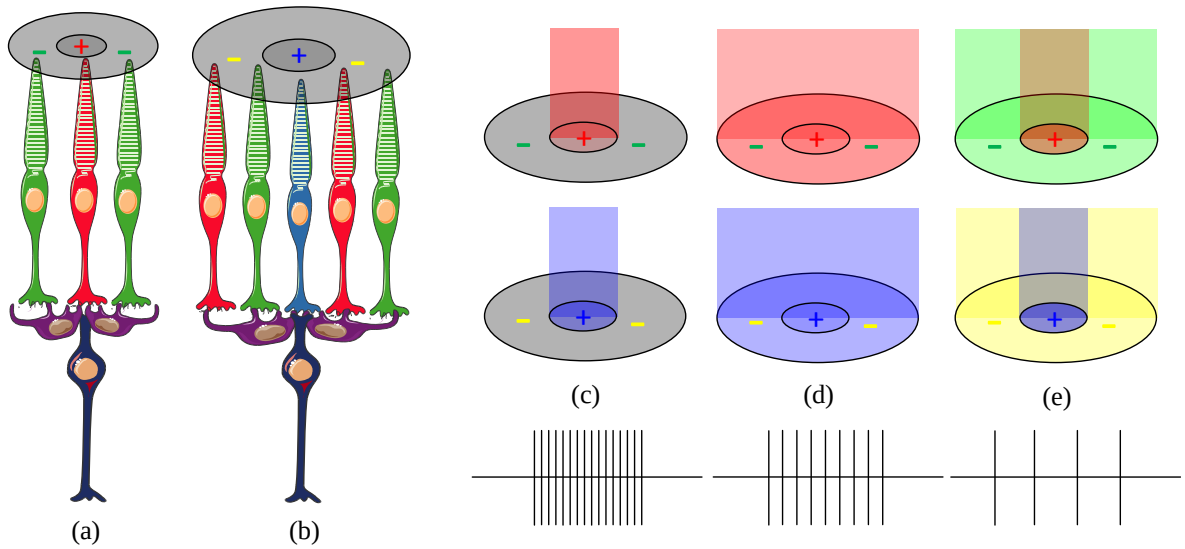
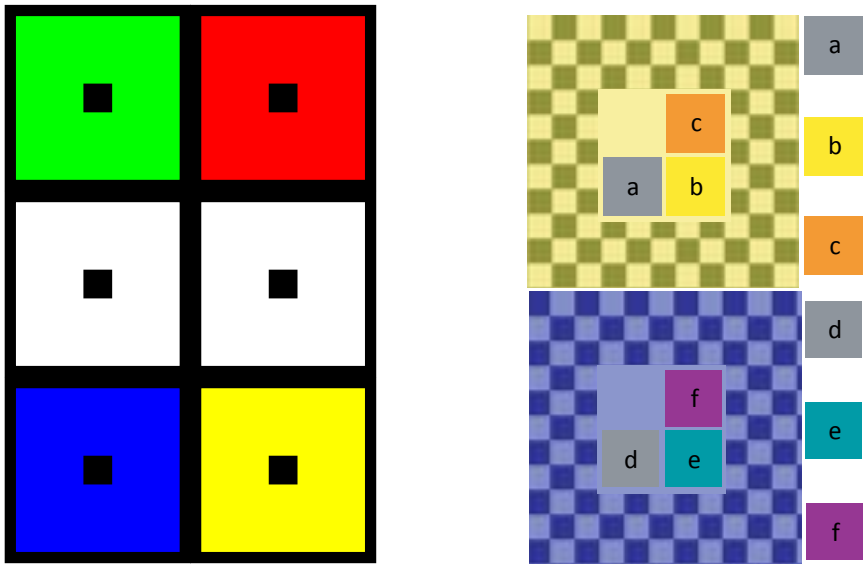


Figura 2.7: Organização dos campos receptivos em oposição de cores: (a) configuração simples para uma célula bipolar centro-ON para a cor verde em oposição à cor vermelha na periferia; (b) mesmo padrão para (a) mas para a oposição entre as cores azul, centro-ON, e amarela, periferia; (c) resposta para um estímulo luminoso convergente à cor da região central de um campo receptivo centro-ON; (d) frequência disparos cai em relação a (c) quando todo o campo é submetido a cor que caracteriza o seu centro; (e) uma maior inibição nos disparos de impulsos nervosos é causada pela incidência de luz na periferia do campo receptivo, na cor de oposição a do centro.

Os campos receptivos de cores, bem como as características tempo de resposta de suas células, podem imprimir um padrão de cores opostas na retina, quando essa é estimulada por uma única cor por um longo período de tempo, e em seguida se troca esse estímulo pela cor branca, que contém todos os comprimentos de onda (Figura 2.8(a)). Esse efeito seria explicado pela saturação de campos receptivos estimulados por uma mesma cor (Figura 2.7(b)), que ao ser trocado pela luz branca, que contém a cor de estímulo e a sua opositora, a qual é ressaltada.

As percepções visuais equivocadas quanto a oposição de cores não se limitam a imposições temporais. A oposição de cores para uma região e suas vizinhanças pode produzir a sensação de que uma região tem cor diferente da real, uma vez posicionada em uma vizinhança com certa cor. Na Figura 2.8(b), os quadrados ‘b’ e ‘d’ dentro da região quadriculada aparenta ter cores próximas, entretanto, visualizando-se os quadrados fora da região, percebe-se que os quadrados ‘a’ e ‘d’ são idênticos e que b diverge, efeito fruto da oposição azul/amarelo. Isso indica que os campos receptivos se estendem em várias escalas, da possibilidade de realce de bordas (Figura 2.6(b)) ao reforço de oposição para regiões (Figura 2.8(b)).



(a) (b)

Figura 2.8: Ilusões de óptica de cor: (a) ao se focar o olhar no centro de um dos quadrados coloridos por cerca de 30 s e se voltar o olhar para um dos quadrados brancos, logo acima ou abaixo, este fica preenchido pela a cor em oposição à focada anteriormente; (b) cor da vizinhança dos quadrados determina como a ele é percebido, note que o quadrado 'a' e 'd' têm cores idênticas, mas quando expostos a regiões distintas, produzem efeitos distintos.

2.4 RESPOSTA NEUROLÓGICA

As informações são enviadas pela retina por intermédio dos neurônios ganglionares, que incidem no núcleo geniculado lateral (NGL) [16]. O NGL (Figura 2.9(a)) de cada um dos hemisférios, recebe informações referentes ao lado oposto do campo visual, por exemplo, toda a imagem referente ao lado esquerdo do campo visual, seja oriunda do olho esquerdo ou do olho direto, segue pelo NGL direito.

Estudos sobre os potenciais de ação no NGL indicam que os campos receptivos ali são quase idênticos àqueles que o estimulam [16]. Em sua aparente disposição em camadas, as células do NGL parvocelular se assemelham a células ganglionares do tipo-P, com áreas centro-periferia pequenas e apresentando resposta a oposição de cores, verde/vermelho e azul/amarelo, em oposição luz/escuridão. Em contraste, as células do NGL magnocelular apresentam centro-periferia extensos e insensíveis a diferenças na frequência da luz de estímulo.

De forma semelhante ao que acontece com o NGL em relação às aferências dos neurônios ganglionares, grande parte de uma das camadas córtex estriado (Figura 2.9(a)), a camada IVC, responde de acordo com as células NGL magnocelular e parvocelular. Em outra camada do córtex estriado, a IVCα, neurônios possuem campos receptivos insensíveis à luz, e na camada IVCβ, neurônios apresentam centro-periferia operando em oposição de cores.

Os campos receptivos aparentam ter uma grande contribuição na forma em que o cérebro interpreta os estímulos visuais. Os estudos laboratoriais quanto ao funcionamento do córtex estriado ganharam força com os neurobiologistas David Hubel e Torste Wiesel, no início da década de 1960 [20]. Seus achados apontaram para uma gama de propriedades da percepção visual entre mamíferos, inclusive a organização de campos receptivos binoculares, essencial em seres humanos.

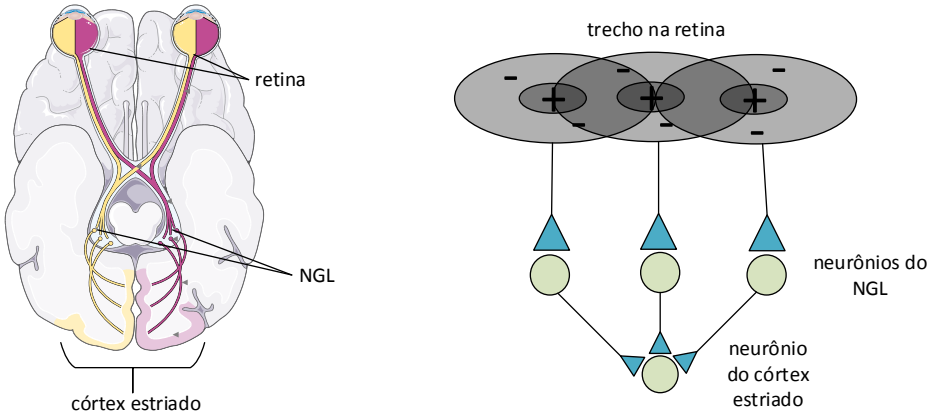


Figura 2.9: Percurso dos estímulos visuais pelo cérebro humano: (a) representação do caminho dos estímulos nervosos da retina, passando pelo núcleo geniculado lateral (NGL) até o córtex estriado; (b) ilustração para uma célula simples de um campo receptivo do NGL, que combina campos receptivos com origem na retina, que estimulam um neurônio no córtex visual.

Os trabalhos de Hubel e Wiesel constataram também uma seletividade quanto à direção de movimento, em que se observava resposta do córtex quando uma barra de luz se movimentava em certa direção, o mesmo efeito não era observado quando barra se movimentava na direção oposta, indicando a presença de neurônios no córtex especializados na análise de movimento.

2.4.1 Seletividade quanto à orientação

Nos estudos relacionados à seletividade de neurônios por certas orientações, destaca-se o experimento com uma barra de luz, que era posicionada em uma orientação ótima para o campo receptivo de um nerônio em análise, ressaltando o que é possivelmente uma das propriedades mais importantes na análise de objetos, essa seletividade quanto à orientação(Figura 2.10).

Grande parte dos neurônios da camada V1 do córtex estriado, bem como alguns da camada IVC, é seletiva à orientação. Uma organização simples de campos receptivos (Figura 2.9 (b)) pode explicar a predileção por uma certa orientação, entretanto, não se aplica a campos receptivos que respondem a orientação independentemente da posição da barra ao longo do campo, esse tipo de organização foi rotulada por Hubel e Wiesel como célula complexa.

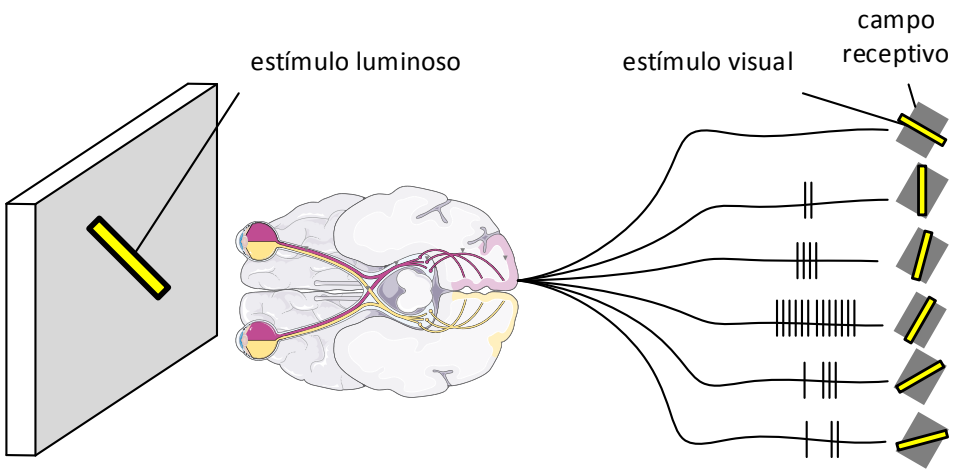


Figura 2.10: Uma célula simples de um campo receptivo no NGL promove diferentes respostas para diferentes orientações de um estímulo luminoso. A intensidade da resposta ao estímulo luminoso é codificada em frequência, quanto maior o número de impulsos, maior o estímulo visual, que tem seu pico quando a barra luminosa tem orientação igual à orientação de predileção do campo receptivo.

Tomando ensejo nas descobertas quanto ao funcionamento do córtex visual, [21] hipotetizou que a liberdade de posicionamento de células complexas ao definir a orientação de um estímulo, seria o ponto chave para o reconhecimento e casamento de características de objetos 3D ao longo de várias vistas. O SIFT [10] foi baseado nessa ideia, preservando a orientação posicional ao descrever um ponto por meio de sua vizinhança.

2.5 O ALGORITMO SIFT

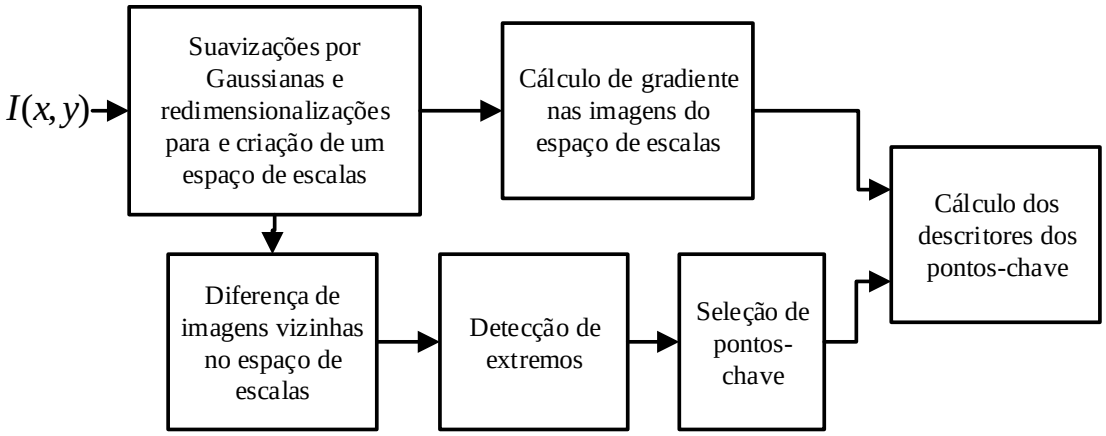


Figura 2.11: Diagrama de funcionamento do SIFT. O algoritmo tem como ponto principal a criação de um espaço de escalas para a determinação de mapas de gradiente e extração de pontos-chaves, para os quais serão criados os descritores locais.

O algoritmo SIFT segue 4 passos principais: (1) detecção e seleção de extremos em um espaço de escalas; (2) localização de pontos-chave; (3) definição da orientação e magnitude dos pontos-chave; e (4) criação de um descritor para os pontos-chave (Figura 2.11). Serão discutidos neste capítulo aqueles passos que inspiram os métodos adotados, não dando ênfase a passos sem um correspondente no algoritmo proposto, como a seleção de pontos-chaves.

2.5.1 Espaço de escalas

No primeiro passo, as escalas representam um conjunto de imagens oriundas da convolução da imagem da $I(x, y)$ com Gaussianas $G(x, y, k\sigma)$ de diferentes desvios padrão, determinados por um fator de escala k:

$$L(x, y, k\sigma) = I(x, y) * G(x, y, k\sigma), \tag{2.1}$$

em que:

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x^2+y^2)}{2\sigma^2}}. \tag{2.2}$$

Esse procedimento simula redimensionamentos consecutivos na imagem para a seleção de pontos que se preservam nessas mudanças de escala, pontos-chave. O espaço de escalas é divido por uma constante de escala k que varia em passo de $k = 2^{1/s}$ até atingir o valor de 2, ou seja, até dobrar o valor inicial de σ. A partir desse ponto, a imagem é reduzida por um fator 2 e o processo reiniciado, configurando uma oitava.

Segundo Lowe, para cobrir uma oitava objetivando invariância à escala, deve-se produzir $s + 3$ imagens dentro da oitava. Uma vez completada a oitava, pega-se a imagem que teve seu σ dobrado e inicia-se uma outra oitava com essa imagem subamostrada, promovendo uma redução em $2\times$ de suas dimensões (Figura 2.12). A detecção de extremos e pontos-chave é realizada nas diferenças entre a pilha de imagens suavizadas, uma oitava. Visto que as imagens de uma oitava estão suavizadas, uma maneira mais eficiente de se obter as diferenças é realizando a subtração em pares das imagens:

$$D(x, y) = L(x, y, k\sigma) - L(x, y, \sigma). \tag{2.3}$$

Denomina-se mais eficiente o método utilizando a equação (2.3), pois a convolução realizada com diferenças entre Gaussianas é uma boa aproximação do Laplaciano de uma Gaussiana normalizado $\sigma^2 \bigtriangledown^2 G$ originalmente utilizado proposto na referência [22]:

$$(k-1)\sigma^2 \bigtriangledown^2 G(x, y, \sigma) * I(x, y) \approx (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y), \tag{2.4}$$

uma vez que:

$$\big(G(x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma), \tag{2.5}$$

evita-se uma novas filtragens utilizando as imagens $L(x, y)$. A Figura 2.12 exibe uma ilustração que esquematiza os espaços de escala quanto as Gaussianas e suas diferenças.

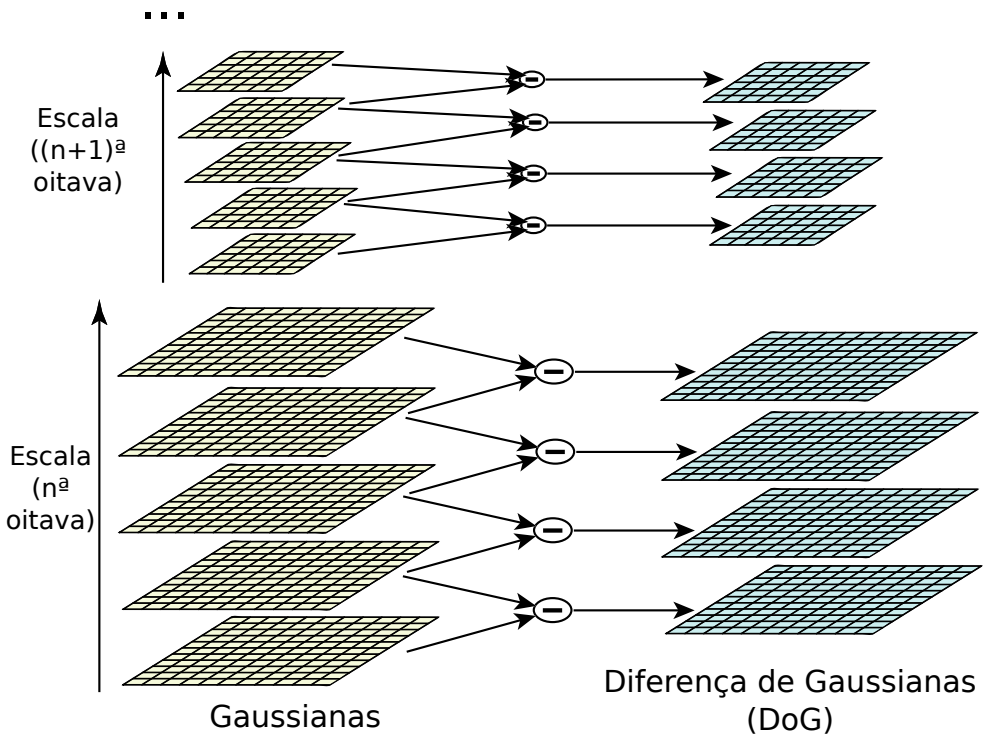


Figura 2.12: Espaço de escalas é criado a partir da diferença de Gaussianas que suavizam um imagem original, simulando uma redimensionalização dessa imagem. O crescimento na escala implica em uma maior suavização da imagem, ao atingir certo nível de suavização, a imagem tem dimensões reduzidas pela metade, configurando uma nova oitava (adaptado de [10]).

2.5.2 Detecção de extremos e seleção de pontos-chave

A detecção de extremos é feita em uma vizinhança 26-conectividade em torno de um pixel em uma pilha de três imagens de diferenças da oitava em sequência, estando este pixel na imagem central. Se um pixel central tem maior ou menor valor de diferença em relação aos seus 8 vizinhos na mesma imagem e seus 18 vizinhos nas outras duas imagens da pilha, ele é considerado um extremo.

Os pontos definidos como extremo são então candidatos a pontos-chave. Exclui-se desses extremos, pontos que possuam pouco contraste em relação a vizinhança, sensíveis a ruído, ou que representem arestas, que formariam descritores pobres em informação. Uma vez selecionados os pontos, esses restantes são chamados de pontos-chave e são fonte para a criação dos descritores.

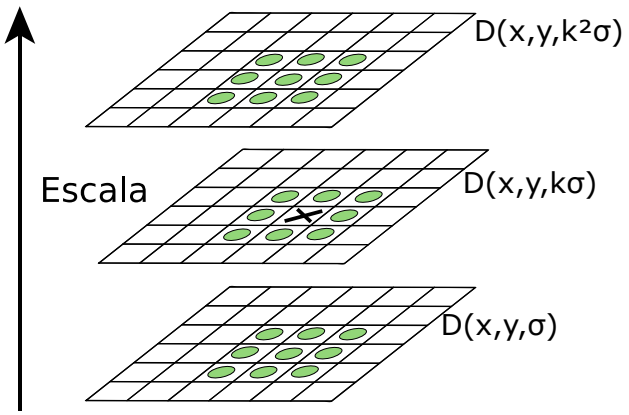


Figura 2.13: Para restrição no número de descritores criados para uma imagem, pontos extremos são detectados entre as imagens do espaço de escalas. Um pixel é considerado relevante se tem maior magnitude dentro de uma vizinhança de 26, 8 com origem na imagem da escala em análise e 18 pixels da imagens das escalas acima e abaixo do pixel em análise.

2.5.3 Determinação da orientação dos pontos-chave

A definição da orientação e magnitude dos pontos-chave, terceiro passo, é realizada na respectiva imagem do espaço de escalas do ponto-chave. Mapas de gradiente são criados, dos quais as projeções horizontal e vertical,

$$m_x = L(x+1, y) - L(x-1, y), \tag{2.6}$$

$$m_y = L(x, y+1) - L(x, y-1) \tag{2.7}$$

respectivamente, geram uma magnitude

$$m(x, y) = \sqrt{m_x^2 + m_y^2} \tag{2.8}$$

e uma orientação

$$\theta(x, y) = \tan^{-1}(m_y/m_x). \tag{2.9}$$

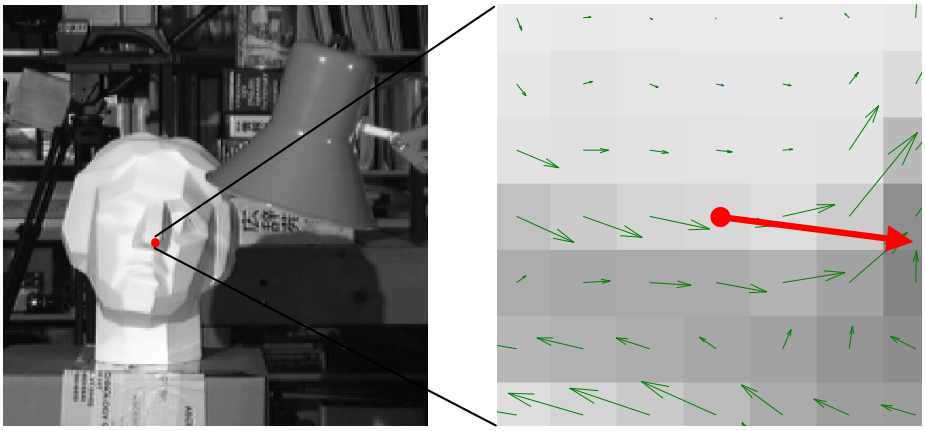
Pode-se representar essas propriedades vetorialmente, em que a magnitude e orientação de cada pixel é dada por:

$$\vec{\mathbf{m}}_i = \sum_{j \in V_i^4} L_j \vec{\mathbf{u}}_{i,j}, \tag{2.10}$$

sendo V_i^4 a vizinhança 4-conectividade do pixel i, L_i sua magnitude e $\vec{\mathbf{u}}_{i,j}$ é o vetor unitário que define a direção entre i e j, neste caso, o conjunto $[(\pm 1, 0), (0, \pm 1)]$. Essa forma vetorial de representação será útil em termos de analogia para os métodos propostos neste trabalho.

Um histograma de orientações é criado para uma região ao redor do ponto-chave (Figura 2.14(a)). Cada amostra adicionada ao histograma é ponderada pela magnitude do gradiente e por uma Gaussiana circular simétrica em relação à localização do ponto-chave. A orientação

de com maior magnitude dentro do histograma é selecionada como a orientação do ponto-chave (Figura 2.14(b)). A partir de um limiar, até três picos de orientações e suas respectivas magnitudes são relacionados à localização do ponto (Figura 2.15).



(a) (b)

Figura 2.14: Determinação da orientação de um ponto-chave: (a) selecionado um ponto-chave, ponto destacado com a cor vermelho; (b) determina-se a sua orientação de acordo com a orientação vencedora para histograma calculado com base mapa de gradiente $m(x, y)$ da escala em análise, o vetor em vermelho representa a orientação do ponto-chave, que é visivelmente a direção de maior frequência e amplitude para o mapa de gradiente ao redor do ponto-chave.

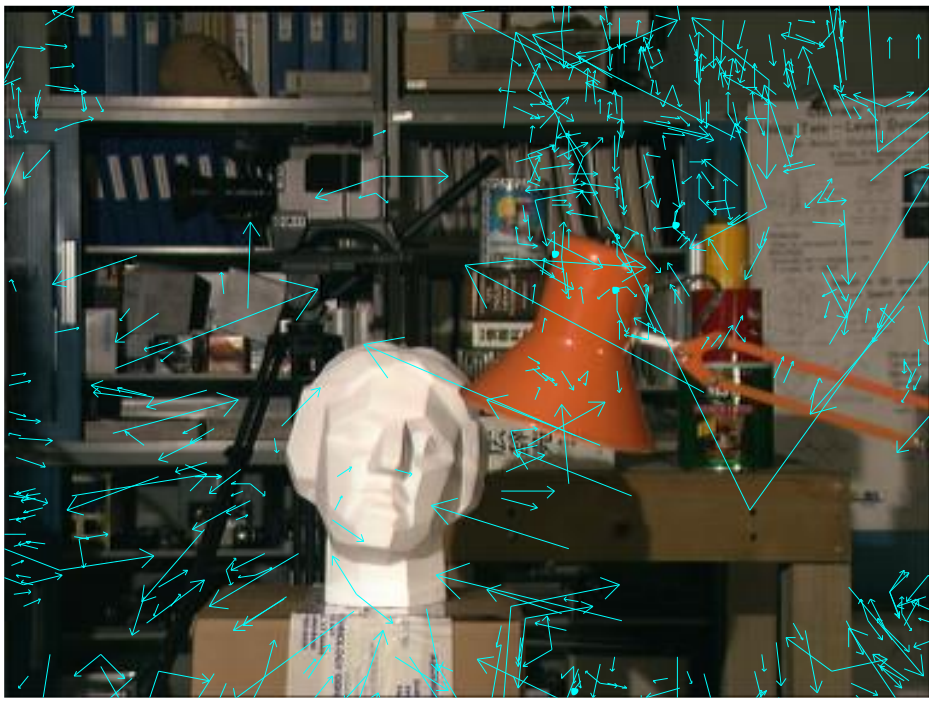


Figura 2.15: Pontos-chaves e suas respectivas direções e escalas representadas por vetores em três imagens distintas. A orientação do ponto-chave é representada pela direção do vetor e a escala pela magnitude do vetor, quanto menor o vetor, menor a escala, quanto maior, maior a escala. Uma mesma posição pode estar associado a mais de um ponto-chave.

2.5.4 Criação dos descritores

Novos histogramas são calculados em setores ao redor do ponto-chave, com as orientações dos gradientes referenciadas pela orientação do ponto-chave, o que gera invariância à rotação. Os histogramas, com elementos também ponderados pelas magnitudes dos gradientes e pela janela Gaussiana, são dispostos em um único vetor, definido como descritor do ponto-chave. Esse descritor é normalizado em valores entre 0 e 1, objetivando a invariância às mudanças na iluminação.

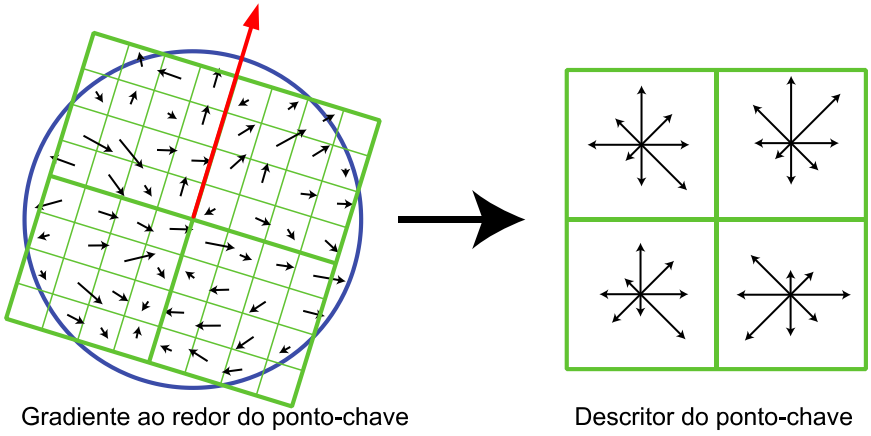


Figura 2.16: Grade retangular para cálculo dos histogramas de orientação do mapa gradiente ao redor de um ponto-chave. À esquerda o mapa de magnitudes e orientações (gradiente) é dividido em quatro setores ao redor do ponto-chave, no caso, cada uma com 16 pixels. Essa grade de setores deve ter a mesmo alinhamento do ponto-chave em análise, representada pela seta em vermelho. Para cada setor, a magnitude e orientação desses pixels, referente ao mapa de gradiente da escala e oitava em análise, são amostrados em um histograma de orientação ponderado por uma Gaussiana de desvio padrão proporcional a escala (indicado com círculo azul) e pela magnitude do vetor. O histograma de cada setor para 8 orientações é observado na imagem da direita (traduzido de [10]).

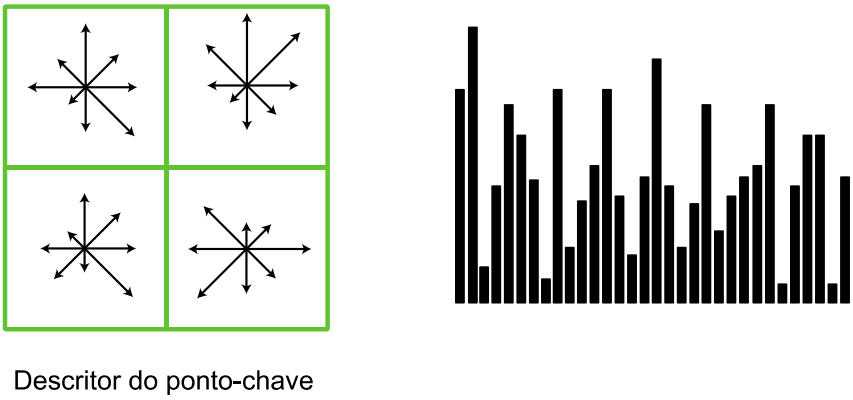


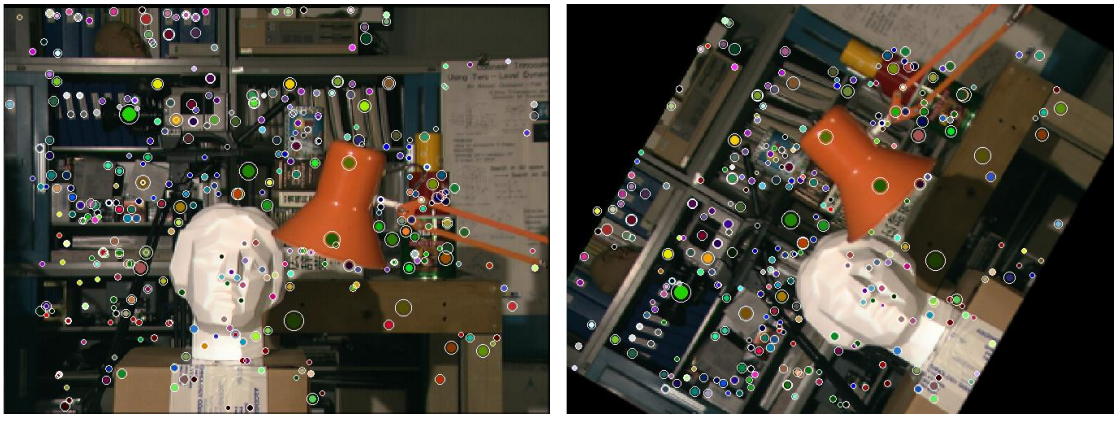
Figura 2.17: Criação do descritor por meio dos histogramas de orientação. As componentes dos histogramas (à esquerda) calculados em setores em torno do ponto-chave são distribuídos em componentes de um vetor, que ao ser normalizado, visando invariância à mudanças na intensidade, configura descritor do ponto-chave (adaptado de [10]).

2.6 CONFRONTO E CASAMENTO DE CORRESPONDÊNCIAS

Realizou-se o confronto entre imagens submetidas a transformações descritas anteriormente: em escala (Figura 2.18); em rotação (Figura 2.19) e; em intensidade luminosa (Figura 2.20). Os pontos com produto interno entre descritores que assumem entre si seu maior valor, são considerados correspondências, sendo ilustrados por um padrão de cor e dimensão proporcional à magnitude da ponto-chave.

(a) (b)

Figura 2.18: Confronto entre duas vistas distintas de uma cena, uma delas submetida a uma transformação na escala: (a) vista em seu aspecto original e; (b) vista com dimensões reduzidas pela metade, mostrada no aspecto original para melhor visualização. Os pontos correspondentes entre as duas imagens são marcados por círculos com um mesmo padrão de cor e tamanho. Devido à grande quantidade de pontos, somente são exibidos correspondências com produto interno superior a $0,9$.



(a) (b)

Figura 2.19: Confronto entre duas vistas distintas de uma cena, uma delas submetida a uma rotação: (a) vista em seu aspecto original e; (b) vista rotacionado em 60^o no sentido anti-horário. Os pontos correspondentes entre as duas imagens são marcados por círculos com um mesmo padrão de cor e tamanho. Devido à grande quantidade de pontos, somente são exibidos correspondências com produto interno superior a $0,9$.

(a) (b)

Figura 2.20: Confronto entre duas vistas distintas de uma cena, uma delas submetida a uma transformação na iluminação: (a) vista em seu aspecto original e; (b) vista com magnitude reduzida em $\sqrt{2}\times$. Os pontos correspondentes entre as duas imagens são marcados por círculos com um mesmo padrão de cor e tamanho. Devido à grande quantidade de pontos, somente são exibidos correspondências com produto interno superior a $0,9$.

3 GRAFOS E SEGMENTAÇÃO DE IMAGENS

3.1 INTRODUÇÃO

São definidas neste capítulo as propriedades da teoria dos grafos e de como elas se aplicam em processamento de imagem. A teoria dos grafos tem uma ampla lista de definições atribuídas a partir de suas representações gráficas. Este trabalho abordará apenas os aspectos básicos, suficientes para desenvolver e compreender os métodos propostos. Ao final do Capítulo serão apresentadas as técnicas de segmentação empregadas neste trabalho, *watershed* e SLIC na formação de agrupamentos e *GrowCut* na segmentação de objetos nos vídeos via grafos.

3.2 GRAFOS

A definição da palavra grafo é oriunda da capacidade dessa representação poder ilustrar problemas graficamente, permitindo uma visualização mais intuitiva para soluções de problemas. Isso reflete situações em que se torna conveniente tratar e descrever um problema por meio de um conjunto de pontos conectados por setas ou linhas, sejam esses pontos, pessoas, lugares, átomos, moléculas, etc. Frisando-se que o importante nesse tipo de análise não é a posição dos pontos, há várias maneiras de se desenhar um grafo, no entanto as relações estabelecidas entre os elementos que o constituem, muitas vezes rotulada com algum tipo de ponderação, são preservadas [23].

3.2.1 Histórico

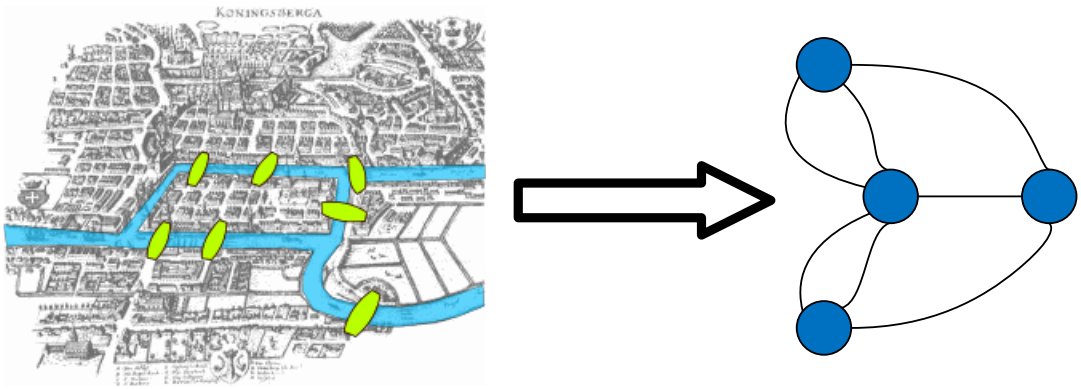


Figura 3.1: Ilustração das pontes de Königsbert sua representação por grafos. O problema envolvendo as pontes de Königsbert consistia em determinar se seria possível atravessar pelas sete pontes, sem repeti-las, e retornar ao ponto inicial (adaptado de [24]).

A origem da utilização dos grafos remete ao ano de 1736, quando Euler provou não solucionável o clássico problemas das pontes de Königsbert. Esse problema retrata duas ilhas conectadas por sete pontes, para as quais se questiona a possibilidade de se atravessar pelas sete, uma única vez em cada, e retornar ao ponto inicial. Dentro de todas as soluções empíricas negativas, Euler generalizou o problema conectando as origens e destinos pontos por linhas (Figura 3.1), um grafo, e mostrando que tal grafo não pode ser cruzado de certas maneiras, provando que o problema não possuía solução [25].

No campo da eletricidade, em 1887, Kirchhoff começou a substituir os elementos de um circuito, resistências, indutores, capacitores, por conexões de pontos feitas por linhas, possibilitando a postulação de um teorema para a análise de sistemas de equações lineares [25]. No século XX, Richard Feynman levou a outra dimensão o mesmo conceito de representação gráfica de problemas (Figura 3.2), definindo soluções de equações em eletrodinâmica quântica por meio de diagramas [26].

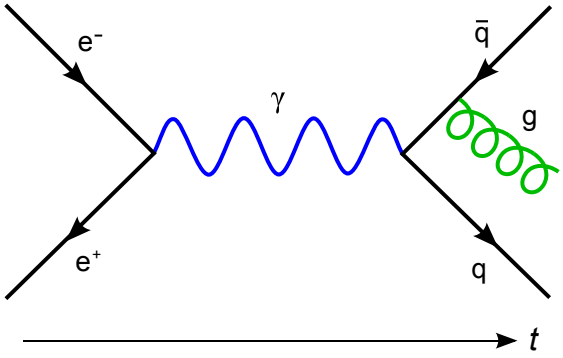


Figura 3.2: Exemplo de um diagrama de Feynman utilizado para a resolução de problemas em eletrodinâmica quântica (retirado de [27]).

Atualmente, o processamento com grafos tem envolvido principalmente grandes volumes de dados, nas quais simplifica-se a relação entre elementos por seu nível de conexão. A análise por grafos diminui a velocidade de processamento em relação a bancos de dados relacionais, onde espera-se que os grafos forneçam uma simplificação para uma estrutura complexa, agilize e flexibilize a solução de problemas [28].

As análises de redes sociais, como Youtube, Facebook e Instragram, também podem ser feitas graficamente, uma vez que as pessoas podem ser relacionadas por meio de conexões [29, 30]. Mais de meio século antes da criação das redes sociais, o escritor Húngaro Frigyes Karinthy, postulou em um dos seus romances uma conjectura de que dois indivíduos no mundo estariam conectados por no máximo 5 conhecidos. Essa teoria hoje pode ser testada e aplicada nos dados de redes sociais, analisadas via teoria dos grafos [31], remetendo-se a um número de conexões menor que o da conjectura.

A teoria dos grafos se faz presente na análise de trajetos mais curtos ou mais rápidos dentro de uma rede de vias automotivas [32, 33], que ligam dois pontos. Dados em estrutura topológica são uma das formas nas quais também se pode representar imagens e vídeos, descrevendo-as como uma série de elementos com conexões definidas por uma matriz adjacente, ou por uma matriz de pesos a qual remete em seus elementos a conexão e a força de conexão entre elementos [34].

3.2.2 Definições e notações

Em geral, a termologia utilizada pelos autores em teoria dos grafos é personalizada, tornando essencial uma boa definição dos conceitos e notações para um bom entendimento do trabalho. Este trabalho não possui um alto nível de complexidade de análise em grafos, fazendo uso das propriedades básicas. Definições quanto à morfologia de um grafo não serão abordadas, as definições apresentadas aqui são inspiradas em [24].

Definição 1 (Grafos) *Um grafo* $G = (V, E)$ *consiste em um par, no qual* V *e* E *são ambos conjuntos finitos de elementos. Os elementos* $v \in V$ *são chamados vértices (nós) e os elementos* $e \in E \subset \{\{i, j\}, i, j \in V, i \neq j\}$ *são chamados de arestas*

A Figura 3.3 representa o grafo do problema das pontes de Königsbert. As pontes em duplicidade foram agrupadas em uma única aresta. As arestas são rotuladas de acordo com seu endereçamento, segundo os vértices que são conectados por elas. Essa relação de vizinhança é chamada de adjacência.

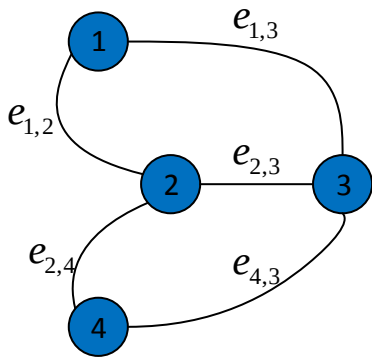
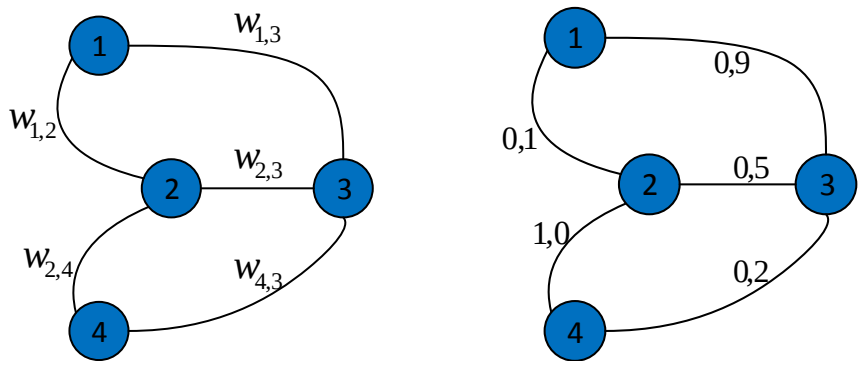


Figura 3.3: Representação do problema de Kirchhoff por um grafo com seus vértices 1, 2, 3 e 4 representados por círculos azuis, conectados por suas respectivas arestas que que ilustram as relações de adjacência do grafo.

Definição 2 (Adjacência) *Se uma aresta* $e_{i,j}$ *conecta* i *a* j*, então estes são adjacentes, ou seja,* i *é vizinho de* j *(e vice-versa)*

As adjacências podem ainda ser caracterizadas por forças de ligação, ou seja, define-se valores de pesos para a relação entre um vértice e outro, formando um grafo ponderado.

Definição 3 (Grafo Ponderado) *No caso de um triplete* $G = (V, E, \mathbf{W})$ *(grafo ponderado),* $w_{i,j} \in W$ *determina a força de ligação de uma aresta* $e_{i,j}$



(a) (b)

Figura 3.4: Representação de um grafo ponderado: (a) as arestas que ligam vértices recebem pesos associados a elas; (b) os valores desse pesos determinam uma força de ligação para aquele par de vértices.

As adjacências e pesos de um grafo podem ser ainda representadas por uma matriz, ou mapa de pesos.

Definição 4 (Mapa de Pesos) *O mapa* $\mathbf{W}$ *é chamado mapa de pesos de* $G = (V, E, \mathbf{W})$. *O mapa* $\mathbf{W}^n$ *denotará o mapeamento do conjunto de arestas* E *de forma que o peso da aresta* $e_{i,j}$ *é igual a* $w^n_{i,j}$

A quantidade de elementos na matriz de ponderação W é determinada pelo número de vértices N do grafo, determinando N^2 relações. No exemplo da Figura 3.4 temos:

$$\mathbf{W} = \begin{bmatrix} 0 & 0,1 & 0,9 & 0 \\ 0,1 & 0 & 0,5 & 1 \\ 0,9 & 0,5 & 0 & 0,2 \\ 0 & 1 & 0,2 & 0 \end{bmatrix}. \tag{3.1}$$

O ordenamento das linhas representa os índices dos vértices, e as colunas as incidências desses vértices, por exemplo, o elemento da segunda linha e primeira coluna, retrata a força de ligação do vértice 2 com o vértice 1, que vale 0,1. Este trabalho utiliza apenas com grafos unidirecionais, ou seja, os elementos da matriz de peso são estritamente positivos, determinando uma relação de simetria $w_{i,j} = w_{j,i}$.

O principal objetivo deste trabalho é selecionar subconjuntos de um grafo que melhor representem o objeto de interesse, uma segmentação. Esse subconjunto é denominado subgrafo.

Definição 5 (Subgrafo) *Dado um grafo* $G = (V, E)$, *o grafo* $G' = (V', E')$ *é chamado de subgrafo de G se e somente se* $V' \subset V$ *e* $E' \subset E$.

3.3 IMAGENS REPRESENTADAS COMO GRAFOS

A representação mais básica de uma imagem consiste no registro de uma matriz contendo dados sobre a magnitude de um pixel, sua intensidade e/ou componentes de cores (Figura 3.5 de (a) a (c)). Esse tipo de registro também é saída de filtragens e processamentos, como suavizações e realce de bordas [34].

Um segundo de nível de registro e representação é feito a partir do agrupamento de pixels em regiões (superpixels), essas regiões representam áreas da imagem com características muito semelhantes de cor, textura, posição, e, possivelmente, pertencentes a um mesmo objeto (Figura 3.5 de (d) a (f)). As duas formas de representação podem ser analisadas a partir das suas vizinhanças, as matrizes de pixels por meio dos grafos de pixels e as regiões por meio dos grafos de regiões.

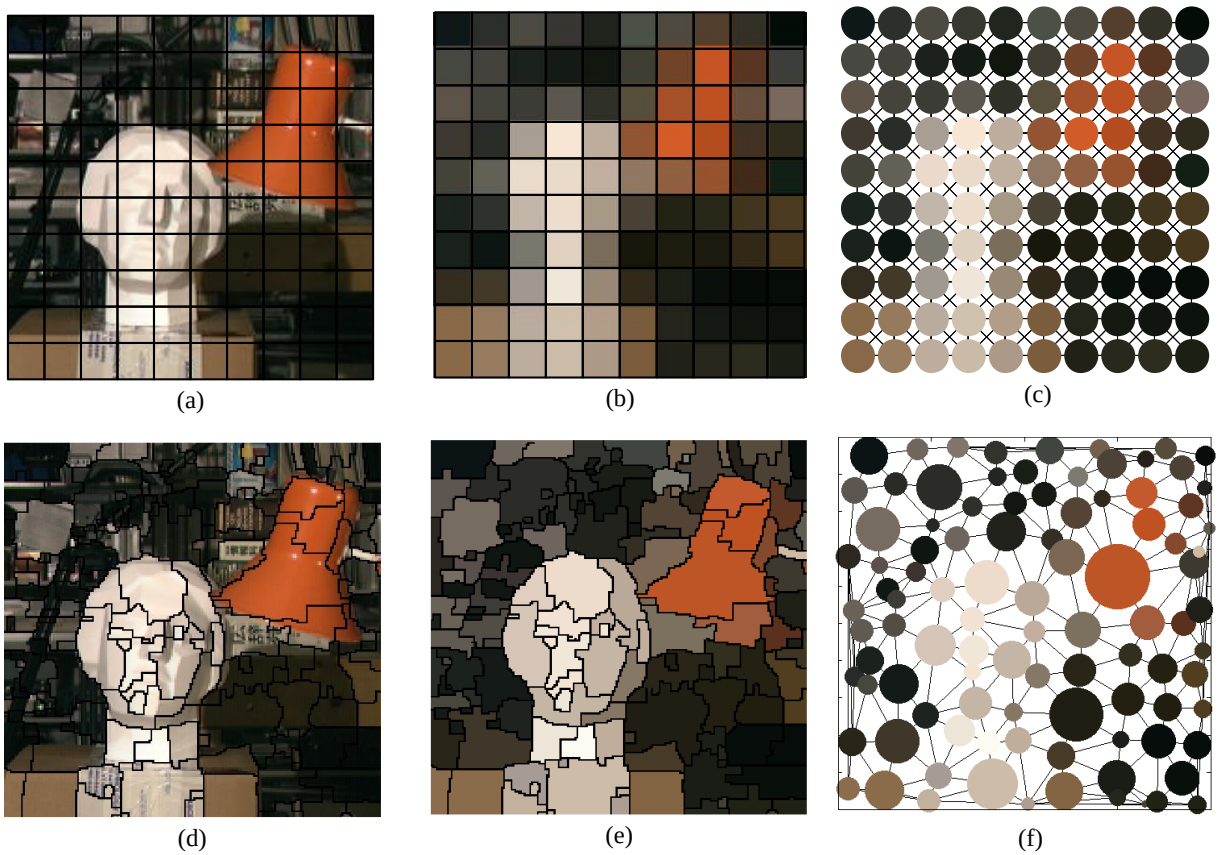


Figura 3.5: Representações de um modelo real em pixels e em regiões: (a) representação por meio de pixels, (b) que são amostras uniformemente distribuídas da cena; (c) em que seu grafo de pixels tem relações de vizinhança que podem ser trivialmente definidas, dada a organização dos dados; (d) o modelo real pode ser representado por regiões; (e) que também se expressam em amostras do modelo real; (f) entretanto, as relações de vizinhança para o respectivo grafo de regiões não são simplesmente definidas, mas tal aproximação supera a versão em pixels pelo grau de representatividade do modelo real com um número similar elementos.

3.3.1 Grafo de pixels

Uma imagem digital $F(x, y)$ determina um tipo especial de grafo (grafo de adjacência de pixels ou grafo de pixels), com vizinhança bem definida por uma grade retangular. Os vértices v desse grafo são representados pelos pixels, as relações de pesos entre os vértices, pixels, adjacentes é, em geral, determinadas pela proximidade espacial e de intensidade luminosa (Figura 3.5(c)).

Métodos de segmentação, como *watershed*, utilizam as propriedades da teoria dos grafos para isolar regiões a partir de um gradiente da imagem original. Outros métodos também se valem da relação entre vizinhança, como *mean shift* e *k-means* [35], com uma atualização constante de pesos em um processo iterativo que busca a minimização da distância entre os elementos e seus agrupamentos. Esses métodos de agrupamento citados anteriormente, são usualmente aplicados na criação de regiões em uma imagem.

3.3.2 Grafo de regiões

A partir de um método de agrupamento aplicado a uma imagem de pixels (sub-grafo de pixels), cada um desses agrupamentos pode ser substituído e representado por uma região homogênea (superpixel), promovendo uma segmentação em baixo nível [24]. Ao contrário de uma imagem de pixels, dispostas em grade bem definida, as uma imagem constituída por regiões possuem, em geral, vizinhanças com relações não triviais. Essas relações podem ser representadas por um grafo de adjacência de regiões (Figura 3.5(f)).

3.4 SEGMENTAÇÃO DE IMAGENS

O processo de segmentação consiste em subdividir uma imagem em regiões ou objetos que a constituem. Seu objetivo é simplificar ou alterar a representação de uma imagem, com a finalidade de facilitar sua análise. Para isso, existem diversos métodos capazes de realizar tal função, nos quais se destacam técnicas baseadas em similaridade (threshold), detecção de descontinuidades, agrupamento de dados (*clustering*). Neste trabalho, serão retratados aqueles processos utilizados nos métodos propostos, as técnicas de agrupamento *watershed* e SLIC que realizam, em conjunto, a formação de regiões ao longo de quadros de vídeos, e a técnica de corte de grafos *GrowCut* para a segmentação do objeto.

3.4.1 Técnicas de agrupamento

Métodos de agrupamento permitem que se extraiam características determinadas de um grupo de dados, separando-os em subgrupos funcionais ou hierarquizando os dados para algum tipo de análise posterior. As técnicas de agrupamento ou análise de agrupamento é o nome dado para o grupo de técnicas computacionais cujo propósito consiste em separar determinados dados pertencentes a um grupo específico, baseando-se nas características que estes dados possuem. A ideia básica consiste em colocar em um mesmo grupo objetos que sejam similares de acordo com algum critério pré-determinado.

O critério de determinação de agrupamento, normalmente, baseia-se em uma função de dissimilaridade. Tal função recebe dois objetos e retorna a distância entre eles. Os grupos determinados por uma métrica de qualidade devem apresentar alta homogeneidade interna e alta separação (heterogeneidade externa). Isto quer dizer que os elementos de um determinado conjunto devem ser mutuamente similares e, preferencialmente, muito diferentes dos elementos de outros conjuntos.

3.4.1.1 Watershed

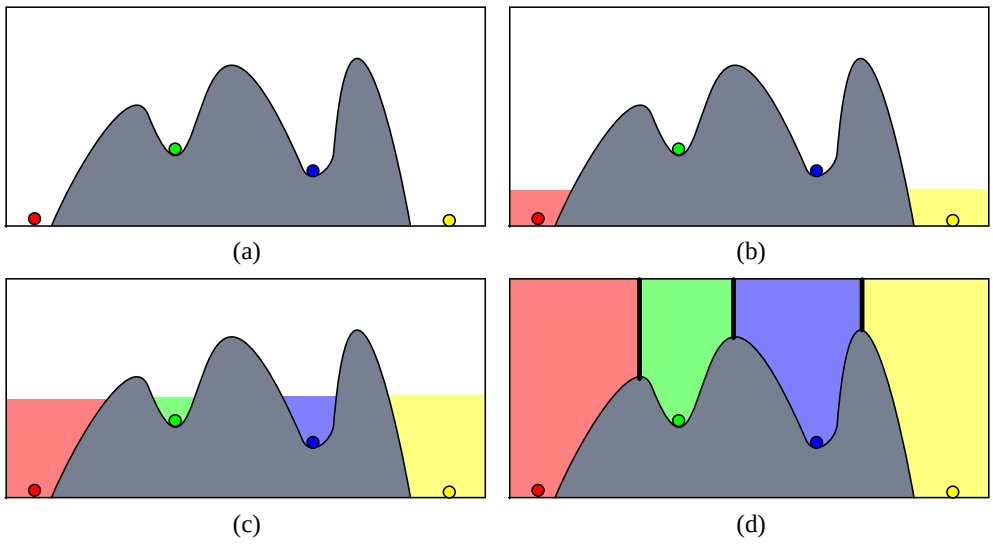


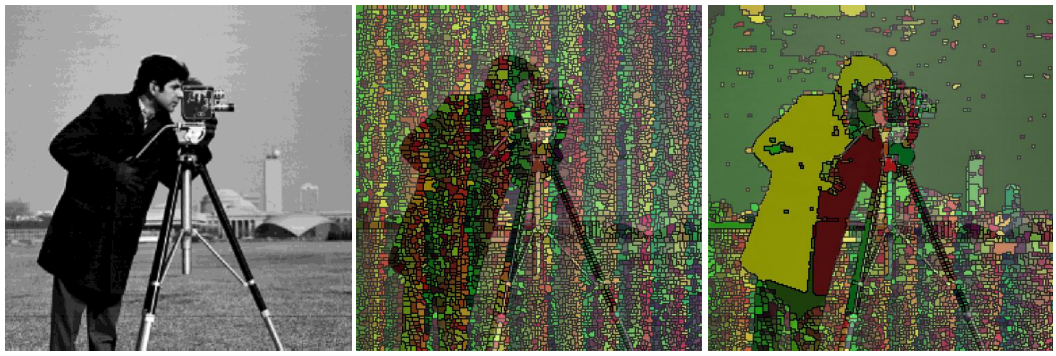
Figura 3.6: Ilustração do processo de agrupamento de regiões pelo algoritmo *watershed*: (a) o mapa de magnitude de uma imagem representa um mapa topológico, no qual os mínimos locais da imagem são estão nas áreas de depressão desse mapa topológico; (b) a partir dos mínimos locais, as depressões começam a ser preenchidas, formando lagos; (c) os lagos crescem de forma a terem uma mesma altura; (d) no ponto de encontro dos lagos, as represas que dividem as regiões são definidas (inspirado em [24]).

A técnica de agrupamentos *watershed*, palavra traduzida para o português como bacia hidrográfica, tem o funcionamento intuitivamente correspondente a sua nomenclatura. Pode-se imaginar os mapas de magnitude de uma imagem como acidentes geográficos, em que os mínimos

locais estão localizados nas depressões. Essas depressões, indicadas em ilustração pelos círculos em diferentes pontos na Figura 3.6(a), são inundadas simultaneamente (Figura 3.6(b) e (c)) de forma que as alturas dos lagos formados nessas depressões estejam sempre niveladas. Regiões são tratadas como um agrupamento, quando os lagos se tocam e uma barragem é construída para definir essa fronteira [24].

Os princípios básicos da *watershed* desencadeiam uma série de algoritmos que possibilitam o tipo de solução desejada. Entre esses algoritmos, destacam-se os métodos de Vicent-Soille, Meyer e técnicas de custo e topológicas, todos envolvendo princípios de teoria dos grafos, uma vez que utilizam as relações entre vizinhanças para a construção de regiões [36]. A *watershed* aplicada ao longo deste trabalho, utiliza o algoritmo de Fernand Meyer [37].

A *watershed* pode ter inicialização feita por marcadores manualmente definidos, entretanto, regiões agrupadas a partir de mínimos locais definidos por um mapa de variações espaciais (gradiente) são usualmente utilizados, por representarem pontos singulares, especiais dentro de uma imagem. Os mínimos locais são evidenciados nos mapas de gradiente da imagem original (Figura 3.7(a)), esse gradiente pode ser obtido ou por meio de aproximações do Laplaciano ou por filtros detectores de borda e é a imagem, em geral, utilizada para a aplicação da *watershed*.



(a) original (b) T = 1 → N = 5843 (c) T = 2 → N = 2494

(d) T = 3 → N = 1748 (e) T = 4 → N = 1111 (f) T = 5 → N = 646

Figura 3.7: *Watershed* aplicada no gradiente de uma (a) imagem fornece diferentes números de regiões N quando um limiar T é definido para esse mapa de gradiente. T é um valor mínimo de variação aceito dentro de um gradiente que diminui o número de mínimos locais, utilizados para inicialização do agrupamento de regiões.

Um limiar mínimo T para as variações detectadas podem diminuir a quantidade de mínimos locais e, por consequência, o número de regiões N formadas a partir da inundação desses platôs (Figura 3.7(b) a (e)).

3.4.1.2 *K-means*

O método de agrupamento *k-means* possui paradigma de aprendizado não-supervisionado, ou seja, procura determinar e identificar automaticamente como os dados estão organizados em um conjunto ou em uma base de dados. O *k-means* é um método de agrupamento que tem como objetivo encontrar k grupos (padrões, ou regiões) na imagem [35].

Estes grupos são representados por centroides, que são médias numéricas de todos os pixels pertencentes ao agrupamento em questão. Deve-se escolher k centroides iniciais, que representam os centros dos k agrupamentos dados por C_1, C_2,..., C_k (Figura 3.8(a)). Cada pixel da imagem é rotulado em relação ao centroide da classe mais similar. Por conseguinte, os centroides têm seus valores atualizados com base nos pixels que passaram a pertencer aos respectivos agrupamentos. Assim, o processo se repete enquanto um critério de parada não for satisfeito (Figura 3.8(b)).

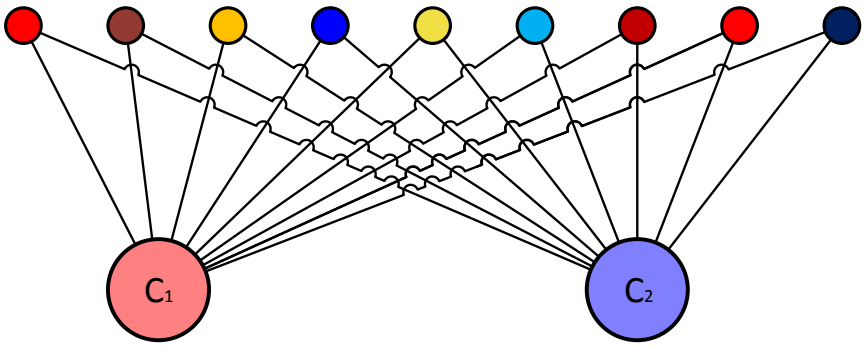
A heurística de agrupamento não hierárquico do *k-means* busca minimizar a distância dos elementos a um conjunto de k centros dado por $X = \{C_1, C_2, ..., C_k\}$ de forma iterativa. Cada centroide C_k é um conjunto de médias que, em geral, tem associado a ele características de posição espacial ou de cor do agrupamento, $C_k = \{\vec{\mathbf{r}}_k, \vec{\mathbf{I}}_k\}$. $\vec{\mathbf{r}}_k$ define a posição espacial média do agrupamento k e $\vec{\mathbf{I}}_k$ o vetor de cores médio desse centroide.

Ao se medir a distância d de conjunto de características de um elemento $p_i = \{\vec{\mathbf{r}}_i, \vec{\mathbf{I}}_i\}$ a um agrupamento representado por um centroide C_k:

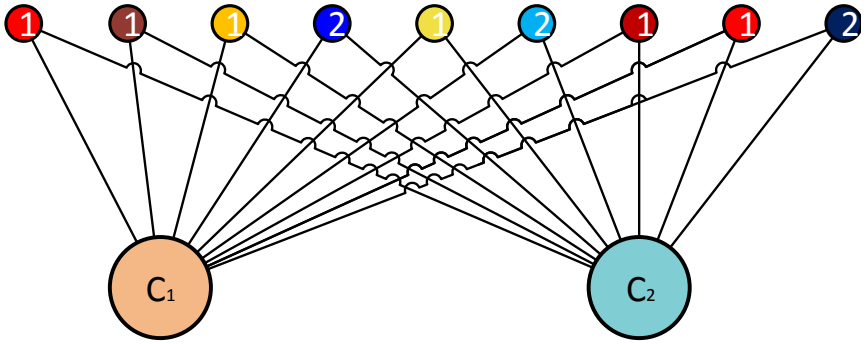
$$k_w = \arg\min_k D(p_i, C_k), \tag{3.2}$$

deseja-se encontrar o centroide C_{k_w} que está mais próximo do elemento i. A cada iteração o valor das propriedades do centroide C_k, posição e cor, são atualizadas com a média dos elementos agrupados por ele.

O algoritmo cessa iterações assim que todos os elementos estão agrupados aos centroides mais próximos a eles. Há uma convergência relativamente rápida para uma solução de equilíbrio, aquela que minimize todas as distâncias dos elementos aos seus respectivos centroides. Essa solução depende dos valores com os quais centroides são inicializados, em geral, determinados de forma aleatória.



(a)



(b)

Figura 3.8: A técnica de agrupamento *k-means* busca relacionar elementos em com centroides, comparando a distância entre si desses elementos e os centroides: (a) os centroides, C_1 e C_2 no exemplo, são inicializados de acordo com a necessidade do usuário e os elementos a serem agrupados podem ou não ter rotulação prévia; (b) a cada iteração os elementos vão sendo rotulados de acordo com sua proximidade com os centroides, os quais vão alterando seus valores com base nos elementos que agrupam.

3.4.1.3 SLIC

A técnica de agrupamento SLIC (agrupamento iterativo linear simples, do inglês *simple linear iterative clustering*) é uma adaptação do algoritmo *k-means* [15]. Os mesmos procedimentos são adotados da forma que descrita anteriormente para o *k-means*, modificando-se a área de atuação de cada centroide. Impõe-se uma restrição, geralmente espacial para imagens, de forma que as distâncias a um centroide são calculadas para um grupo menor de elementos, não todo o conjunto em análise.

Na Figura 3.8, que exemplifica o algoritmo *k-means*, tem-se dois centroides, C_1 e $C2$, competindo por todos os elementos de um conjunto o qual deseja-se separar em grupos. Podemos expandir esse raciocínio para elementos distribuídos em um plano 2-D, como na Figura 3.9(a), na qual todos os elementos sem rótulos são disputados por todos os 4 centroides, C_1, C_2, C_3 e C_4.

Na sua versão adaptada, os centroides disputam os elementos de acordo com áreas de atuação pré-definidas (Figura 3.9(b)). O objetivo dessa restrição na quantidade de elementos em disputa para cada centroide é a redução do esforço computacional.

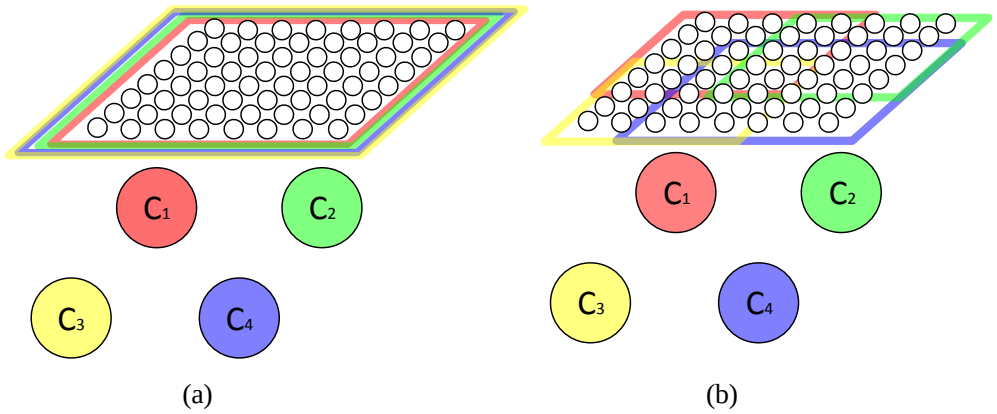


Figura 3.9: Ilustração comparando os algoritmos *k-means* e SLIC: (a) no algoritmo *k-means* os centroides disputam todos os elementos em análise entre si, uma área de atuação ampla, representada pelos quadrados, com arestas seguindo o padrão de cores de seus respectivos centroides; (b) para o SLIC, essa análise fica restrita a áreas pré-definidas, que se sobrepõem, novamente ilustrando casamento das cores das arestas dos quadrados e os centroides que atuam na sua respectiva área.

Da mesma forma descrita para o algoritmo *k-means*, C_k é um conjunto de médias do agrupamento k, contendo informações quanto à posição média $\vec{\mathbf{r}}_k$ desse grupo e seu vetor de cores médio $\vec{\mathbf{I}}_k$, o conjunto p_i do elemento i também tem os mesmos vetores associados a ele, $\vec{\mathbf{r}}_i$ e $\vec{\mathbf{I}}_i$. Em [15], a distância $D_{i,k}$ entre um agrupamento k a um elemento i, é uma combinação da distância espacial:

$$ds_{i,k} = ||\vec{\mathbf{r}}_i - \vec{\mathbf{r}}_k|| \tag{3.3}$$

e a distância entre as cores:

$$dc_{i,k} = ||\vec{\mathbf{I}}_i - \vec{\mathbf{I}}_k||, \tag{3.4}$$

em que $||\cdot||$ é a norma euclidiana entre os vetores.

Ao se normalizar a distância espacial $ds_{i,k}$ entre centroide e elemento por uma constante S, que é a dimensão da aresta da região de atuação dos centroides (Figura 3.9(b)), e se inserir um fator de peso m, define-se a distância $D_{i,k}$ para o SLIC como:

$$D_{i,k} = \sqrt{dc_{i,k}^2 + \left(\frac{ds_{i,k}}{S}\right)^2 m^2}. \tag{3.5}$$

A dimensão S da região de disputa do agrupamento, normaliza a distância espacial $ds_{i,k}$ entre o centroide e o elemento em análise, de forma que os valores desta distância para pontos extremos nesta região de disputa fiquem próximos a 1.

O parâmetro m é um peso determinado pelo usuário que define a maior grau de importância para a distância espacial ou para a distância entre cores, no processo de agrupamento, quanto maior o valor de m maior peso é dado a $ds_{i,k}$. A quantidade de agrupamentos ainda é definida pelo usuário, assim como no *k-means* (Figura 3.10).

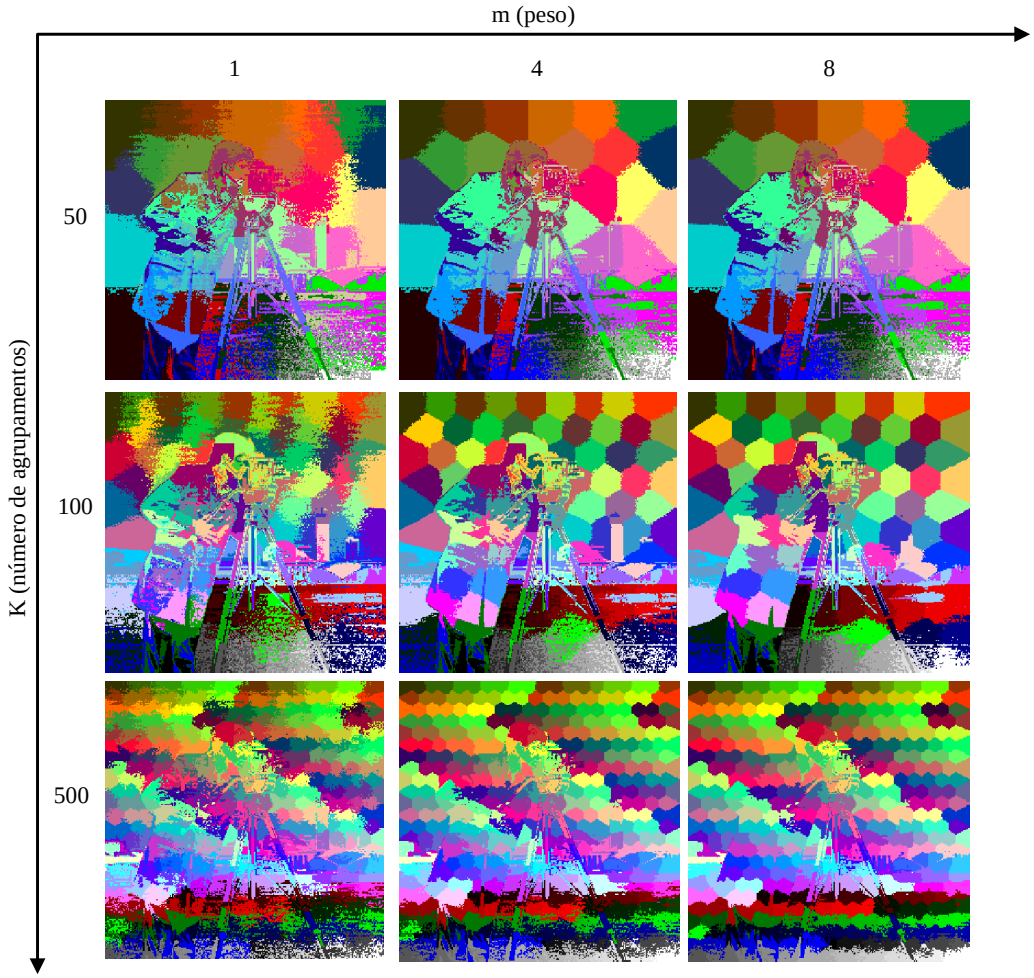


Figura 3.10: Algoritmo de agrupamento SLIC aplicado na imagem da Figura 3.7(a), para diferentes níveis de número de agrupamentos K e peso m. O aumento de m implica no aumento da importância da distância espacial no agrupamento, percebe-se para regiões como alto grau de semelhança, como o céu da imagem, a uniformidade no formato das regiões para os valores mais altos de m. A definição no número de agrupamentos depende do grau de representatividade desejado pelo usuário, quanto menor o número de agrupamentos, maior serão as áreas e os erros envolvendo suas fronteiras.

3.4.2 Cortes em grafos

A segmentação de imagens via cortes consiste na desassociação de um grafo V em dois subconjuntos, A e B, em que $A \cup B = V$ e $A \cap B = \emptyset$. O subconjunto A pode representar um objeto de interesse enquanto B representa o fundo do qual deseja-se separar o objeto.

Em geral, um corte em grafos tem como objetivo a determinação dos subconjuntos A e B que minimizem uma função de custos:

$$C(A, B) = \sum_{i \in A} \sum_{j \in B} w_{i,j}. \tag{3.6}$$

A função de custos C é chamada também de capacitância, e leva em consideração a forma na qual os conjuntos A e B estão conectada, somando o peso de todas as suas conexões $w_{i,j}$. Quanto menor o valor de $C(A, B)$ menor o custo de se desassociar um grafo nos dois subconjuntos A e B.

3.4.2.1 Min cut/Max flow

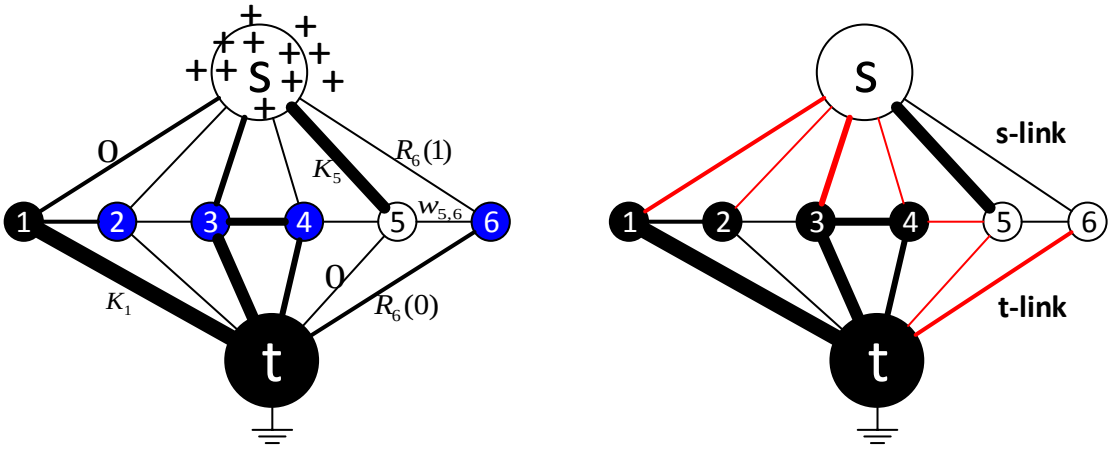
Definindo $\mathbf{a} = (a_1, a_2, ..., a_N)$ como um vetor binário cujas componentes especificam se um elemento p de uma imagem pertence ao conjunto A ($a_p = 1$) ou ao conjunto B ($a_p = 0$), dentro dos N nós do grafo, a ref. [38] propõe uma função de energia para minimização e corte dada como:

$$E(\mathbf{a}) = \lambda R(\mathbf{a}) + C(\mathbf{a}), \tag{3.7}$$

em que a função de capacitância $C(\mathbf{a})$, que exprime as relações de fronteira do objeto, é combinada a uma função de custos $R(\mathbf{a})$, que mede o custo de se tomar um elemento p como parte do objeto ou do fundo, que pode independer das relações de fronteira, podendo estar relacionado aos níveis de magnitude de uma região ou pixel, por exemplo. A combinação entre as duas funções é mediada pela constante λ, quanto menor o seu valor, mais peso se dá as relações de fronteira do objeto.

Um novo grafo é construído com dois nós a mais (Figura 3.11), um nó chamando de *source* (s, fonte) e o outro chamado de *sink* (t, pia), o nó s representa o conjunto A (objeto) e o t o conjunto B (fundo). As conexões desses novos vértices aos nós do grafo é independente, recebendo como peso as relações da função de custos R para nós sem definição de rótulo (se pertencem ao conjunto A ou ao B), um peso K muito alto para aqueles elementos previamente rotulados de acordo com o seu conjunto e um peso nulo caso contrário (Figura 3.11(a)).

Os algoritmos de *maximum flow/minimum cut* se baseiam no teorema de grafos que define que o fluxo (*flow*) entre os vértices s e t é máximo para o corte mínimo (*minimum cut*). Pode-se fazer uma analogia a um circuito elétrico, no qual o nó s é a fonte de energia e o t é a referência para qual as cargas fluem (Figura 3.11(a)). Os pesos para as conexões são as admitâncias, quanto maior o peso, menor a resistência quanto a passagem de corrente. O conjunto de arestas/conexões na qual a corrente satura é o conjunto de conexões na qual a admitância é mínima (Figura 3.11(b)).



(a) (b)

Figura 3.11: No corte de grafos s/t, geralmente utilizado nos algoritmos de *max flow/min cut*, cria-se dois vértices a mais para análise, s (*source*) e t (*sink*): (a) os pesos para as arestas desses novos vértices são determinados por um K, de valor alto, caso o rótulo do vértice seja o mesmo do novo nó, na ilustração branco para pertencente ao s e preto para t, os nós i destinados à segmentação (em azul), recebem uma ligação $R_i(0/1)$ que representam um risco de se tomá-los como parte do objeto (1) ou do fundo (0); (b) na análise ou do fluxo máximo ou do corte mínimo, procura-se um conjunto de arestas (ligações em vermelho) para se desvincular um conjunto do outro a um custo baixo.

3.4.2.2 Normalized cut

Um dos problemas com os algoritmos de mínimo corte é que eles favorecem a segmentação de pequenos conjuntos de nós isolados. Por exemplo, uma vez que o peso de ligação de um elemento de cor muito distinta a sua vizinhança é baixo, um corte mínimo pode isolar apenas esse elemento do restante do grafo, visto que esse fornece uma capacitância pequena. Resolve-se tal deficiência, quando se analisa a desassociação de dois conjuntos A e B de um grafo V pela fração que esses dois conjuntos representam no grafo, nas condições anteriormente definidas, determinando um corte normalizado [39].

Define-se um grau de conectividade d_i do elemento i com o grafo V (Figura 3.12(a)), ao se somar todos os pesos de conexão que esse elemento faz com o grafo:

$$d_i = \sum_{j \in V} w_{i,j}. \tag{3.8}$$

Esse grau de conectividade pode ser estendido a um conjunto A (Figura 3.12(b)), somando-se todos os valores de d_i para $i \in A$:

$$assoc(A, V) = \sum_{i \in A} d_i. \tag{3.9}$$

Com o corte $cut(A, B)$ entre os dois conjuntos A e B, que é a função de capacitância da equa-

ção (3.6), [39] constrói a função $Ncut(A, B)$ utilizada para minimização:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)}. \tag{3.10}$$

Recorrendo novamente ao vetor de rótulos **a** que define quais nós pertencem ao conjunto A e quais pertencem ao B, o corte normalizado que minimiza $Ncut(\mathbf{a})$ pode ser solucionado ao se resolver um problema de autovalores λ e autovetores **u** [39]:

$$\mathbf{D}^{\frac{1}{2}}(\mathbf{D} - \mathbf{W})\mathbf{D}^{\frac{1}{2}}\mathbf{u} = \lambda\mathbf{u}, \tag{3.11}$$

em que **D** é a matriz diagonal do grafo V, em que d_i se torna o elemento da posição $\{i, i\}$ dessa matriz. Por exemplo, a matriz diagonal referente ao grafo da Figura 3.4 com mapa de pesos da equação (3.1) é dada como:

$$\mathbf{D} = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 1,6 & 0 & 0 \\ 0 & 0 & 1,6 & 0 \\ 0 & 0 & 0 & 1,2 \end{bmatrix}, \tag{3.12}$$

As N (número de nós) soluções linearmente independentes para o vetor **u** retornam valores contínuos. Discretizando esse vetor em dois níveis, 0 ou 1, obtém-se uma segmentação para o grafo, uma aproximação para o **a** que minimiza $Ncut(\mathbf{a})$.

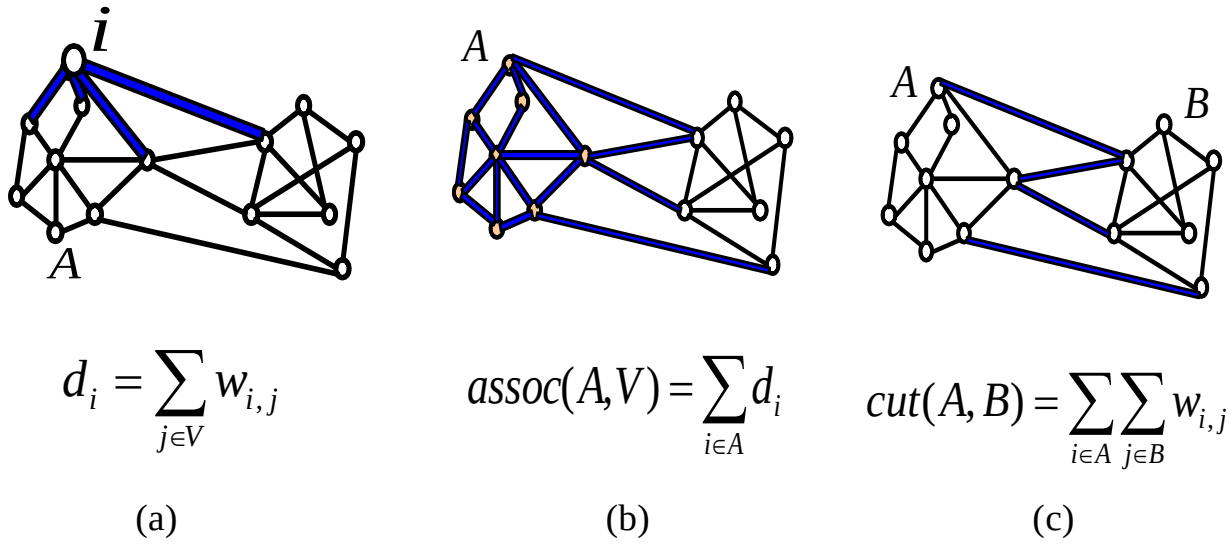
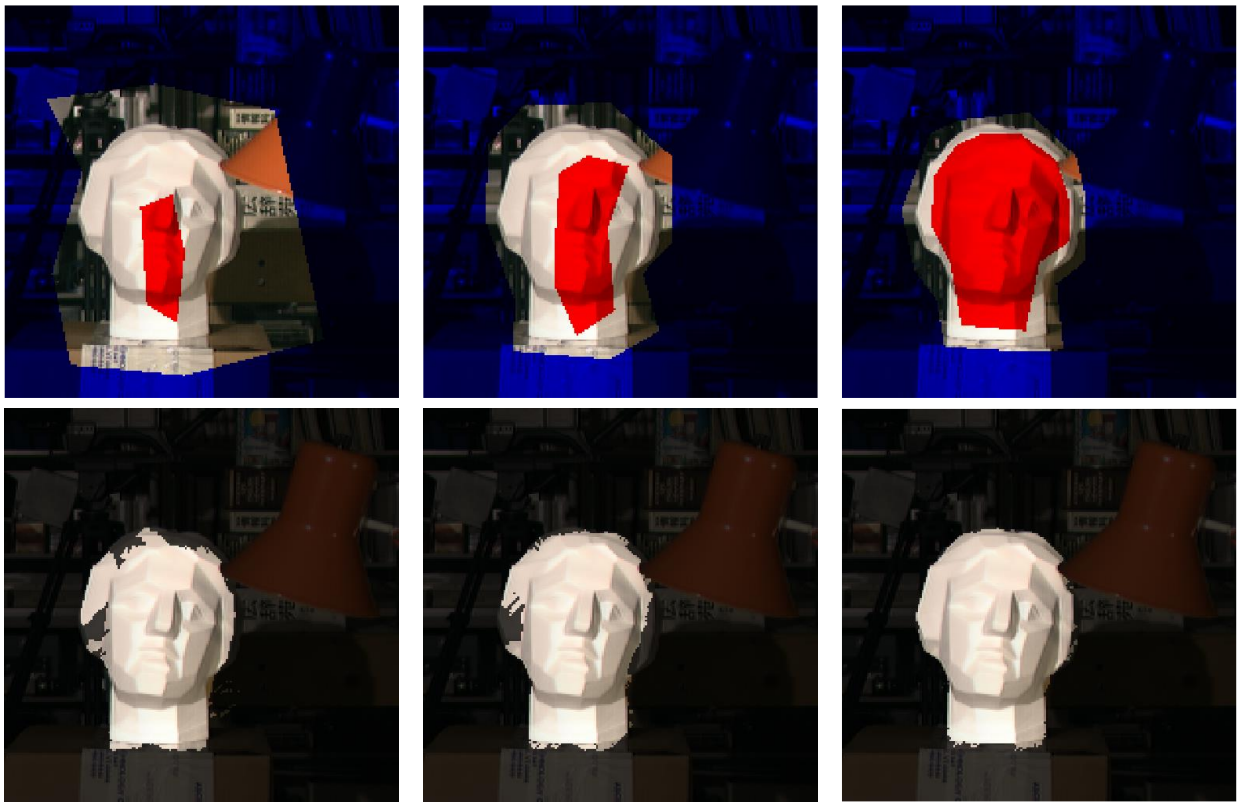


Figura 3.12: Expressões utilizadas no corte normalizado e suas respectivas representações nos grafos: (a) a soma de todas dos pesos de todas as arestas de um vértice i fornece um nível de conectividade d_i desse ao grafo; (b) a ideia pode ser estendida para um subgrafo A, medindo-se o grau de associação desse subconjunto ao grafo V, $assoc(A, V)$, pelo somatório de d_i com $i \in A$; (c) ponderando e combinando um corte $cut(A, B)$ pelo grau de associação de A e B com o grafo V , obtém-se a função $Ncut(A, B)$, utilizada para o corte normalizado.

3.4.2.3 GrowCut

A técnica de cortes em grafos *GrowCut* proposto na ref. [1], tem princípios muito semelhantes aos do SLIC. A competição para rotulação dos elementos ocorre também mediante uma restrição espacial, somente vizinhos próximos a cada elemento são analisados. Diferentemente do SLIC, no *GrowCut* são formados apenas dois agrupamentos, objeto e fundo, e a competição por um elemento ocorre entre os próprios nós do grafo (pixels no caso de imagens), não entre centroides.

Definindo-se valores de um nó i para o vetor de rótulos **a** como $a_i = 1$, caso o elemento pertença ao objeto, $a_i = -1$, caso pertença ao fundo e $a_i = 0$, caso não tenha rótulo definido, o algoritmo *GrowCut* opera iterativamente, atualizando o vetor **a** até todos os elementos do grafo terem um rótulo definido e as condições impostas terem sido respeitadas, 1 ou -1. Para inicialização, alguns rótulos precisam ser previamente definidos, quanto maior a quantidade de rótulos pré definidos melhor a segmentação obtida (Figura 3.13).



(a) (b) (c)

Figura 3.13: Aplicação da técnica de segmentação via grafos *GrowCut* em uma imagem. A segmentação manual de fundo (em azul) e objeto (em vermelho) determinam a qualidade da segmentação pela quantidade de rótulos prévios que são criados, antes da aplicação do *GrowCut*. O algoritmo retorna resultados aceitáveis para a segmentação da estátua, mesmo com diante de uma rotulação prévia pobre, (a) e (b), e retorna um bom resultado para uma rotulação prévia próxima à segmentação desejada (c).

Mantendo a comparação entre *GrowCut* e SLIC, para a técnica de corte, pode-se considerar cada elemento i do grafo como o seu próprio centroide que, a cada iteração, em vez de atacar, é atacado pelos seus vizinhos $j \in Q$, recebendo o rótulo do vizinho mais próximo a ele. Essa distância é medida por uma função de custos:

$$g(i,j) = 1 - \frac{||\vec{\mathbf{I}}_i - \vec{\mathbf{I}}_j||^2}{max||\vec{\mathbf{I}}||^2}, \tag{3.13}$$

Para os valores médio que representariam o centroide no SLIC, cada elemento i no *GrowCut* recebe uma função de transição local θ_i^t, que recebe inicialmente (primeira iteração, $t = 0$) valor 1 para os nós rotulados e 0 para nós não rotulados. A condição para que um nó i seja atacado por um vizinho j e ganhe o seu rótulo é:

$$g(i,j) \cdot \theta_j^t > \theta_i^t. \tag{3.14}$$

O vértice i também ganha um novo valor para a função de transição local baseado no valor do vizinho j:

$$\theta_i^{t+1} = g(i,j) \cdot \theta_j^t. \tag{3.15}$$

O algoritmo cessa iterações quando a condição (3.14) não é mais respeitada, ponto no qual todos os elementos estão classificados/rotulados.

4 PROPOSTA PARA CONSTRUÇÃO DE REGIÕES EM IMAGENS E SEUS DESCRITORES LOCAIS

4.1 INTRODUÇÃO

Neste Capítulo, serão abordados os métodos propostos para a criação de superpixels/regiões que determinam os grafos de regiões, bem como os métodos utilizados para a determinação dos descritores locais referentes a essas regiões. Tanto a criação de regiões quanto a criação dos descritores é inspirada em processos da SIFT, que visam encontrar características nas imagens que se preservem diante de transformações geométricas como o escalonamento e rotação, e em face de mudanças nos níveis de intensidade. A partir do descritor, é proposto um método para determinação de regiões correspondentes e de vetores de estimativa de movimento entre duas imagens. O rastreamento de regiões, pela determinação de correspondências, permite que a criação de regiões seja orientada ao objeto e ao seu movimento.

4.2 *WATERSHED* EM UM ESPAÇO DE ESCALAS

O descritor proposto tem inspiração no algoritmo SIFT, no qual os descritores são obtidos de forma a tentar representar um objeto ou cena em diferentes escalas [10]. Essa representação em várias escalas tem como objetivo o reconhecimento de um objeto/cena apresentada em qualquer tamanho em uma imagem, fazendo-se necessária uma seleção de pontos (pontos-chave) para a redução na quantidade de dados em análise (Figura 2.11).

Para os objetivos propostos neste trabalho, a mudança de escala esperada para um objeto entre dois quadros consecutivos é pequena. Isso restringe a aplicação do algoritmo, para o qual se escolheu construir os descritores apenas para regiões de uma imagem dentro de um espaço de escalas. A imagem é definida pelos parâmetros n, σ e T, e selecionada visualmente pelo usuário, de forma que a sobre-segmentação oferecida represente bem o objeto de interesse (Figura 4.1).

Diferentemente de alguns algoritmos utilizados para criação de agrupamentos (regiões) em uma imagem, como *k-means* e *mean shift*, a técnica *watershed*, quando baseada nos mínimos locais da imagem, não depende de um usuário para determinação do número de agrupamentos. A quantidade de regiões formada está diretamente ligada a uma propriedade intrínseca da imagem, os mínimos locais definidos pelo módulo do gradiente da imagem, dos quais regiões emergem até a formação de barreiras entre elas [18].

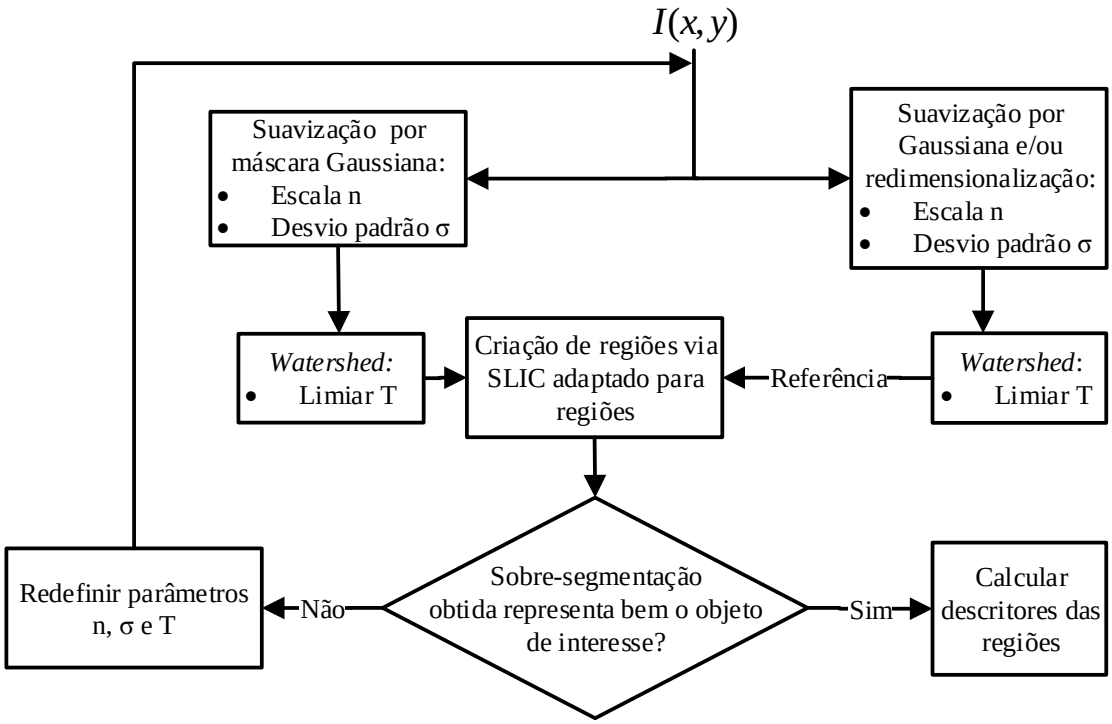


Figura 4.1: Diagrama do algoritmo proposto para a determinação de regiões e descritores em uma imagem $I(x, y)$. A imagem recebe suavizações que dependem da escala n e de um desvio padrão σ (caminho à esquerda), para a aplicação da técnica de *watershed* para a criações de regiões em uma aproximação de gradiente de $I(x, y)$, que tem valores nulos a partir limiar T. De forma semelhante, uma suavização mais intensa é aplicada à $I(x, y)$, que ainda pode ser redimensionada antes da criação de regiões via *watershed* (caminho à direita), essas regiões são tomadas como referência para o reagrupamento das regiões previamente definidas. Uma vez selecionados os parâmetros n, σ e T que produzam uma imagem com sobre-segmentação satisfatória para o objeto de interesse, os descritores locais são calculados para esta imagem selecionada.

Se assemelhando ao SIFT [10] na busca de pontos especiais para objetos em imagens sujeitos a transformações, sejam elas geométricas ou em magnitude, a técnica *watershed* se mostrou a mais indicada para a criação de regiões de maneira não supervisionada e para a determinação do descritor proposto. Apesar da preservação de pontos (mínimos locais), o algoritmo *watershed* perde em desempenho na irregularidade das bordas e no tamanho das regiões formadas a partir desses pontos [15]. Neste trabalho, foram obtidas regiões com bordas e tamanhos mais estáveis ao se mesclar o algoritmo SLIC [15] ao *watershed* em um agrupamento por escalas.

No caso proposto, o gradiente para aplicação do *watershed* é aproximado pela combinação da convolução do filtro detector de bordas Sobel com os três canais de cores no padrão CIELAB [40] da imagem. A adição de propriedades de oposição de cores no descritor proposto, é um diferencial em relação ao algoritmo SIFT, que leva em consideração somente o mapa de luminância da imagem.

Pares de vistas de cenas (Figura 4.2) serão utilizadas ao longo deste capítulo para exemplificar e testar o algoritmo proposto. A escolha dessas cenas, que possuem um leve deslocamento horizontal entre si, recai na semelhança das transformações entre cenas com as de uma transição entre quadros subsequentes de um vídeo.

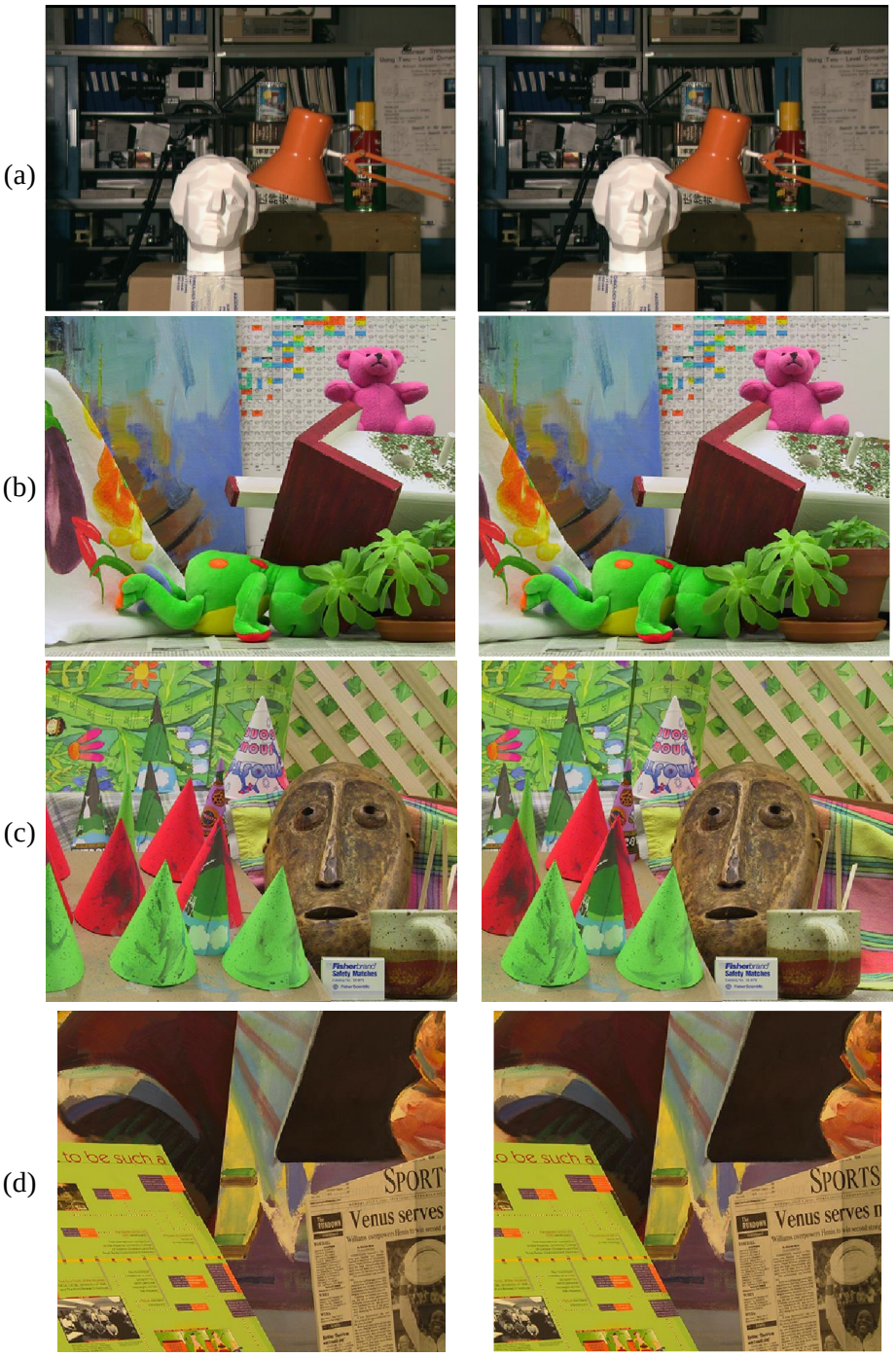


Figura 4.2: Pares de vistas distintas de cenas utilizados para avaliação do algoritmo proposto. Uma das imagens dos pares será submetida a transformações geométricas e de intensidade. Ao longo deste capítulo as cenas também exemplificam o algoritmo proposto, cenas intituladas como: (a) *Statue*; (b) *Teddy*; (c) *Cones*; e (d) *Venus*.

4.2.1 Aproximação do gradiente

O filtro Sobel é aplicado sobretudo em problemas de detecção de bordas em uma imagem ao longo de uma das suas direções. Por meio de uma combinação dessas resultantes em cada direção, uma soma euclidiana, por exemplo, pode-se obter uma imagem que ressalta suas bordas independentemente da sua disposição. A máscara Sobel que representa uma detecção ao longo do eixo horizontal x, é expressa como:

$$h_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \tag{4.1}$$

A variação abrupta da máscara Sobel (equação 4.1) ao logo do eixo vertical, ressalta os pontos com variações no mapa de magnitude de uma imagem ao longo desse eixo, por intermédio da convolução entre a máscara e imagem. Quando essas variações perduram ao longo do eixo horizontal, bordas são ressaltadas nessa direção. A convolução de uma imagem $I(x, y)$ (Figura 4.2(a) à esquerda) e a máscara h_x retorna uma imagem $D_x(x, y)$ de componentes horizontais:

$$D_x(x, y) = h_x * I(x, y) \tag{4.2}$$

Para a obtenção de componentes verticais, aplica-se a máscara h_y (que é a transposta de h_x) à imagem:

$$D_y(x, y) = h_y * I(x, y). \tag{4.3}$$

O módulo do gradiente aproximado por essas duas componentes, D_x e D_y, retorna a imagem:

$$U(x, y) = \big\| \big(D_x(x, y), D_y(x, y)\big) \big\|. \tag{4.4}$$

Do canal de luminância $I_L(x, y)$ (Figura 4.3(a)), obtém-se a resultante $U_L(x, y)$ (Figura 4.3(d)), por meio das componentes horizontais (Figura 4.3(b)) e verticais (Figura 4.3(c)).

O procedimento é aplicado aos outros dois canais de cores individualmente, $I_a(x, y)$ e $I_b(x, y)$ (Figura 4.4 (a) e (b)). É possível observar que os módulos resultantes $U_a(x, y)$ e $U_b(x, y)$ (Figura 4.4 (c) e (d)), respondem de maneira mais enfática aos pontos de variação de cor, os quais podem não ser detectados analisando-se apenas a luminosidade (Figura 4.3(d)).

A técnica de *watershed* poderia ser aplicada às três imagens de módulo do gradiente obtidas pelos três canais CIELAB, o que forçaria a uma análise conjunta de três camadas de regiões. Optou-se em aplicar a transformação na combinação de módulos dos gradientes para os três canais:

$$U(x, y) = \sqrt{U_L(x, y)^2 + U_a(x, y)^2 + U_b(x, y)^2}. \tag{4.5}$$

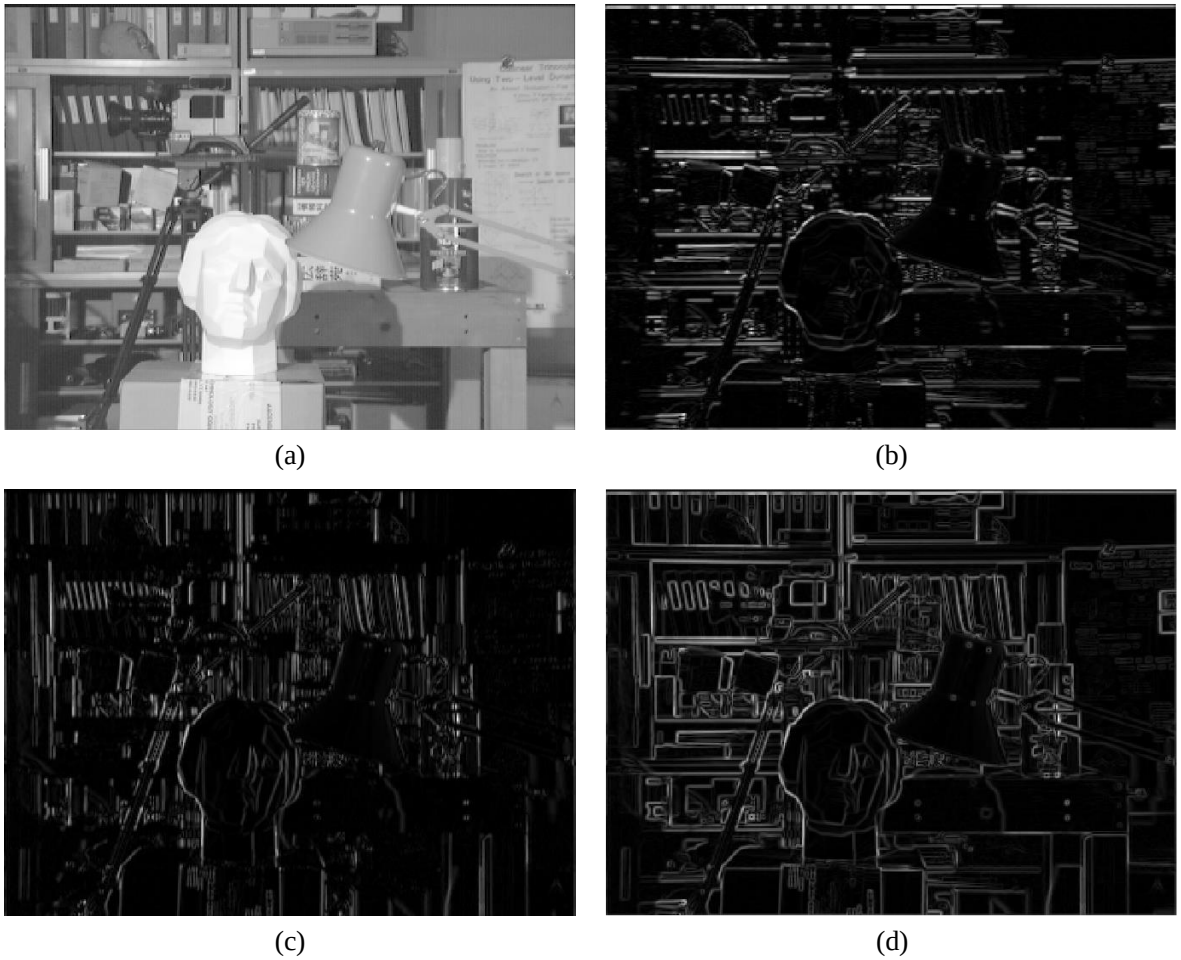
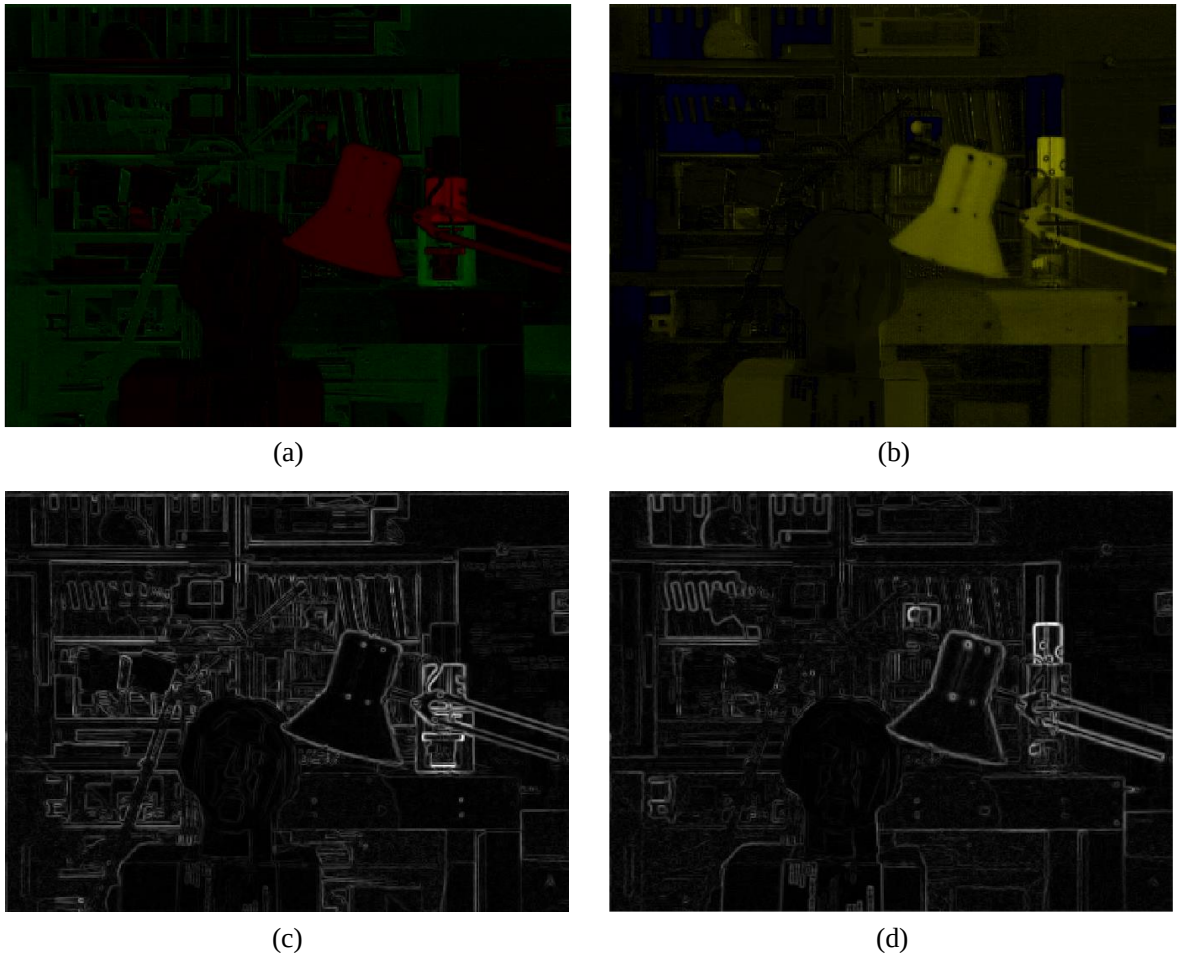


Figura 4.3: Ilustração da aplicação do filtro Sobel para aproximação do gradiente no (a) canal de iluminação para detecção de bordas (b) horizontais e (c) verticais, (d) que são combinadas em uma soma euclidiana.

A utilização de informações quanto às cores das imagens para a formação das suas regiões, reflete uma nova propriedade agregada à adaptação da SIFT para os grafos de regiões. A ideia parte dos campos receptivos presente no sistema visual humano, apresentados no Capítulo 2 e que são inspiração para o algoritmo SIFT. Os campos receptivos se apresentam tanto para as diferenças em iluminação quanto para a oposição de cores, verde em oposição ao vermelho e azul em oposição ao amarelo, enquanto que o SIFT utiliza apenas os aspectos de iluminação [10].

A Figura 4.4 ilustra em (a) e (b) a oposição de cores para a imagem da esquerda da Figura 4.2(a), oposição fornecida pelo sistema CIELAB de cores. Na Figura, (a) exibe a oposição de cores entre o verde e vermelho, e (b) a oposição entre o azul e amarelo. Esses mapas são registrados em valores que variam na faixa de $[-100, 100]$; o verde e o azul recebem os valores negativos e o vermelho e o amarelo os valores positivos. A Figura 4.4 (c) e (d) mostra o módulo do gradiente (equação (4.4)) para componentes vertical e horizontal aproximadas pelo filtro Sobel aplicado nos mapas de oposição de cores.



(a) (b)

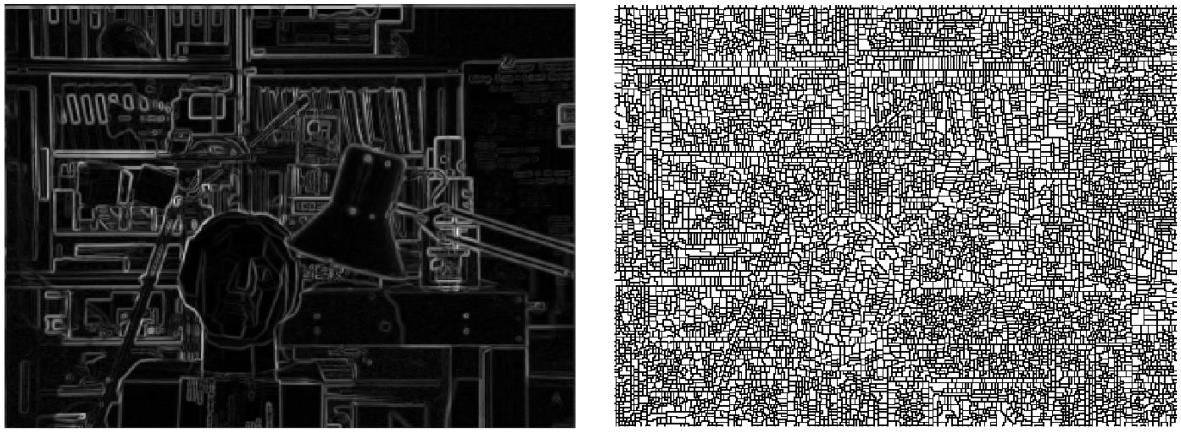
(c) (d)

Figura 4.4: Canais de cores no sistema CIELAB em (a) e (b), e seus respectivos módulos dos gradientes aproximados por um filtro detector de bordas Sobel em (c) e (d).

4.2.2 Formação de regiões pela *watershed* e definição de suas propriedades

O agrupamento de pixels em regiões (usualmente chamadas de superpixels) via *watershed*, está associado a uma redução de elementos, que pode ser observada na Figura 4.5: (a) quando aplicada à combinação de módulos do gradiente de uma imagem originalmente de dimensões 385×288, 110592 pixels; (b) obtém-se uma imagem com 7784 regiões, que é menos de um nono da quantidade de pixels na imagem original.

A redução do número de regiões em relação número de pixels para um valor abaixo de um nono não é arbitrária. O algoritmo de *watershed* adotado utiliza uma análise de vizinhança 8-conectividade [41]. Na formação de uma inundação a partir de um mínimo local são necessários no mínimo 8 pixels ao desse ponto para formação de uma represa. Isso limita a criação de regiões em uma região para cada composição de 3×3 pixels analisada. Um método para uma redução mais relevante no número de elementos dentro de uma imagem de regiões é apresentado neste mesmo Capítulo.



(a) 385 × 288 = 110592 pixels (b) N = 7784 regiões

Figura 4.5: Aplicação da *watershed*: (a) imagem fruto da combinação dos módulos do gradiente dos três canais de cores aproximados pelo filtro Sobel; (b) superpixels em branco delimitados pelas "barreiras"representadas pelas linhas em preto.

O intuito de se representar uma imagem via grafos de região é simplificar a análise dessa imagem com uma quantidade menor de elementos, de forma que esses essa redução ainda represente de maneira satisfatória o problema. Essa simplificação também é valida para as propriedades de cada agrupamento de pixels representado por uma região SP.

Assim como um pixel, um superpixel i tem associado a ele uma posição $\vec{\mathbf{r}}_i$ e um vetor de componentes de cores $\vec{\mathbf{I}}_i$. Além dessas duas características comuns a um pixel, está associada a um superpixel i uma área A_i, que é a quantidade de pixels que a compõe. A posição de um superpixel é obtida pela posição média dos pixels que o formam:

$$\vec{\mathbf{r}}_i = \frac{1}{A_i} \sum_{p \in SP_i} \vec{\mathbf{r}}_p, \tag{4.6}$$

em que $\vec{\mathbf{r}}_p$ é a posição (x_p, y_p) dos pixels que compõem a região e A_i a sua área, que computa a quantidade de pixels dentro desse superpixel SP_i.

Na geração de um mapa de magnitude de uma imagem digital, a quantidade de fótons que atinge um receptor eletrônico é convertido por uma função de potência inversa de γ antes de ser registrada. Desta forma, para intensidade em regiões é interessante considerar a soma de potências de γ, visto que:

$$\frac{1}{A_i} \sum_{p \in R_i} I(x_p, y_p) \leq \left(\frac{1}{A_i} \sum_{p \in R_i} I(x_p, y_p)^\gamma \right)^{\frac{1}{\gamma}}, \tag{4.7}$$

para $\gamma > 1$. A média simples das intensidades é menor que a raiz da média potencializada por γ, que é a intensidade real capturada. Uma possibilidade é usar a popular norma quadrática para a

média, de forma que:

$$I_i = \sqrt{\frac{1}{A_i} \sum_{p \in SP_i} I(x_p, y_p)^2}, \tag{4.8}$$

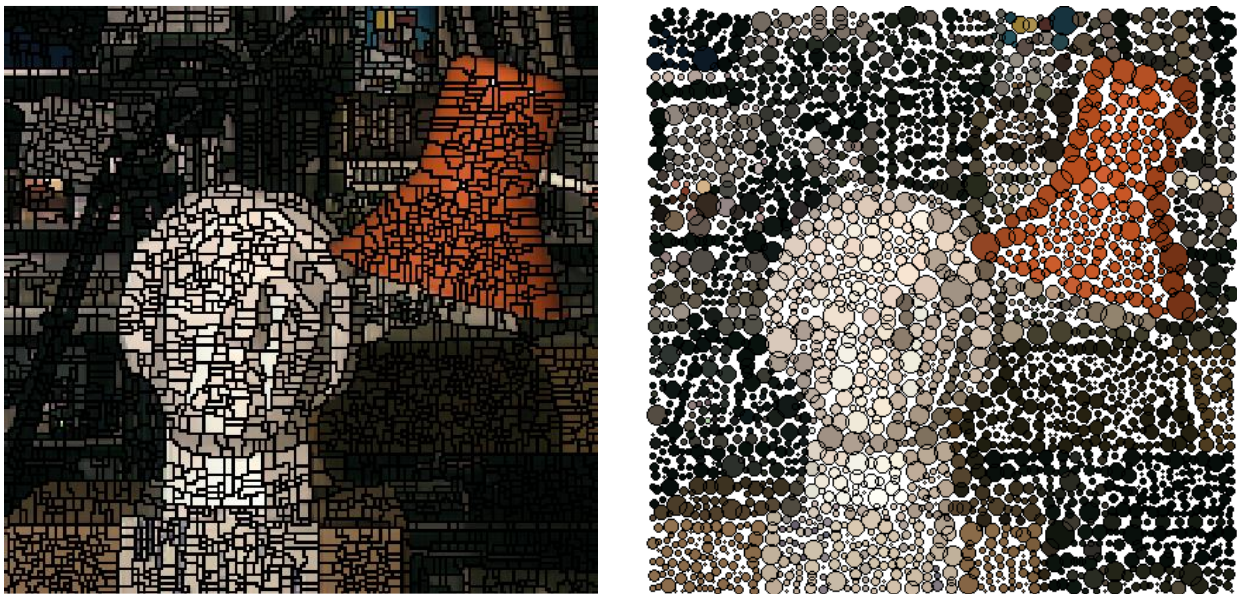
em que $I(x_p, y_p)$ é a magnitude de um canal de cores para o pixel p contido no superpixel SP_i, o processo então deve ser repetido para os 3 canais de cores.

Apesar da criação das regiões no sistema de cores CIELAB, o formato RGB foi utilizado para obter a média de intensidades e cores das regiões, evitando assim distorções causadas pela não linearidade do processo de conversão do sistema RGB para o CIELAB. A representação desse valor médio para intensidades pode ser observado na Figura 4.6(a), em que as regiões formadas são divididas pelos contornos em preto.

A imagem da Figura 4.6(a) fornece uma boa visualização dos pixels concatenados pelas respectivas regiões, entretanto, para fins práticos, cada região i possui apenas três propriedades, uma posição $\vec{\mathbf{r}}_i$, uma cor $\vec{\mathbf{I}}_i$ e uma área A_i, a qual pode ser convertida em um raio equivalente:

$$\overline{R_i} = \sqrt{\frac{A_i}{\pi}}, \tag{4.9}$$

tal simplificação é ilustrada na Figura 4.6(b).



(a) (b)

Figura 4.6: Ilustração dos agrupamentos formados após aplicação da *watershed*: (a) cada região agrupa uma certa quantidade de pixels que determina a sua área A_i, suas fronteiras são determinadas pelas linhas pretas e no interior é representado pela cor média; (b) uma forma mais simplificada de se representar esses elementos que na análise via grafos se tornam vértices, pontos, é os representando por um círculo de área A_i, os cálculos para gradiente e pesos de ligações nos Capítulos futuros, dependerão apenas desses três atributos, posição $\vec{\mathbf{r}}$, cor $\vec{\mathbf{I}}$ e área A.

Quando representada por um grafo, cada região da Figura 4.6(a) se torna um simples ponto, um nó. A Figura 4.6(b) ilustra esses nós com círculos dispostos nas posições calculadas para cada região e com cores médias determinadas. Esses círculos têm raio igual ao raio equivalente calculado pela equação(4.9), essa representação é interessante para o cálculo de gradiente para a construção dos descritores que envolvem apenas essas três propriedades.

4.2.2.1 Suavização e Limiar

A sobre-segmentação oferecida pela *watershed* é essencial para se analisar um objeto por meio do conjunto de diversas regiões que o compõe. Entretanto quando aplicada diretamente à imagem de gradiente $U(x, y)$, essa sobre segmentação pode retornar uma quantidade de elementos ainda indesejada. A forma encontrada para controlar essa produção de regiões foi a suavização da imagem antes da aplicação do filtro detector de bordas e a determinação de um limiar T para $U(x, y)$ antes de se aplicar *watershed*.

A suavização consiste na convolução da imagem original, ainda no sistema de cores RGB, com uma Gaussiana de desvio padrão σ:

$$L(x, y, \sigma) = I(x, y) * G(x, y, \sigma), \tag{4.10}$$

onde:

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{4.11}$$

Esse procedimento de suavização pode ser comparado a uma desfocalização da imagem que chega a retina ou à densidade de receptores capazes de resolver essa imagem, que é aliado a uma transformação:

$$\overline{U}(x, y) = \begin{cases} U(x, y) & \text{caso} \quad U(x, y) \geq T, \\ 0 & \text{caso} \quad U(x, y) < T. \end{cases} \tag{4.12}$$

da imagem $U(x, y)$ a partir de um limiar T. Elementos com magnitude menor que o limiar definido são anulados, limitando a detecção de bordas e a formação de regiões a uma sensibilidade T. Quanto maior o seu valor, menor a capacidade de detecção de diferenças e, por consequência, menor o número de regiões.

A escolha do limiar T fica a critério do objeto de segmentação definido. No caso de objetos que se destaquem em cor ou brilho em relação a sua vizinhança, um limiar alto promove uma boa delimitação das regiões objeto, sem se confundir com a vizinhança. Já no caso de objetos com pouco contraste, tanto em intensidade quanto em cor, em relação à vizinhança, sensibilidades mais baixas são recomendadas. O valor de T está limitado à escala de cores CIELAB da imagem $L(x, y)$ e ao filtro Sobel que dá ganhos às variações de L para gerar $U(x, y)$.

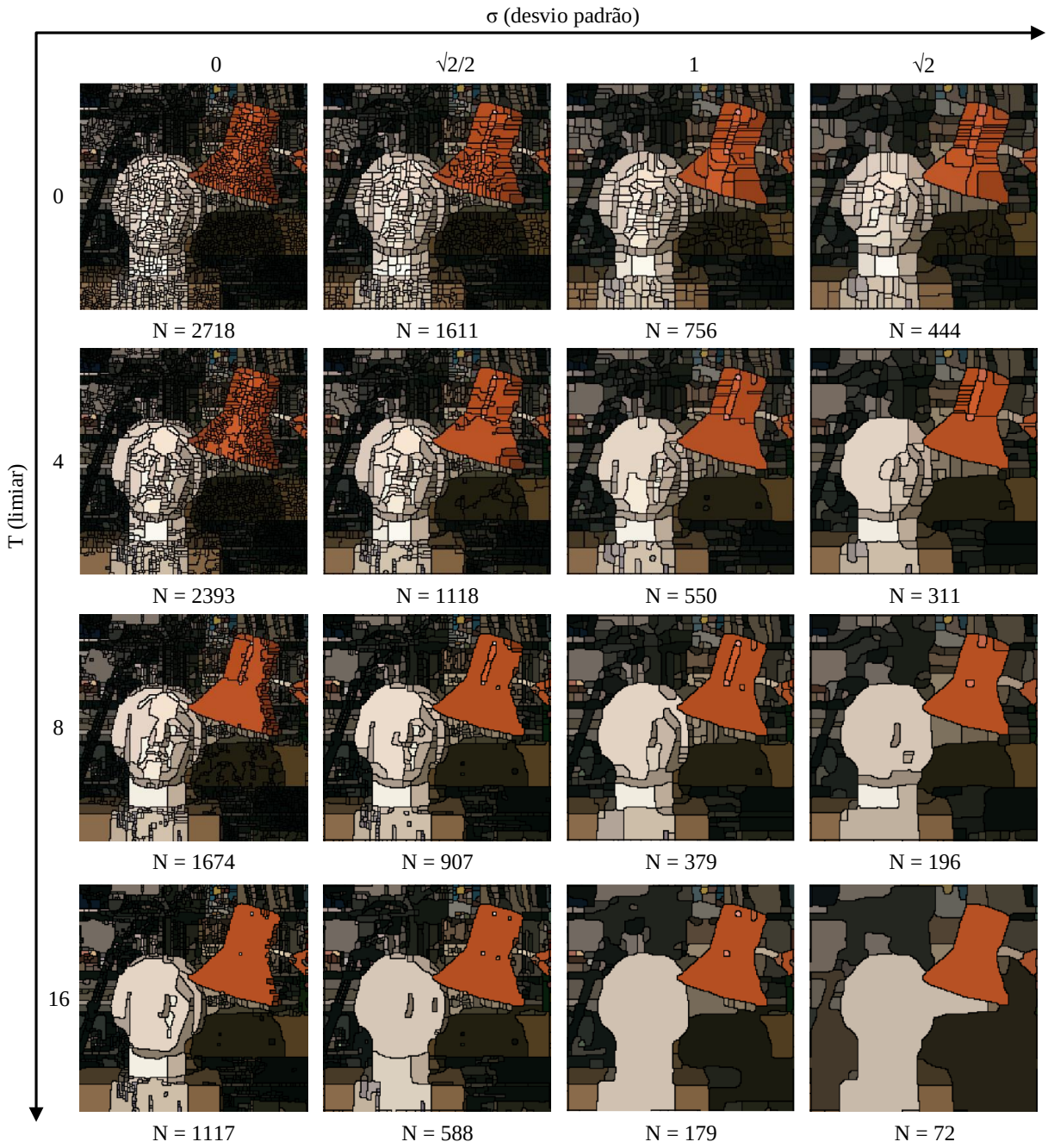
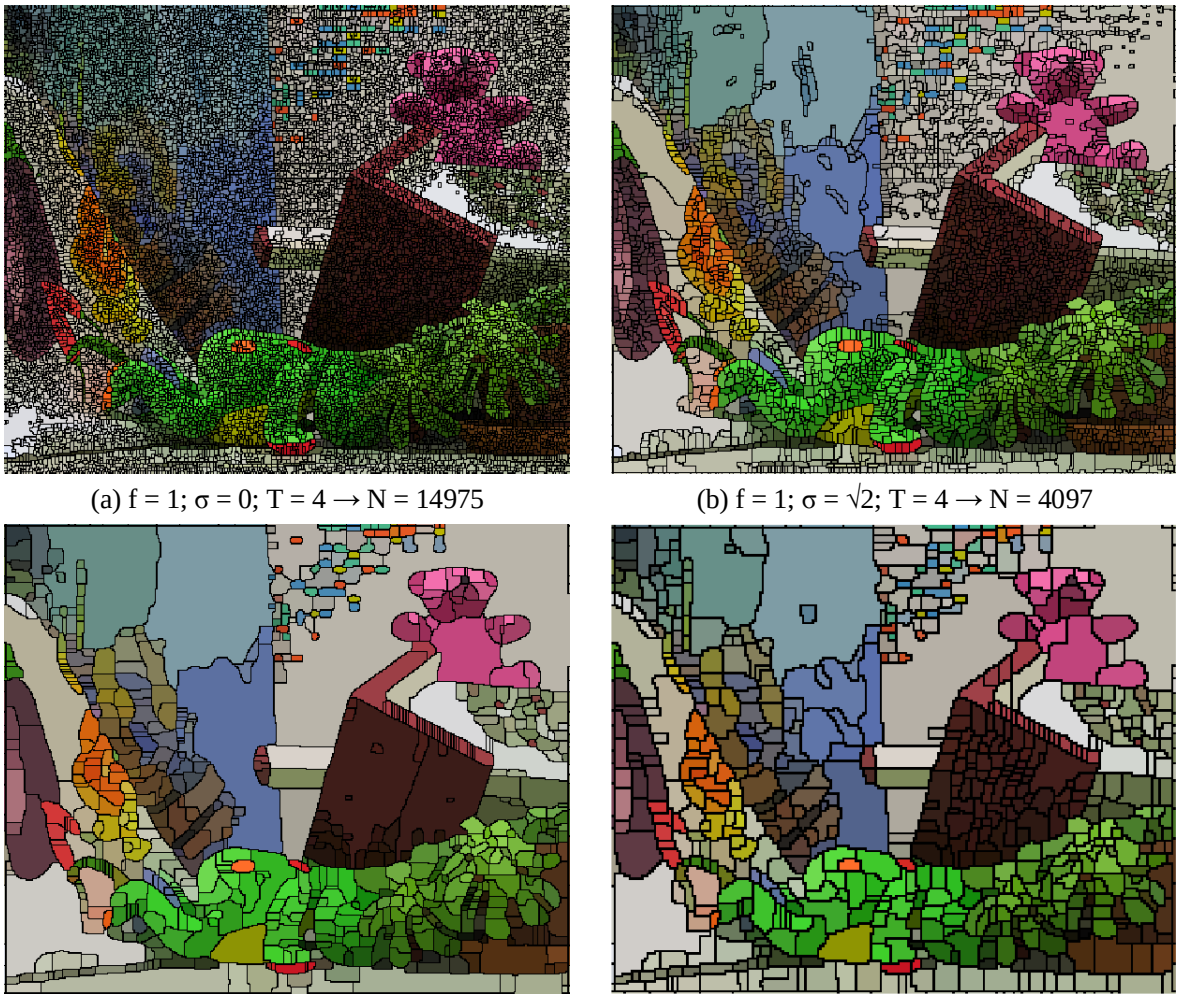


Figura 4.7: Relação entre suavização da imagem e aplicação de um limiar para o gradiente, antes de realizar um agrupamento via *watersehd*. A suavização é aplicada na imagem original, antes da determinação de seu gradiente, por um filtro gaussiano de desvio padrão σ. O limiar é determinado para a imagem de gradiente $U(x, y)$ que é a combinação do gradiente dos três canais de cores no padrão CIELAB. Nota-se quantidade de regiões formadas, uma relação inversa tanto para o crescimento do limiar quanto para o crescimento desvio padrão. A suavização diminui a relevância das bordas detectadas, aquelas que formam as barreiras na *watershed*, e o limiar diminui a quantidade de mínimos locais dos quais as inundações são iniciadas.

Um aumento de T promove uma diminuição do número de regiões formadas e, por consequência, do número de elementos para o grafo obtido final destinado à análise. Essa relação de decréscimo depende das características da imagem em estudo (Figura 4.7).

4.2.3 Agrupamento por Escalas

A relação entre suavização e diminuição no número de regiões aparenta ser uma relação direta, alheia às características da imagem (Figuras 4.7 e 4.8). Um borramento na imagem necessariamente altera as relações entre pixels de suas bordas, tornando transições mais suave e indiscrimináveis pelo limiar T adotado após filtragem com a máscara Sobel.



(a) f = 1; σ = 0; T = 4 → N = 14975

(b) f = 1; σ = √2; T = 4 → N = 4097

(c) f = 1; σ = 2√2; T = 4 → N = 827

(d) f = 2; σ = √2; T = 4 → N = 746

Figura 4.8: Técnica *watershed* aplicada à cena *Teddy* com um mesmo limiar T e diferentes graus de suavização e redimensionalização: (a) quando aplicada a imagem sem borramento em seu tamanho original, mesmo na presença de um limiar, a *watershed* produz uma grande quantidade de elementos, comparada à; (b) e (c) versões suavizadas em tamanho original, quanto maior o desvio padrão σ do filtro gaussiano, menor a quantidade de regiões; (d) entretanto uma redução na imagem original aliada a uma suavização por um filtro gaussiano, retorna uma quantidade de regiões semelhante a uma versão suavizada com maior σ da imagem em seu tamanho original (c), implica um ganho de homogeneidade nas regiões formadas.

Um fator limitante na detecção de uma borda usando uma máscara de Sobel (equação (4.1)) em uma imagem borrada, ou qualquer outra transição naturalmente mais suave dentro de uma imagem, é o tamanho dessa máscara detectora, 3×3.

A Figura 4.8 ilustra como uma imagem suavizada por Gaussianas de desvios padrões distintos, $\sigma = \sqrt{2}$ em (b) e $\sigma = 2\sqrt{2}$ em (c), promove uma significante redução no número de regiões em relação à imagem sem suavização, (a) para um mesmo limiar $T = 4$. Quando se aplica os mesmos princípios em uma imagem reduzida por um fator f $= 2$ (d), utilizando uma interpolação bicúbica e um filtro *anti-aliasing* [42], constata-se a produção de um número menor de regiões que nas outras três imagens, regiões que estão distribuídas de maneira mais homogênea.

Diminuindo-se o tamanho da imagem, verifica-se a capacidade de versões reduzidas produzirem regiões mais homogêneas (Figura 4.8 (d)), dada uma quantidade semelhante de regiões comparando-se versões mais amplas apenas suavizadas. Entretanto, existe uma degradação de bordas oriunda da redução da imagem e da perda de seus detalhes. Tal degradação foi superada combinando-se a *watershed* de imagens em seu tamanho original com suas versões borradas ou reduzidas e adaptando-se o algoritmo SLIC para fornecer um ajuste fino das bordas das regiões (Figura 4.9).

A Figura 4.9 esquematiza o funcionamento do agrupamento proposto, (a) em que as regiões formadas pela aplicação de *watershed* em uma imagem no seu tamanho original são (b) reagrupadas conforme as regiões definidas pela *watershed* em uma imagem mais borrada ou reduzida, (h) formando agrupamentos que aproveita a redução e homogeneidade de regiões de (b) e os detalhes oferecidos pela *watershed* aplicada à imagem original em (a).

Primeiramente, determina-se quais sub-regiões de (a) têm seu centro médio $\vec{\mathbf{r}} = (x, y)$, determinado pela equação (4.6), concatenados por qual região da imagem (b) após um processo de erosão (c). Nessa etapa espera-se classificar rapidamente um grande número de elementos, definido se esses pertencem a uma região e a qual região somente pela sua posição espacial. Esses elementos espacialmente classificados são chamados concatenados, V_c (Figura 4.9(d)).

Os elementos não concatenados V_{nc} (Figura 4.9(e)) são aqueles cujos centroides das regiões formadas por eles, se posiciona nas fronteiras, nas linhas na cor preta da imagem em (c). Por retratarem as bordas das regiões finais, essas sub-regiões do conjunto V_{cn} passam por um processo de agrupamento mais sofisticado, o SLIC (Capítulo 3), para definir a qual região pertencem. A imagem de referência para a formação de novas regiões (c) pode ser uma imagem no mesmo tamanho da original, mas com um maior borramento, ou uma versão reduzida e borrada da imagem original. Claramente, ao se comparar as posições das regiões da imagem original com as regiões de uma versão reduzida, há a necessidade de um escalonamento nas posições $\vec{\mathbf{r}} = (x, y)$ da magnitude do fator de redução f.

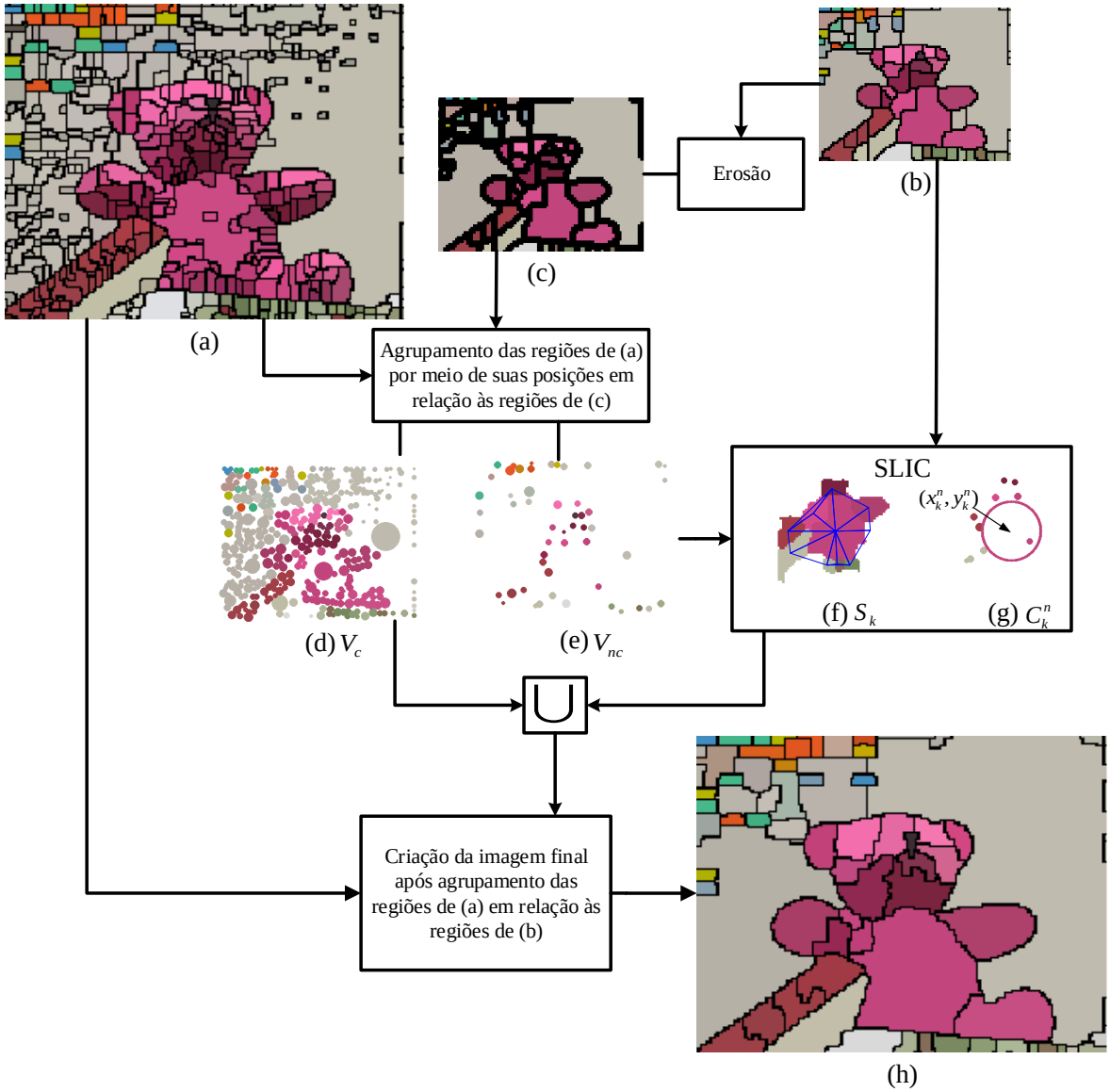


Figura 4.9: Diagrama ilustrativo para o processo de agrupamento por escalas: (a) uma imagem em seu tamanho original tem *watershed* aplicada a ela, para um σ_1 e T definidos; (b) deseja-se para um mesmo limiar T, redistribuir as regiões de (a) pelas regiões em menor número de uma imagem mais borrada $\sigma_2 > \sigma_1$ ou reduzida por um fator f > 1, uma referência para a criação novos agrupamentos; (c) um processo de erosão na imagem em (b) amplia as fronteiras (linhas na cor preta) e diminui o alcance das regiões, nas quais os elementos de (a) que tiverem seus centroides $\vec{\mathbf{r}} = (x, y)$ ali posicionados; (d) são rotulados como elementos concatenados V_{nc}, com uma região definida pela sua localização espacial, dentro ou fora da região; (e) os elementos não concatenados, V_{nc} são aquelas regiões de (a) que têm seus centros posicionados nas fronteiras definias em (c), as bordas na cor preta, região de dúvidas; (f) os elementos V_{nc} são agrupados por meio do algoritmo SLIC, uma restrição espacial para as regiões em análise para cada agrupamento k é definida por uma vizinhança Delaunay em torno da região k em (b); (g) a quantidade de elementos envolvidos no processo de agrupamentos via SLIC cai consideravelmente para a cada iteração em um agrupamento k, o centroide inicial C_k^o é definido pelos elementos concatenados V_{nc}, esse centroide atualiza seus valores de posição $\vec{\mathbf{r}}_k^n$, cor $\vec{\mathbf{I}}_k^n$ e área A_k^n a cada iteração n; (h) o agrupamento das regiões de (a) em relação às de (b) produz as regiões desejadas.

4.2.3.1 Considerações para o SLIC em regiões

Quando se aplica o SLIC em superpixels/região, algumas considerações se fazem necessárias, uma vez que os elementos possuem tamanhos distintos, diferentemente do caso via pixels. A divergência no tamanho de regiões interfere na distância $D_{i,k}$ de um elemento i a um centroide k, para a equação (3.5) descrita no Capítulo 3 e utilizada na proposta original [15].

Neste trabalho, optou-se por eliminar o termo m, determinando a distância de uma região i a um centroide k como:

$$D_{i,k} = dc_{i,k}\frac{ds_{i,k}}{\overline{R}_k + \overline{R}_i}. \tag{4.13}$$

Obtém-se uma expressão na qual a distância entre cores fica ponderada pela distância espacial normalizada pela soma entre os raios equivalentes do centroide k, $\overline{R}_k = \sqrt{\frac{A_k}{\pi}}$, e da região i, $\overline{R}_i = \sqrt{\frac{A_i}{\pi}}$. Elimina-se uma variável de entrada, produzindo-se resultados satisfatórios.

A normalização pela soma dos dois raios equivalentes coloca a distância espacial $ds_{i,k}$ em termos das dimensões do centroide e da região. Considera-se um contato entre um centroide k e uma região i, quando esta dista do centroide em magnitude menor ou igual a $\overline{R}_i + \overline{R}_k$, nessas condições a distância $D_{i,k}$ é fortemente dependente da distância entre cores $dc_{i,k}$ e, para efeitos de competição, se torna independente da dimensão das regiões e dos centroides.

A normalização pelos raios equivalentes assemelha-se àquela globalmente feita pelo termo S (equação (3.5)), que é o tamanho da aresta do quadrado que fornece a área esperada do superpixel na ref. [15], sendo que a competição entre centroides se estende a uma área quadrada $2S \times 2S$. Para o método proposto, a área de atuação S_k de um centroide C_k (Figura 4.9(f)) é determinado pela vizinhança Delaunay [43] da região k de origem (Figura 4.9(b)) do agrupamento, todas os elementos não concatenados V_{nc} que têm seu centro dentro desta área, ficam em disputa pelo centroide k e por outros centroides que por ventura se sobreponham a essa área.

Outra consideração se refere a média dos centroides, que deve contabilizar a contribuição de cada elemento $i \in C_k$, ou seja, sua área A_i. A área de um centroide é a soma das áreas das regiões que o compõe:

$$A_k = \sum_{i \in C_k} A_i. \tag{4.14}$$

Para a posição $\vec{\mathbf{r}}_k$ de um centroide k, por exemplo, é dado por:

$$\vec{\mathbf{r}}_k = \frac{1}{A_k}\sum_{i \in C_k} A_i\vec{\mathbf{r}}_i, \tag{4.15}$$

que é a média ponderada entre as posições das regiões. Para o vetor de cores $\vec{\mathbf{I}}_k$ recorre-se novamente ao sistema RGB para o cálculo da norma quadrática média em cada componente de

cor:

$$I_k = \sqrt{\frac{1}{A_k} \sum_{i \in C_k} A_i I(x_i, y_i)^2}. \tag{4.16}$$

Os valores de posição, cor e área para os centroides são atualizados a cada iteração. O critério de parada de 4 iterações se mostrou suficiente para um bom reagrupamento das regiões não concatenadas V_{nc}.

4.2.4 Espaço de escalas

Os agrupamentos formados pela *watershed* de uma imagem em escalas distintas, ajustados pelo algoritmo SLIC, permitem a criação de um espaço de escalas semelhante ao adotado na ref. [10] (Figura 4.10). O agrupamento por escalas proposto é definido por duas imagens, uma que serve como referência para os novos agrupamentos (Figura 4.9(b)) de elementos oriundos de uma versão mais detalhada da imagem (Figura 4.9(a)).

No espaço de escalas proposto, a imagem de regiões detalhada é uma *watershed* aplicada à imagem em seu tamanho original, que recebe um borramento de acordo com a oitava do agrupamento. Por exemplo, a primeira imagem de regiões da primeira oitava não recebe suavização, suas regiões são distribuídas por uma imagem de referência que é suavizada (equação 4.10) por uma gaussiana de desvio padrão $\sigma > 0$. Um exemplo de agrupamento por escala resultante desta primeira oitava é exibido na primeira linha da Figura 4.10.

A transição de uma oitava para outra é determinada quando cria-se uma referência a partir da imagem original redimensionada por um fator de redução f $= \sqrt{2}$, ou seja, as dimensões da imagem ficam menores por uma razão $\sqrt{2}$ e sua área cai pela metade. Uma oitava n tem como imagem de referência *watersheds* aplicadas à imagem original reduzida por um fator:

$$\text{f(n)} = (\sqrt{2})^{(n-1)}. \tag{4.17}$$

A imagem de referência da Figura 4.9(b) tem metade das dimensões da imagem original, ou seja, define uma terceira oitava, f(3) $= \sqrt{2}^{(3-1)} = 2$.

Para uma primeira oitava adota-se $\sigma = 0$ (sem borramento), para uma oitava $n > 1$ o desvio padrão da suavização é dado por $\sigma = \frac{\sqrt{2}}{2}\text{f}(n)$ (equação 4.11). O limiar T para todas as situações, referência ou imagem destinada a novos agrupamentos, é o mesmo. O espaço de escalas ilustrado na Figura 4.10 tem limiar T definido como 8 para todas as *watersheds* aplicadas no processo de agrupamentos por escalas. Observa-se uma diminuição gradual no número de elementos com o crescimento da escala e oitava.

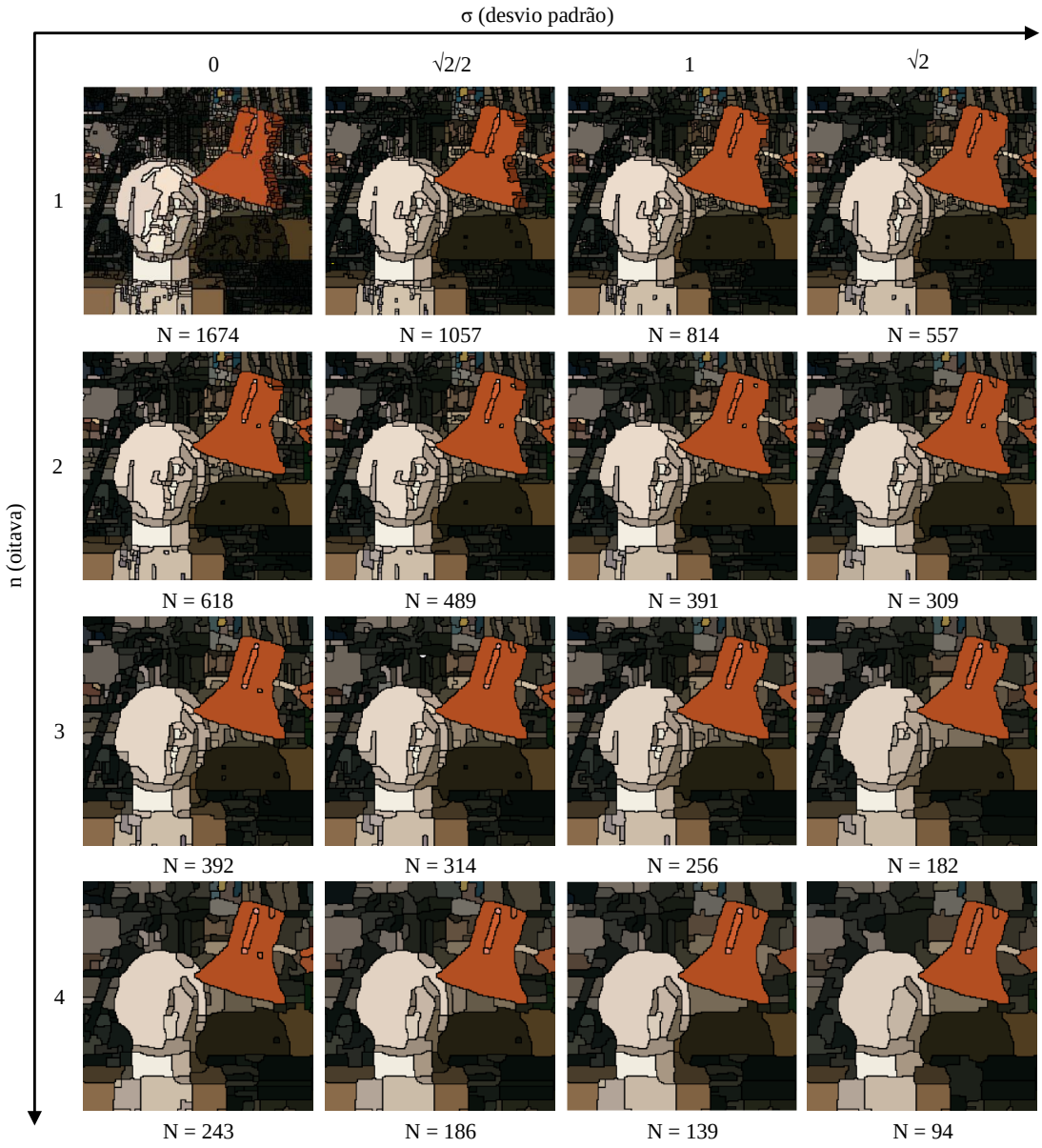


Figura 4.10: Ilustração para o espaço de escalas definido pelo método de agrupamento por escalas proposto, o crescimento da escala implica em um aumento dos agrupamentos. Uma nova oitava é determinada pela redução da imagem original para aplicação da *watershed* e criação de uma imagem de regiões, essa imagem reduzida é tomada como referência para a formação de agrupamentos via SLIC.

Diferentemente da proposta original [10] que visa a determinação de pontos-chave dentro de um espaço de escalas, isto é, pontos especiais que se preservem diante de transformações na escala, o espaço de escalas para o algoritmo proposto visa orientar a criação de regiões, em nível de detalhamento e número de elementos. Na adaptação proposta, não é utilizado todo o espectro de imagens da escala para a criação dos descritores em seus pontos-chave, escolhida uma imagem no espaço de escalas, que satisfaça o usuário em número de regiões e grau de representatividade do objeto de interesse, descritores são calculados para todas as regiões dessa imagem.

4.3 DESCRITOR LOCAL PROPOSTO

Para adaptação do algoritmo SIFT para uma representação por superpixels, foi necessário definir procedimentos que garantissem as mesmas proposições do algoritmo original. A invariância às mudanças de escala, intensidade e rotação devem ser preservadas pela nova expressão proposta para o cálculo de gradientes, bem como pela região na qual o descritor é construído.

4.3.1 Cálculo do gradiente em regiões

O vetor gradiente indica a direção e magnitude do crescimento em um ponto de uma função de mais de uma variável. No caso de uma imagem digital, retrata a variação entre os valores dos pixels vizinhos a esse ponto de análise. Para a proposta original [10], o gradiente é aproximado pela equação (2.10), em que são analisadas as diferenças de intensidade entre os vizinhos 4-conectividade de um pixel. A diferença entre os valores dos pixels imediatamente acima e abaixo do pixel em análise determina a componente vertical desse gradiente, e a diferença horizontal determina a componente nessa direção.

As etapas adotadas para obter expressão para o gradiente neste trabalho são detalhadas na sequência com base na equação da proposta original (equação (2.10)), em que V_i^4 é a vizinhança 4-conectividade do pixel i, cuja intensidade não é levada em consideração para cálculos. Em geral, para os grafos de regiões as vizinhanças não são bem definidas, fazendo necessária a escolha de um outro tipo de vizinhança para análise. O gradiente obtido pela equação (2.10) é bastante sensível a ruído, que é superado pela construção de histogramas [10].

No caso do algoritmo SIFT, os pontos-chave são selecionados e a direção e magnitude associados a eles é determinada pela seleção de orientações de predileção, obtidos em histogramas de orientação de uma região ao redor do ponto. Para o descritor proposto, todos os as regiões formadas, são tratadas como pontos-chave. Optou-se que direção e magnitude associadas a uma região deveriam sair diretamente da expressão proposta.

Uma possibilidade para o cálculo do vetor gradiente $\vec{\mathbf{m}}_i$ para um vértice i de um grafo seria:

$$\vec{\mathbf{m}}_i = \sum_{j \in V} \frac{L_j - L_i}{||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||} \vec{\mathbf{u}}_{i,j}, \tag{4.18}$$

em que se considera a soma vetorial de todas as diferenças de intensidade $L_j - L_i$ entre os elementos j pertencentes ao grafo V, com o vértice i. A magnitude de cada contribuição na soma é ponderada pela distância $||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||$ entre o elemento i e o j, e a direção é dada pelo vetor unitário que conecta os dois vértices, definido como:

$$\vec{\mathbf{u}}_{i,j} = \frac{\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i}{||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||}. \tag{4.19}$$

A equação 4.18 é menos sensível a ruídos, por levar em consideração todos os elementos pertencentes ao grafo, ponderados por sua distância até o ponto de análise. A equação leva em consideração a intensidade L_i do ponto i de análise, entretanto não considera como as vizinhanças desses elementos se interagem e nem a possibilidade da distância $||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||$ retornar um valor próximo a zero.

Para expressão adotada neste trabalho, adiciona-se na equação (4.18) uma normalização na distância entre os elementos e um deslocamento unitário. Sendo L o canal luminância da respectiva região:

$$\vec{\mathbf{m}}_i = \sum_{j \in V} \frac{L_j - L_i}{\frac{||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||}{(\overline{R_i} + \overline{R_j})} + 1} \vec{\mathbf{u}}_{i,j}. \tag{4.20}$$

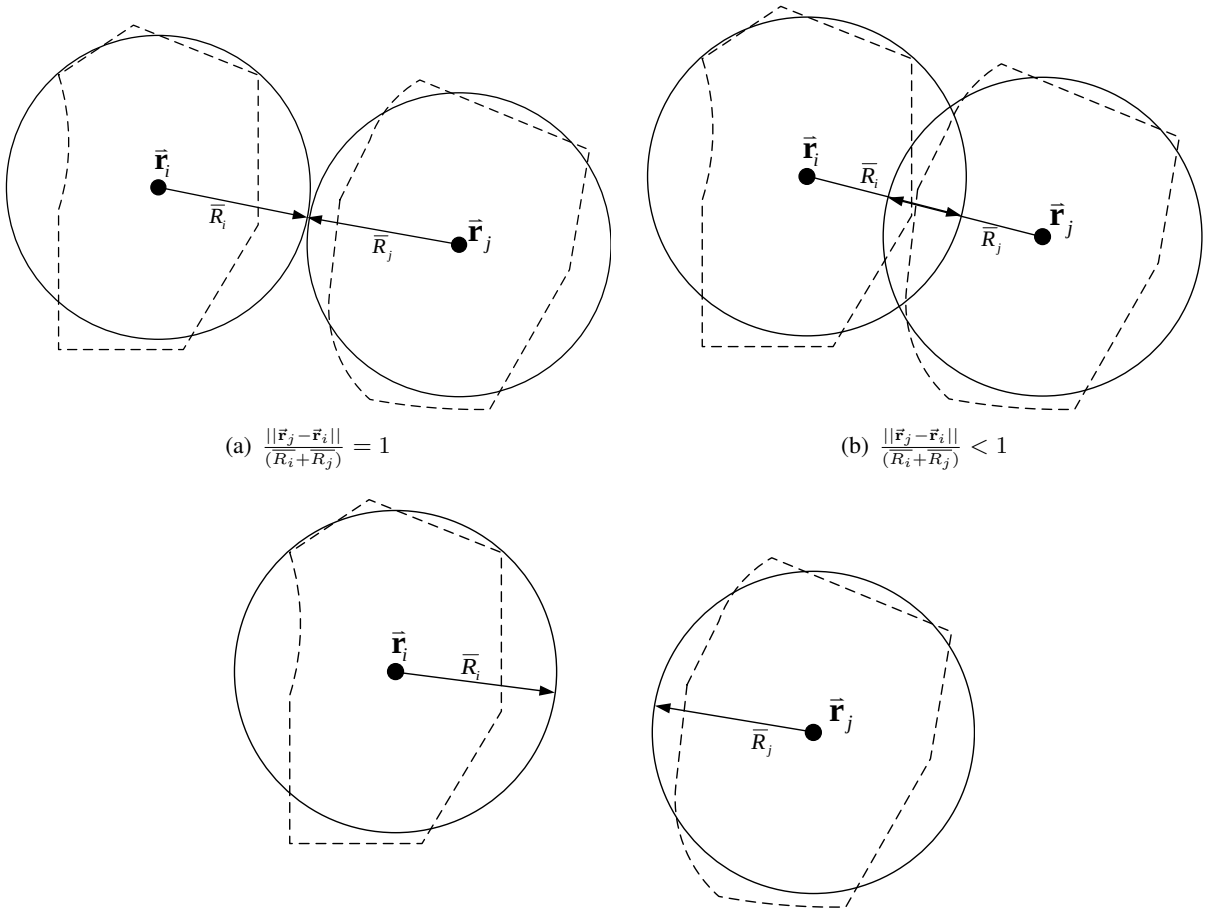
Ao se normalizar a distância entre os elementos, $\vec{\mathbf{r}}_i$ e $\vec{\mathbf{r}}_j$, pela soma de raios equivalentes, $\overline{R_i} + \overline{R_j}$, obtém-se uma expressão que tenta aproximar a interação entre duas regiões, pela proximidade dos círculos que as representam, tendo dimensões fornecidas pelos seus respectivos raios equivalentes (Figura 4.11).

Como foi salientado, a simplificação de uma imagem em superpixels carrega consigo uma simplificação nas relações de fronteira entre as regiões. Cada região i passa a ser representada por três propriedades: uma posição $\vec{\mathbf{r}}_i = (x_i, y_i)$, uma cor $\vec{\mathbf{I}}_i = (L_i, a_i, b_i)$ (CIELAB) e uma área A_i.

Ao se normalizar a distância entre elementos (equação (4.20)), busca-se colocar a distância entre duas regiões em termos de seus raios equivalentes. Uma unidade dessa distância normalizada representa um tangenciamento dos círculos que representam as regiões (Figura 4.11(a)). Valores menores que 1 para essa distância, indicam que fronteiras as regiões são muito próximas (Figura 4.11(b)) e uma distância normalizada maior do que 1 indica que as fronteiras das regiões se toquem em poucos pontos ou até estejam desconectadas (Figura 4.11(c)).

Para valores menores do que 1 para a distância normalizada (Figura 4.11(b)), ainda há a possibilidade de uma situação extrema em que $||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i|| \approx 0$. Essa proximidade entre alguns dos centroides das regiões é fruto das suas conformações geométricas. Nessa situação limite, a expressão (2.10) retorna valores muito altos, podendo diminuir a força do descritor, visto que esta singularidade pode ser oriunda de uma instabilidade na formação das regiões. Na equação (4.20), evita-se essa situação ao se somar uma unidade no denominador da equação (2.10).

A normalização efetuada também visa preservar a invariância às mudanças de escala na imagem, ou possíveis distanciamentos ou aproximações de objetos dentro de uma cena. Regiões de maior área têm magnitude de gradiente maior que regiões de menor área, dada sua área de atuação, e acabam servindo como pontos de referência na construção dos descritores próximos a ele.



(a) $\frac{||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||}{(\overline{R_i} + \overline{R_j})} = 1$

(b) $\frac{||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||}{(\overline{R_i} + \overline{R_j})} < 1$

(c) $\frac{||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||}{(\overline{R_i} + \overline{R_j})} > 1$

Figura 4.11: Normalização pela soma dos raios equivalentes $\overline{R_i}$ e $\overline{R_i}$ da distância $||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||$ entre dois elementos i e j: (a) uma unidade dessa distância normalizada equivale a um tangenciamento entre os círculos que representam as regiões, círculos com dimensões determinadas pelos raios equivalentes $\overline{R_i}$ e $\overline{R_j}$; (b) valores menores do que 1 indicam uma proximidade e interação entre as fronteiras da região; (c) valores maiores do que 1 indicam um distanciamento entre as fronteiras das regiões.

Além do canal de luminância L (Figura 4.12(b)), o gradiente é calculado para os outros dois canais de cores no sistema CIELAB, medindo o distanciamento do verde ao vermelho (Figura 4.12(c)) e do azul ao amarelo (Figura 4.12(c)). Na Figura 4.12, temos uma imagem dividida em regiões (a) que são representadas por círculos em seus três canais, luminância (a), verde/-vermelho (b), azul/amarelo(c). A simplificação em círculos, com dimensões determinadas pelas áreas das regiões e posicionados nos centroides das mesmas, exemplifica a adoção de uma distância normalizada (equação (4.20)).

Para o canal de luminância (Figura 4.12(b)), por exemplo, as regiões que fazem fronteira com o setor esquerdo da estátua, são mais escura que ela, os vetores de gradiente indicam isso com sua orientação nessa direção. Se fosse adotada uma distância não normalizada (equação 4.18), a magnitude da interação entre essas regiões descritas seria menor, dada a dimensão da região referente a parte esquerda da cabeça da estátua e a distância entre os elementos.

A utilização dos canais de cores que representam a oposição de cores verde/vermelho (Figura 4.12(c)), azul/amarelo (Figura 4.12(d)), é um diferencial em relação ao SIFT que utiliza apenas os mapas de magnitude das imagens. Assim como o SIFT, esse diferencial é baseado nos campos receptivos presentes no sistema visual humano (Capítulos 2).

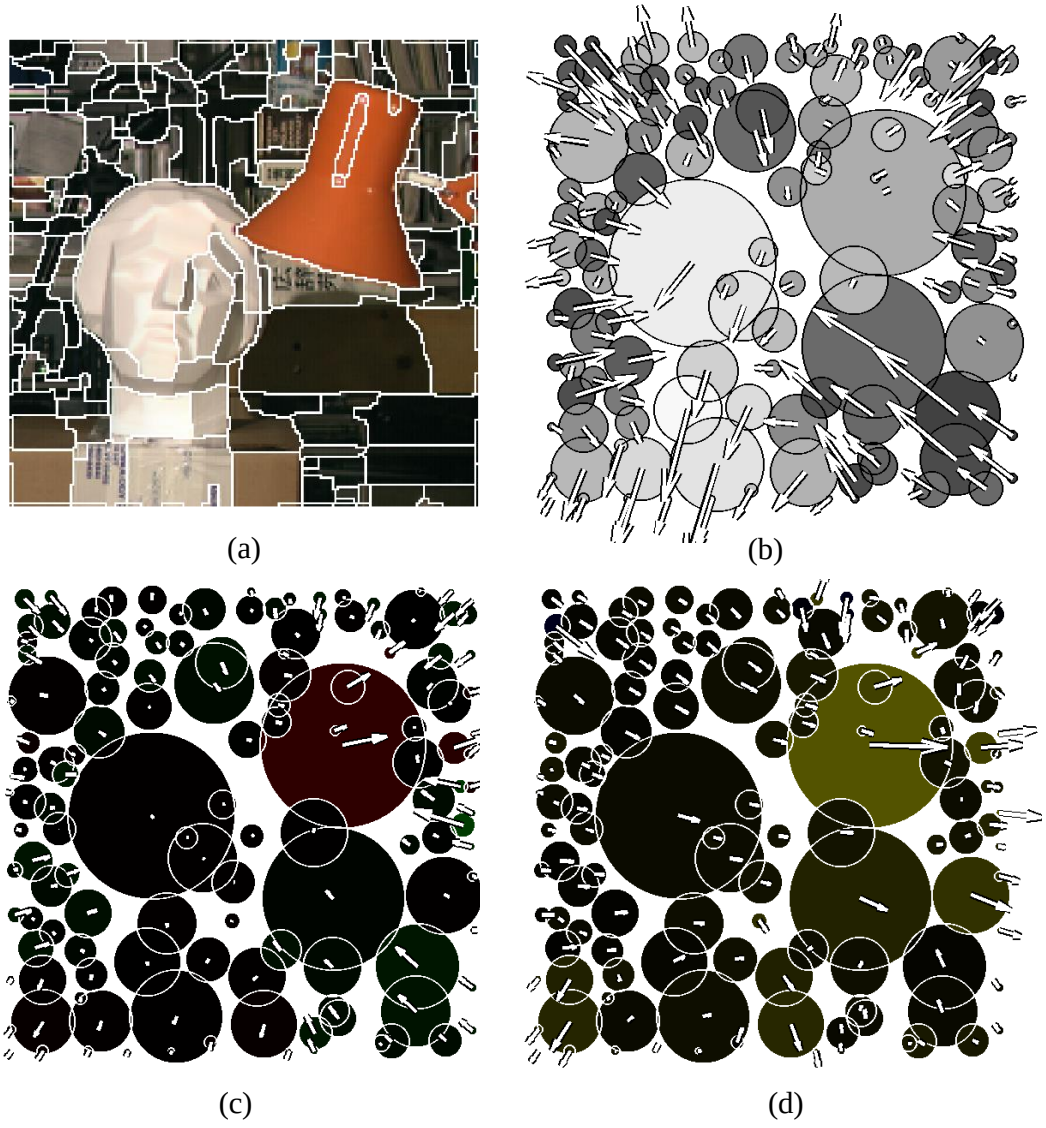
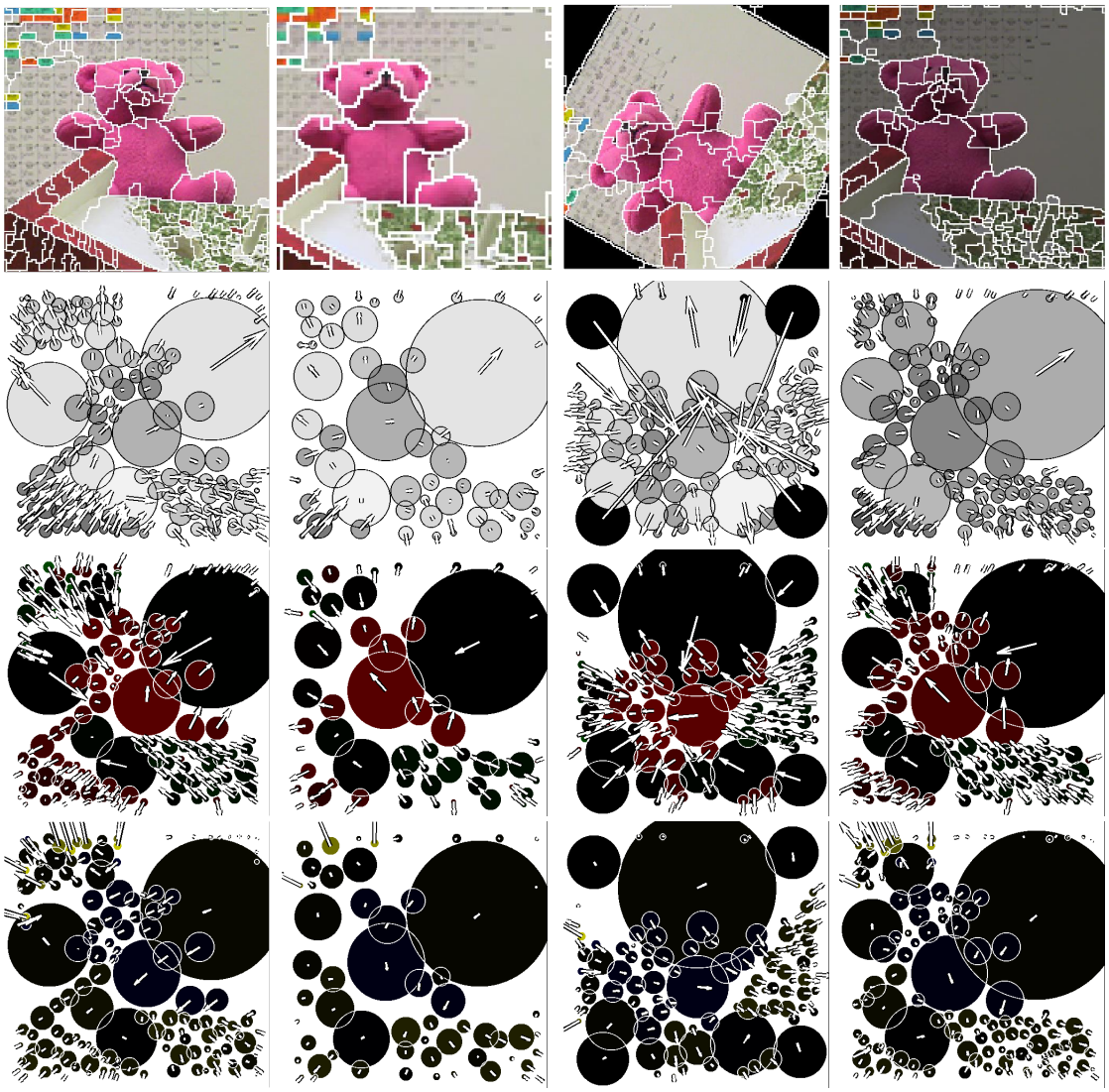


Figura 4.12: Representação em vetores para oo gradiente de regiões: (a) a imagem original é dividida em regiões conforme as técnicas de agrupamento definidas neste Capítulo, a simplificação dessas regiões por círculos; (b) podem representar o gradiente para canal de luminância; (c) de oposição das cores verde/vermelho e; (d) de oposição das cores azul/amarelo.

Aplicando transformações geométricas de mudança de escala (Figura 4.13 (b)), orientação (Figura 4.13 (c)), e de iluminação (Figura 4.13 (d)), percebe-se a manutenção das orientações da maior parte dos vetores gradiente em relação a imagem em seus aspectos originais (Figura 4.13 (a)), principalmente as orientações relacionadas a regiões de grandes áreas. Isso indica robustez do método de cálculo de gradiente proposto, que dentro de mudanças em escala, intensidade e orientação das imagens, preserva as características nos vetores de gradiente.



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

Figura 4.13: Comparativo para o gradiente proposto diante de transformações geométricas e em iluminação, a imagem no topo é aquela fornece as regiões para definição dos gradientes abaixo de cada imagem os gradientes dos três canais de cores no sistema CIELAB, todas regiões definidas uma mesma oitava n, escala σ e limiar T: (a) imagem original; (b) antes do agrupamento em regiões, a imagem original com uma redução pela metade em suas dimensões, exibida no mesmo aspeto original para facilitar a visualização; (c) imagem em (a) é rotacionada em 60^o no sentido anti-horário; (d) imagem em (a) recebe uma redução atenuação seu mapa de magnitude, de forma que essa cai para metade do valor.

4.3.2 Região de definição do descritor

Para dar uma orientação ao ponto-chave, no SIFT é necessário buscar uma orientação e uma magnitude vencedora dentro de um histograma de orientações bem detalhado do mapa de gradientes, em uma região em torno dos pontos-chave. Diferentemente do algoritmo SIFT [10], todos os vértices da imagem analisada são considerados pontos-chave, com orientação e magnitude determinadas pelo próprio vetor de gradiente $\vec{\mathbf{m}}_i$. No SIFT, além daqueles para determinar as orientações dos pontos-chave, histogramas são calculados ao longo de regiões/setores ao redor do ponto-chave, cujas componentes são combinadas para formar o descritor do ponto-chave.

Os histogramas que compõem o descritor na SIFT englobam a contagem de vetores em 8 direções distintas. Neste trabalho, foram adotadas 4 orientações para os histogramas, θ_1 a θ_4 (Figura 4.14(a)). A orientação ou não de um vetor de gradiente $\vec{\mathbf{m}}_i$ em relação a θ_k é dada pela função:

$$\Theta(\vec{\mathbf{m}}_j, \theta_k) = \begin{cases} 1, & \text{se } \vec{\mathbf{m}}_j \text{ está orientado na direção de } \theta_k \\ 0, & \text{caso contrário} \end{cases} \tag{4.21}$$

A função $\Theta(\vec{\mathbf{m}}_j, \theta_k)$ retorna o valor 1, quando a orientação mais próxima do vetor de gradiente $\vec{\mathbf{m}}_i$ é θ_k e 0 caso contrário.

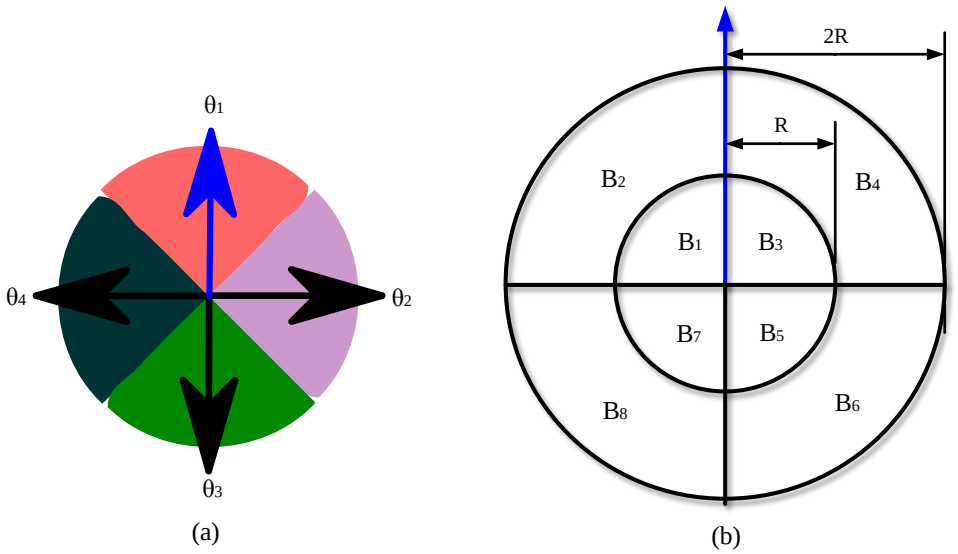


Figura 4.14: Determinação das orientações e das regiões para os cálculo dos histogramas de orientações: (a) 4 orientações, θ_1 a θ_4, para as quais serão criados histogramas em (b) cada setor de uma grade polar dividida em dois raios e quatro quadrantes, totalizando 8 setores, B_1 a B_8. O raio da grade polar é determinado é proporcional a R, definido pelas dimensões da região para qual se calcula o descritor.

O cálculo dos histogramas que compõem o descritor é feito por setores dentro de regiões ao redor do ponto-chave, entretanto, em vez da distribuição dos setores em uma região retangular [10], foi definida uma distribuição de setores por uma grade polar (Figura 4.14(b)).

A grade polar de análise é dividida em 4 quadrantes e dois raios, R e $2R$ (Figura 4.14(b)), contabilizando $4 \times 2 = 8$ setores B_l, $l = \{1, 2, ..., 8\}$. O raio equivalente $\overline{R}_i$ do superpixel i, para qual se está calculando o descritor, determina o raio da grade, definiu-se $R = 10\overline{R}_i$ uma vez que tal proporção apresentou os melhores resultados em testes.

Antes de contabilizar os histogramas, deve-se alinhar o contador (Figura 4.15(a)) e a grade polar (Figura 4.15 (b)) em relação ao vetor de gradiente $\vec{\mathbf{m}}_i$, do superpixel i para o qual está sendo computado o descritor. A definição da direção dos vetores de gradiente (Figura 4.15 (a)) é feita ao se alinhar a direção θ_1 com $\vec{\mathbf{m}}_i$ (Figura 4.15 (b)). A grade polar é posicionada com o centro convergente ao centro do superpixel i e alinhada com o vetor $\vec{\mathbf{m}}_i$ (Figura 4.15 (c)).

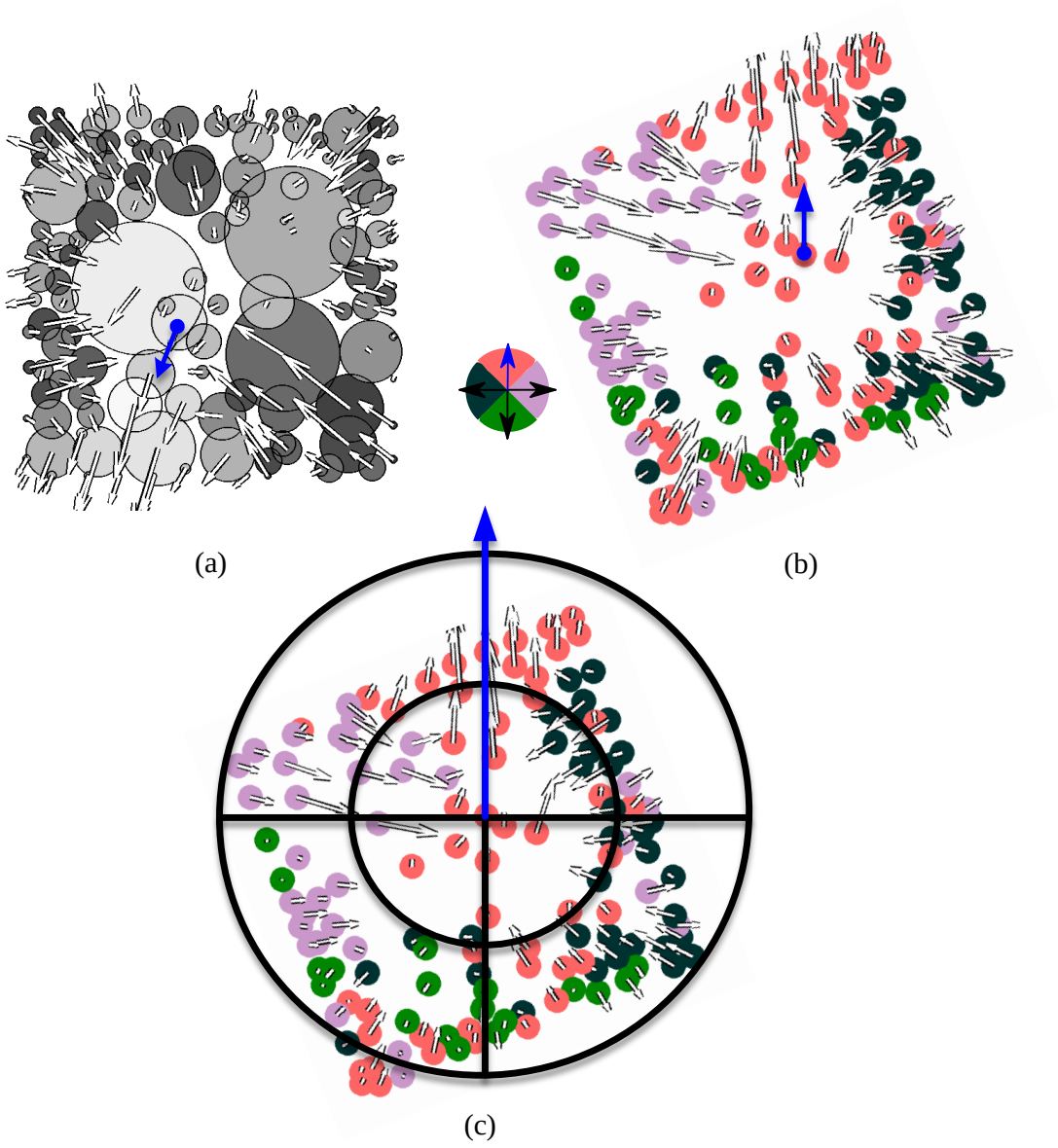


Figura 4.15: Processo de criação do descritor: (a) mapa de gradientes do canal de luminância (Figura 4.12) com um vetor de gradiente (cor azul) em destaque; (b) os vizinhos a esse vetor têm nele uma direção de referência, θ_1; (c) bem como a grade polar na qual são construídos histogramas de orientação por setor.

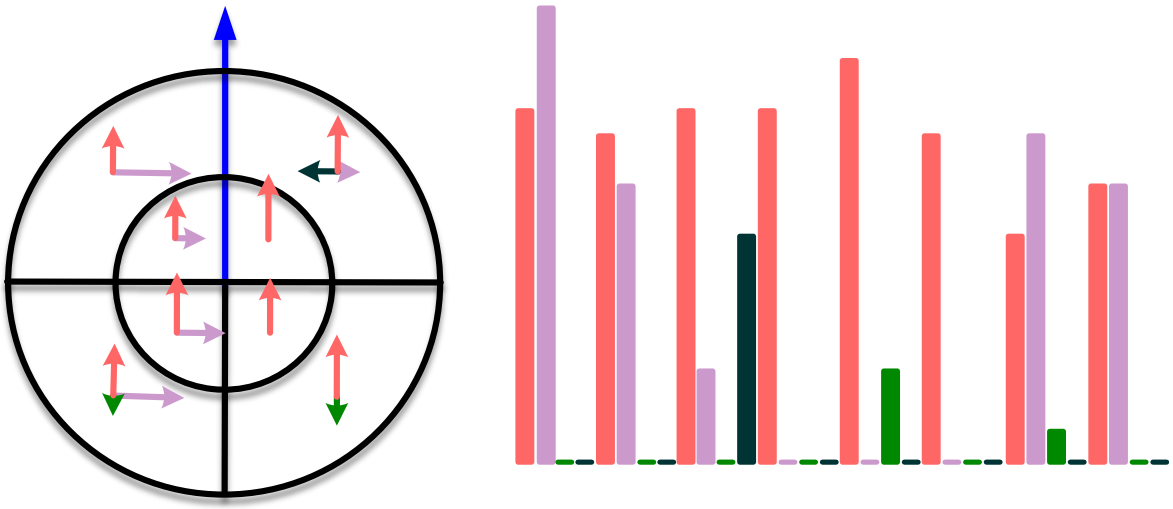
Cada componente do histograma, de cada um dos 8 setores, computa a quantidade de vetores orientados em uma das quatro possibilidades θ_1, θ_2, θ_3 e θ_4. Uma vez definido como pertencente a uma certa orientação, a contribuição de um vetor $\vec{\mathbf{r}}_j$, da região j dentro da grade polar de construção do descritor da região i, é ponderada pela magnitude do gradiente e por uma Gaussiana circular simétrica em torno da posição $\vec{\mathbf{r}}_i$ de i:

$$c_i(\theta_k, B_l) = \sum_{i \in B_l} \Theta(\vec{\mathbf{m}}_j, \theta_k) ||\vec{\mathbf{m}}_j|| e^{-\left(\frac{||\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i||}{2R}\right)^2}. \tag{4.22}$$

A orientação do setor de acordo com a orientação do vetor do elemento de referência, visa dar ao descritor invariância à rotação ou mudanças de orientação de objetos e cenas.

A ilustração da Figura 4.16 mostra os histogramas das 4 orientações (a) para cada um dos 8 setores, suas componentes são dispostos em um vetor (b) de tamanho $8 \times 4 = 32$ componentes. Esse descritor da região i para um canal de cores pode ser escrito como:

$$\vec{\mathbf{D}}_i = [c_i(\theta_1, B_1), c_i(\theta_2, B_1), ..., c_i(\theta_3, B_8), c_i(\theta_4, b_8)], \tag{4.23}$$



(a) (b)

Figura 4.16: Representação do descritor em um vetor: (a) 8 histogramas posicionados em seus respectivos setores que contabilizam para 4 orientações, os vetores de gradiente, as componentes desses histogramas, representados pelas setas em cada setor; (b) são dispostos em um vetor que ao ser normalizado, representa o descritor $\vec{\mathbf{D}}$ de um canal de cor para uma região.

O descritor final utilizado neste trabalho é formado por uma composição dos três descritores, com um tamanho $3 \times 32 = 96$, referentes aos seus respectivos canais de cor. Antes de serem dispostos em um único vetor, esses histogramas são individualmente normalizados, e quando dispostos em um único vetor, esse vetor também é normalizado. Esse vetor de norma unitária tem contribuição peso de $1/3$ para cada canal de cor. Esse procedimento de normalização objetiva fornecer ao descritor invariância quanto a mudanças na iluminação da imagem.

4.4 DETERMINAÇÃO DE REGIÕES CORRESPONDENTES

Para determinação de regiões correspondentes entre duas imagens com o descritor proposto, optou-se por utilizar o produto interno em vez de uma distância euclidiana, pela facilidade de implementação ao se confrontar conjuntos de descritores de duas imagens por uma multiplicação de matrizes e pelos valores nulos para os confrontos entre descritores ortogonais. Esse produto interno revela correspondências de forma que:

$$j_w(i) = \underset{j}{\operatorname{argmax}}\, \vec{\mathbf{D}}_i \cdot \vec{\mathbf{D}}_j, \tag{4.24}$$

sendo que i a região de uma imagem V_1, e $j_w(i)$ a região de maior produto interno com i dentro dos elementos de uma segunda imagem V_2. $j_w(i)$ só é considerado uma correspondência, se a partir da relação inversa:

$$i_w(j) = \underset{i}{\operatorname{argmax}}\, \vec{\mathbf{D}}_j \cdot \vec{\mathbf{D}}_i, \tag{4.25}$$

$[j_w, i_w]$ formam um par recíproco.

A Figura 4.17 ilustra os pares $[1, 4]$ e $[3, 7]$ como pares de correspondências. O produto interno de $i = 1$ da imagem V_1 tem maior produto interno na imagem V_2 com a região 5 e a relação inversa é respeitada.

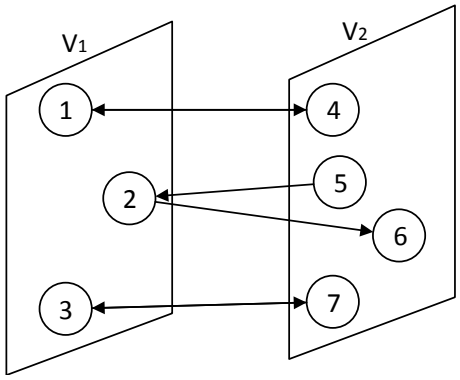
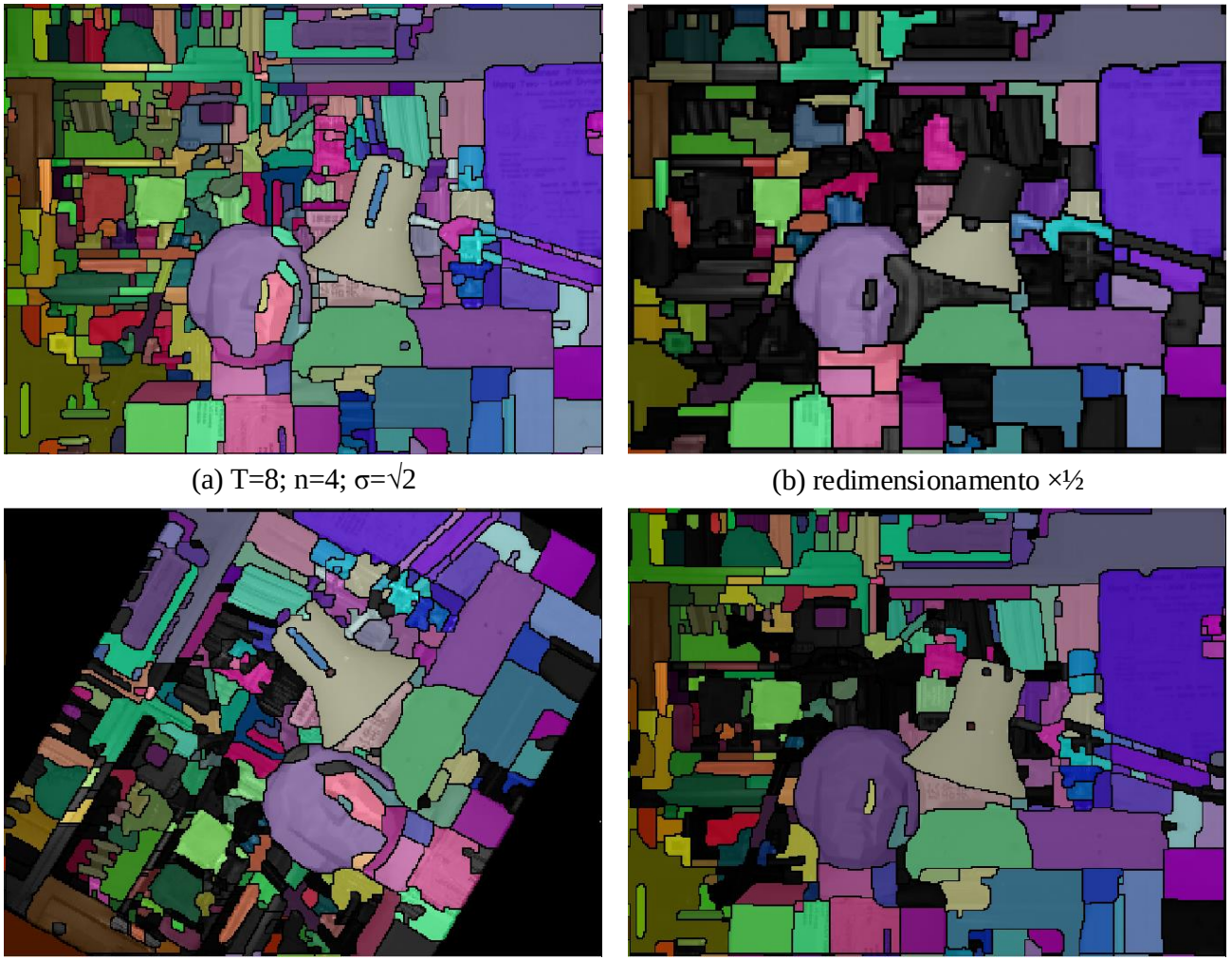


Figura 4.17: Ilustração para a determinação de regiões correspondentes entre duas imagens que tem regiões confrontadas com a aplicação do descritor proposto. $[1, 4]$ e $[3, 7]$ configuram pares de correspondência, pois formam um par recíproco na determinação do produto interno máximo no confronto de elementos da imagem V_1 com a V_2 e vice-versa.

As Figuras 4.18, 4.19, 4.20 e 4.21 exibem o confronto de regiões de imagens tratadas como referência (a) e suas versões em transformações geométricas de escala (b), orientação (c) e mudanças na iluminação (d). Percebe-se uma boa convergência entre as regiões da imagem original (a) para com seus respectivos pares de vistas sob transformações. No próximo Capítulo é definido um método de cálculo de fluxo óptico que refina esse confronto de regiões com a inclusão de informações de posição e cor das regiões.



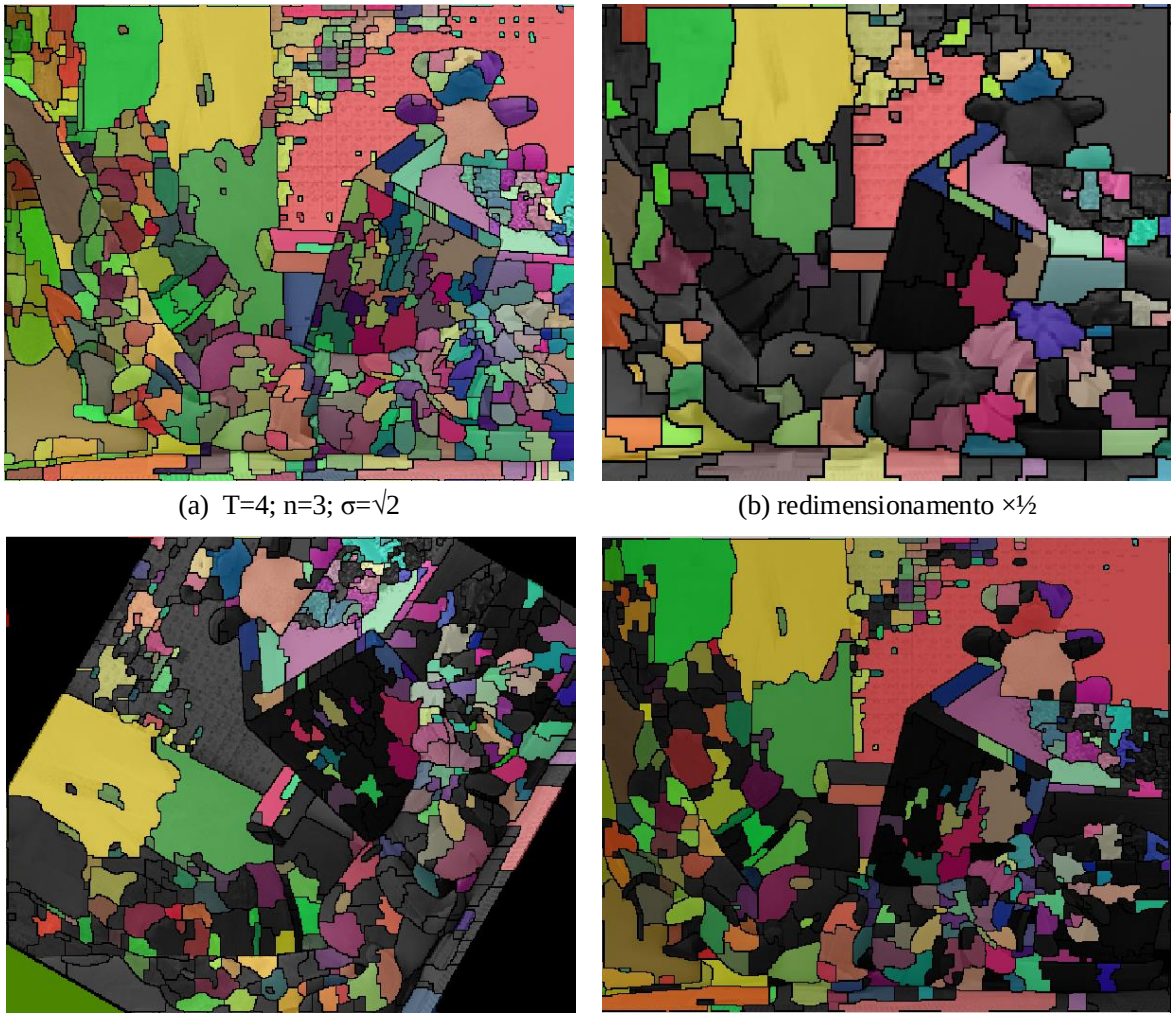
(a) T=8; n=4; σ=√2

(b) redimensionamento ×½

(c) rotação de 60º

(d) magnitude reduzida em 2×

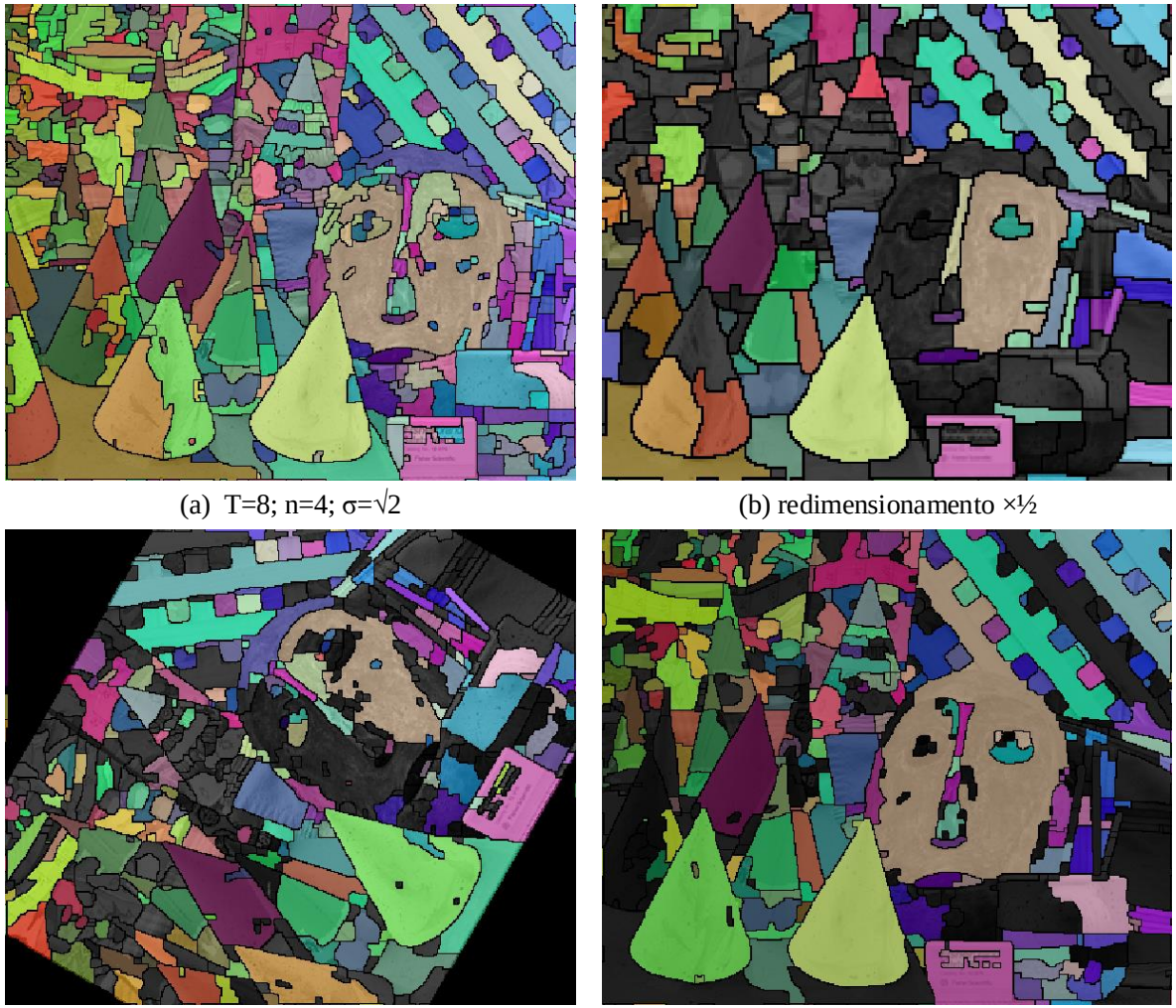
Figura 4.18: Confronto entre duas imagens da cena *Statue* uma em seu aspecto original e a outra submetida a três tipos de transformação: (a) uma vista em seus aspectos originais com regiões criadas pelo método de agrupamento proposto com parâmetros T, n e σ que são mantidos para as imagens confrontadas, todas as regiões recebem uma cor que se mantém para as regiões correspondentes nas imagens confrontadas; (b) a segunda vista tem dimensões reduzidas pela metade, exibida no tamanho original para melhor visualização; (c) aplicada uma rotação de 60^o e; (d) uma atenuação no seu mapa de magnitude na ordem de $2\times$.



(a) T=4; n=3; σ=√2 (b) redimensionamento ×½

(c) rotação de 60º (d) magnitude reduzida em 2×

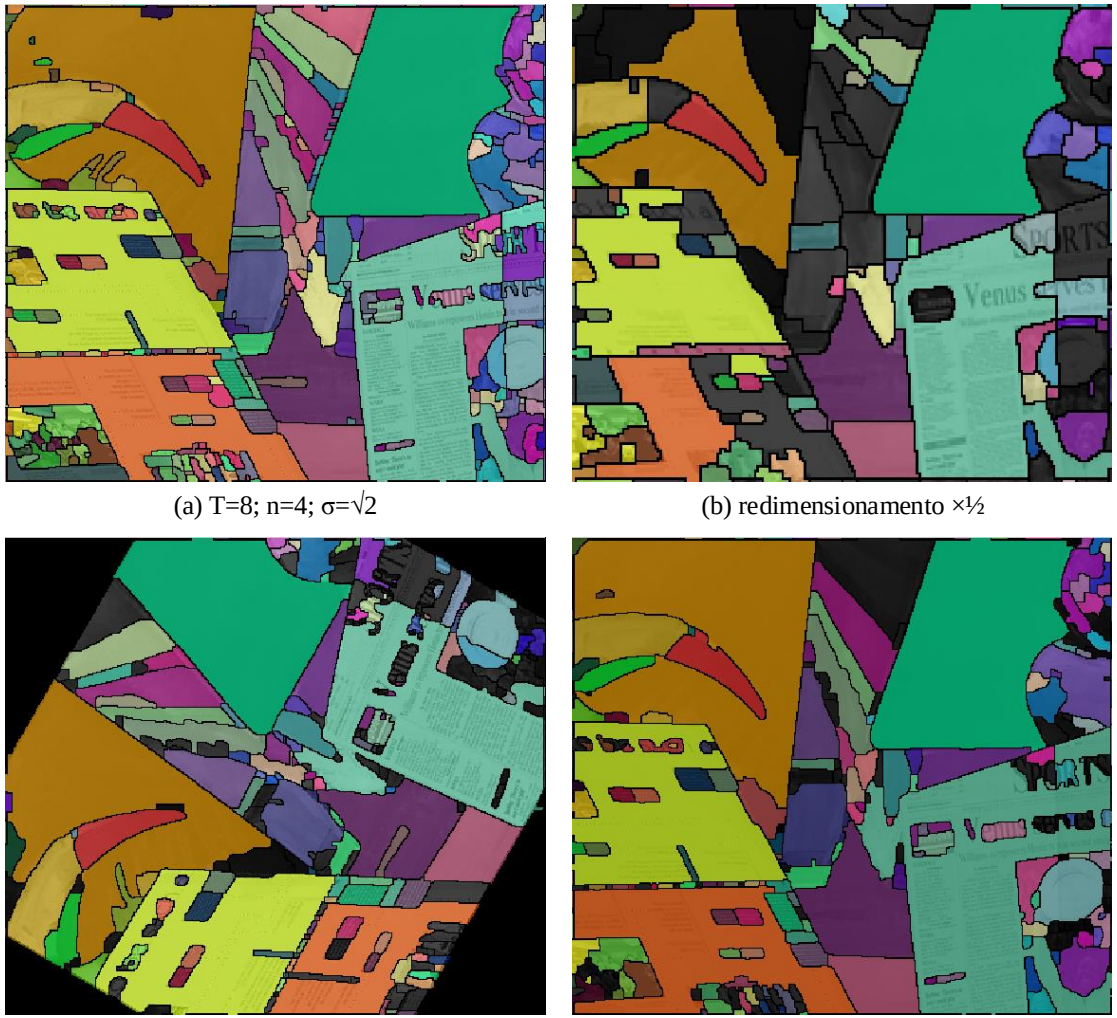
Figura 4.19: Confronto entre duas imagens da cena *Teddy* uma em seu aspecto original e a outra submetida a três tipos de transformação: (a) uma vista em seus aspectos originais com regiões criadas pelo método de agrupamento proposto com parâmetros T, n e σ que são mantidos para as imagens confrontadas, todas as regiões recebem uma cor que se mantém para as regiões correspondentes nas imagens confrontadas; (b) a segunda vista tem dimensões reduzidas pela metade, exibida no tamanho original para melhor visualização; (c) aplicada uma rotação de 60^o e; (d) uma atenuação no seu mapa de magnitude na ordem de $2\times$.



(a) T=8; n=4; σ=√2 (b) redimensionamento ×½

(c) rotação de 60º (d) magnitude reduzida em 2×

Figura 4.20: Confronto entre duas imagens da cena *Cones* uma em seu aspecto original e a outra submetida a três tipos de transformação: (a) uma vista em seus aspectos originais com regiões criadas pelo método de agrupamento proposto com parâmetros T, n e σ que são mantidos para as imagens confrontadas, todas as regiões recebem uma cor que se mantém para as regiões correspondentes nas imagens confrontadas; (b) a segunda vista tem dimensões reduzidas pela metade, exibida no tamanho original para melhor visualização; (c) aplicada uma rotação de 60^o e; (d) uma atenuação no seu mapa de magnitude na ordem de $2\times$.



(a) T=8; n=4; σ=√2 (b) redimensionamento ×½

(c) rotação de 60º (d) magnitude reduzida em 2×

Figura 4.21: Confronto entre duas imagens da cena *Venus* uma em seu aspecto original e a outra submetida a três tipos de transformação: (a) uma vista em seus aspectos originais com regiões criadas pelo método de agrupamento proposto com parâmetros T, n e σ que são mantidos para as imagens confrontadas, todas as regiões recebem uma cor que se mantém para as regiões correspondentes nas imagens confrontadas; (b) a segunda vista tem dimensões reduzidas pela metade, exibida no tamanho original para melhor visualização; (c) aplicada uma rotação de 60^o e; (d) uma atenuação no seu mapa de magnitude na ordem de $2\times$.

4.4.1 Ajuste fino de correspondências e estimativa de movimento

Foi desenvolvido um método que, a partir do descritor proposto, estima o movimento das regiões entre quadros, visando o aumento da exatidão nas convergências determinadas. A técnica utilizada se aproveita da capacidade do descritor em relacionar regiões em diferentes graus de transformação geométrica, ajustando e refinando os deslocamentos das regiões em um processo iterativo. O método iterativo proposto está ilustrado na Figura 4.22, na qual estima-se os deslocamentos das regiões de um quadro de referência (a) em relação ao seu subsequente (b).

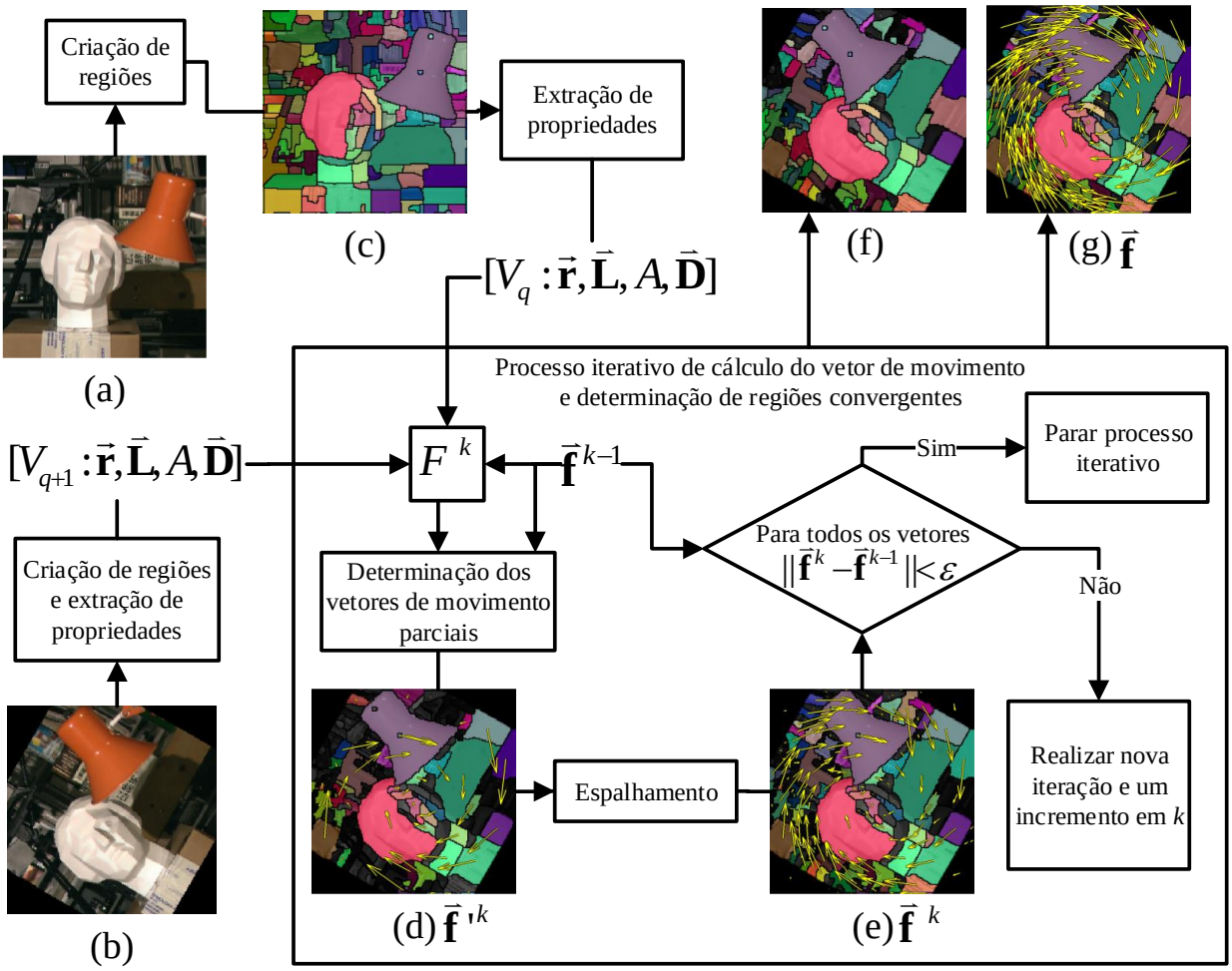


Figura 4.22: Diagrama ilustrando o processo de estimativa de movimento entre regiões: (a) uma imagem/quadro é tomada como referência, V_q; (b) o método proposto estima para um quadro subsequente, V_{q+1} (ilustrado como uma rotação do par estéreo da imagem de referência), o movimento em relação ao quadro de referência; (c) V_{q+1} é representada por regiões e parâmetros são extraídos delas, posições, áreas, cores e descritores, o mesmo é feito para V_q; (d) o processo iterativo calcula e atualiza as correspondências e o vetor de movimento parcial das regiões do quadro subsequente; (e) a informação dos vetores de fluxo definidos pelas correspondências é espalhada para as vizinhas a partir de uma filtragem; (f) após cerca de 15 interações, as correspondências e; (g) os vetores permanecem estáveis.

Uma vez determinado o descritor $\vec{\mathbf{D}}_i$, do elemento i pertencente ao quadro V_q, e o descritor $\vec{\mathbf{D}}_j$, do elemento i pertencente ao quadro subsequente V_{q+1}, um vetor de ajuste, ou vetor de movimento $\vec{\mathbf{f}}_j$, é calculado para corrigir a posição $\vec{\mathbf{r}}_j$ da região/elemento j em relação ao vetor posição $\vec{\mathbf{r}}_i$ do elemento i. O vetor de fluxo $\vec{\mathbf{f}}_j$ é obtido por um processo iterativo.

Calcula-se inicialmente uma força de conexão F^k entre os elementos para determinação das correspondências, que são atualizadas a cada iteração. O produto interno que define essa força de conexão entre os descritores, é então ponderado pela multiplicação de duas gaussianas, uma referente à distância entre as componentes de cor, $\vec{\mathbf{L}}_j$ e $\vec{\mathbf{L}}_j$, e outra referente à distância espacial entre os elementos, $\vec{\mathbf{r}}_i$ e $\vec{\mathbf{r}}_j$, ajustada pelo vetor de movimento $\vec{\mathbf{f}}_j^k$ na iteração k:

$$F_{i,j}^k = \vec{\mathbf{D}}_i \cdot \vec{\mathbf{D}}_j e^{-\left(\frac{||\vec{\mathbf{L}}_i - \vec{\mathbf{L}}_j||}{T}\right)^2} e^{-\left(\frac{(k-\vec{\mathbf{D}}_i \cdot \vec{\mathbf{D}}_j)||\vec{\mathbf{r}}_i - (\vec{\mathbf{r}}_j + \vec{\mathbf{f}}_j^k)||}{\overline{R}_i}\right)^2}. \tag{4.26}$$

A cada iteração, há na equação (4.26) um incremento de uma unidade em k, que é inicializado como 1. Nesta inicialização, $k = 1$, temos a seguinte aproximação para $F_{i,j}^1$:

$$F_{i,j}^1 \approx \vec{\mathbf{D}}_i \cdot \vec{\mathbf{D}}_j e^{-\left(\frac{||\vec{\mathbf{L}}_i - \vec{\mathbf{L}}_j||}{T}\right)^2}, \tag{4.27}$$

uma vez que os valores do produto interno de regiões semelhantes $\vec{\mathbf{D}}_i \cdot \vec{\mathbf{D}}_j$ estão próximos de 1, quando subtrai-se $k = 1$ por esse produto interno, o expoente da gaussiana referente a posição se aproxima de zero na equação (4.26), ou seja, na inicialização as correspondências não são influenciadas pela posição dos elementos, somente pela sua proximidade entre cores e descritores. Para todo $j \in V_{q+1}$, $\vec{\mathbf{f}}_j^0$ é inicializado com valor 0.

Em uma primeira iteração, por exemplo, para imagem ou quadro de referência e as regiões que a definem (Figura 4.22(a) e (b), respectivamente) determina-se para o quadro subsequente (Figura 4.22(c)) correspondências parciais (Figura 4.22(e)). Essas correspondências parciais são determinadas como descrito para as equações (4.24) e (4.25), encontra-se um par vencedor $[i, j] \in V_{wp}$ determinado para a função F, que representa o produto interno de descritores ponderado.

Encontradas as correspondências parciais para os vencedores (Figura 4.22(c)) atualiza-se um vetor de movimento parcial $\vec{\mathbf{f}}_j^k$ para um elemento j_w em relação ao seu par i como:

$$\vec{\mathbf{f}}_j'^k = \begin{cases} \vec{\mathbf{r}}_i - \vec{\mathbf{r}}_j, & \text{se } [i,j] \in V_{wp} \\ \vec{\mathbf{f}}_j^{k-1}, & \text{caso contrário} \end{cases}, \tag{4.28}$$

se $[i, j]$ não representam um par vencedor, $\vec{\mathbf{f}}_j'^k$ toma o valor do vetor na iteração anterior $\vec{\mathbf{f}}_j^{k-1}$.

Na primeira iteração (Figura 4.22(c)) há uma atualização do vetor de fluxo apenas para as correspondências, enquanto os elementos sem correspondências continuam com o argumento nulo. A informação referente aos vencedores é difundida para os elementos próximos (Figura 4.22(f))

e, por consequência, para elementos sem par de correspondência, calculando-se a média entre todos os pontos dentro de uma vizinhança:

$$\vec{\mathbf{f}}_j^k = \left(\sum_{\forall u \in V_{del}^j} \vec{\mathbf{f}}_u'^k A_u \right) / \left(\sum_{\forall u \in V_{del}^j} A_u \right). \tag{4.29}$$

A equação (4.29) opera de maneira semelhante a um filtro média móvel no domínio de uma imagem de pixels. Obtém-se uma média dos elementos em uma vizinhança, no caso da representação por regiões, utilizou-se a vizinhança V_{del}^j, que são os elementos conectados à j por uma triangulação *Delaunay* [43], incluindo o próprio j. Multiplicando-se cada vetor $\vec{\mathbf{f}}_u'^k$ por sua respectiva área A_u, cria-se uma soma ponderada que é média de vetores dentro da vizinhança; quanto maior a área da região, maior sua contribuição .

O crescimento de k a cada iteração (equação 4.26) restringe espacialmente as possibilidades de casamento. A gaussiana referente à distância espacial entre os elementos i e j, tendo j sua posição ajustada por $\vec{\mathbf{f}}_j^k$, determina essa restrição quando assume valores muito próximo de zero para elementos mais afastados de i.

A mudança nas correspondências passa a ser irrelevante a cada iteração, bem como a mudança no vetor de movimento antes (Figura 4.22(g)) e depois da filtragem (Figura 4.22(h)). O processo é encerrado quando de uma iteração para a outra, o máximo módulo do deslocamento dentre todos os elementos $j \in V_{q+1}$ é menor que ϵ:

$$||\vec{\mathbf{f}}_j^{k-1} - \vec{\mathbf{f}}_j^k|| < \epsilon, \tag{4.30}$$
$$\forall j \in V_{q+1}$$

Um valor ϵ unitário representa um deslocamento inferior a um pixel. Este valor foi adotado por se mostrar um bom critério de parada. Valores menores do que 1 para o ϵ se mostraram inatingíveis em algumas situações, e valores maiores implicam em incoerências espaciais, ou seja, correspondências erroneamente definidas. Os vetores final $\vec{\mathbf{f}}$ têm valores iguais a $\vec{\mathbf{f}}^k$ na última iteração.

4.4.1.1 Restrições

Visando maior exatidão no casamento de correspondências e, por consequência, no vetor de movimento determinado, duas restrições para o casamento de regiões foram empregadas. Tais restrições têm contribuição para a diminuição no tempo de execução do algoritmo, uma vez aumentado o grau de exatidão para as primeiras iterações.

Foi determinada uma restrição quanto à distância entre a posição dos i e j, $\vec{\mathbf{r}}_i$ e $\vec{\mathbf{r}}_j$, em relação a suas dimensões, $\overline{R}_i$ e $\overline{R}_j$. F na equação (4.26) assume valores nulos caso a seguinte condição

seja respeitada:

$$||\vec{\mathbf{r}}_i - (\vec{\mathbf{r}}_j + \vec{\mathbf{f}}_j^k)|| - \rho(\overline{R}_i + \overline{R}_j) > 0. \tag{4.31}$$

A restrição em (4.31) visa eliminar casamentos de regiões que tenham suas fronteiras muito distantes umas das outras. Estima-se a distância entre as fronteiras subtraindo-se o vetor posição entre do centro das regiões, $\vec{\mathbf{r}}_i$ e $\vec{\mathbf{r}}_j$ corrigido pelo vetor de fluxo $\vec{\mathbf{f}}_j^k$ na iteração k, pela soma dos raios equivalentes, $\overline{R}_i$ e $\overline{R}_j$ (Figura 4.23), multiplicada por um fator ρ. Adotou-se o valor $\rho = 4$, por exibir uma boa captura de movimentos relevantes dos objetos, e, ao mesmo tempo, excluir falsos positivos que possam influenciar no cessar de iterações do processo.

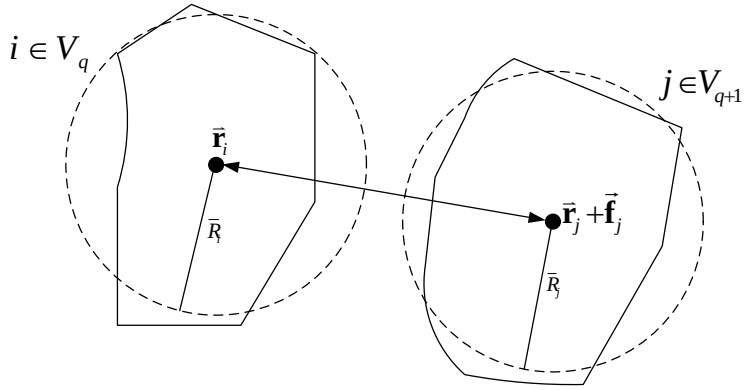


Figura 4.23: Ilustração para a restrição espacial para casamento de regiões, determinada pela equação 4.31. Apesar de estarem representadas em um mesmo plano, as regiões i e j pertencem a dois quadros distintos V_q e V_{q+1}, respectivamente. Após a correção da posição de j com um vetor de fluxo $\vec{\mathbf{f}}_j^k$, nota-se que as fronteiras da região não se interceptam, indicando que regiões não se tratam de correspondências. Uma forma de se aproximar e simplificar essa relação de fronteiras é aproximando as regiões por circunferências de raio igual ao equivalente $\overline{R}_i$ e $\overline{R}_j$, representado pelas circunferências tracejadas ao redor do centroide das regiões.

Outra restrição aplicada é para a mudança de escala da região, ou seja, um aumento ou diminuição de área de um quadro para o outro que não se adéqua a um padrão estabelecido, leva a função F da equação (4.26) a valores nulos, pela condição:

$$\frac{|\overline{A}_i - \overline{A}_j|}{\overline{A}_i + \overline{A}_j} > \nu. \tag{4.32}$$

Quando as áreas A_i e A_j, referentes as regiões i e j, respectivamente, têm uma área muito próxima, o valor do lado esquerdo da inequação (4.32) tente a 0, enquanto para áreas com grande desproporcionalidade esse valor tende a 1.

Quando, por exemplo, a área A_j assume o dobro do valor de A_i, $A_j = 2A_j$, ou A_i assume o dobro do valor de A_j, o lado direito da inequação (4.32) assume valor $1/3$. Adotou-se um $\nu = 0,4$ que permite o casamente de regiões que sofram deformações máximas em área de aproximadamente $2,5$ vezes. O valor visa contemplar as transformações em escala e a instabilidade gerada pela segmentação *watershed*.

4.4.2 Correspondências e sementes

Para experimentos e testes dos métodos propostos, foi utilizado um algoritmo de segmentação de simples implementação, o *GrowCut* [44]. Esse método, apresentado no Capítulo 3, utiliza sementes para inicialização e rotulação de elementos que são pertencentes a um objeto (OB) ou ao fundo da imagem (BK). O método foi originalmente desenvolvido para ser aplicado em uma imagem, em que as sementes s são pixels de inicialização, que dão suporte ao agrupamento de novos pixels a cada iteração, para a segmentação do objeto de interesse.

Para este trabalho, as sementes têm origem no primeiro quadro da sequência a ser segmentada, ou seja, todos os elementos do primeiro quadro possuem uma rotulação prévia, se pertence ao conjunto OB ou ao BK. Essa rotulação é determinada pelo *groud truth* (GT) desse primeiro quadro. A Figura 4.24 demonstra a propagação das sementes relativas ao objeto (círculos vermelhos) e ao fundo (marcações 'x'em azul), que tem origem no primeiro quadro (esquerda superior). Como o rastreamento é feito quadro a quadro, essa propagação de sementes e as correspondências indicadas pela padronização de cores são exibidos em pares de quadros na (Figura 4.24).

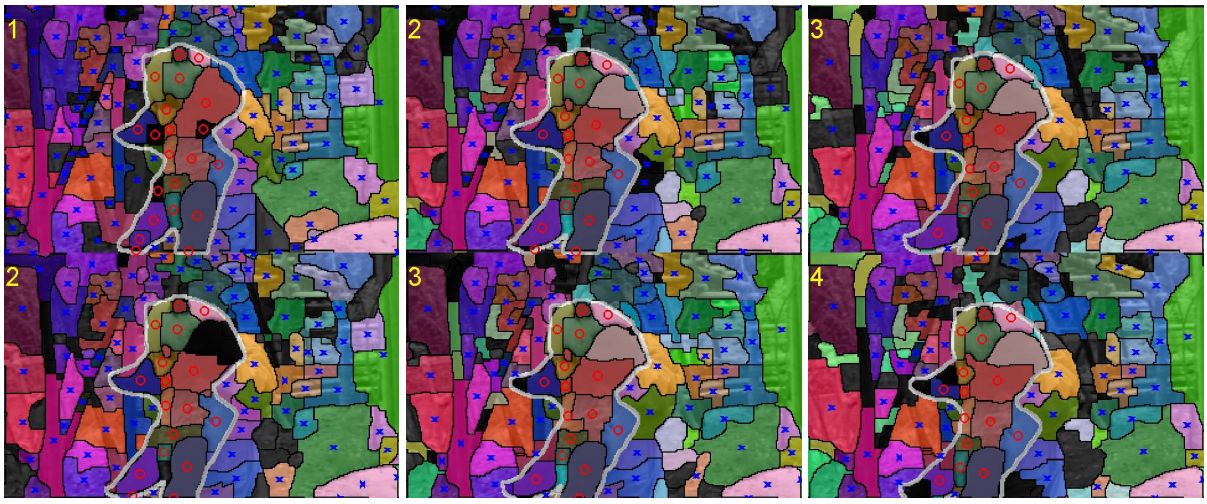
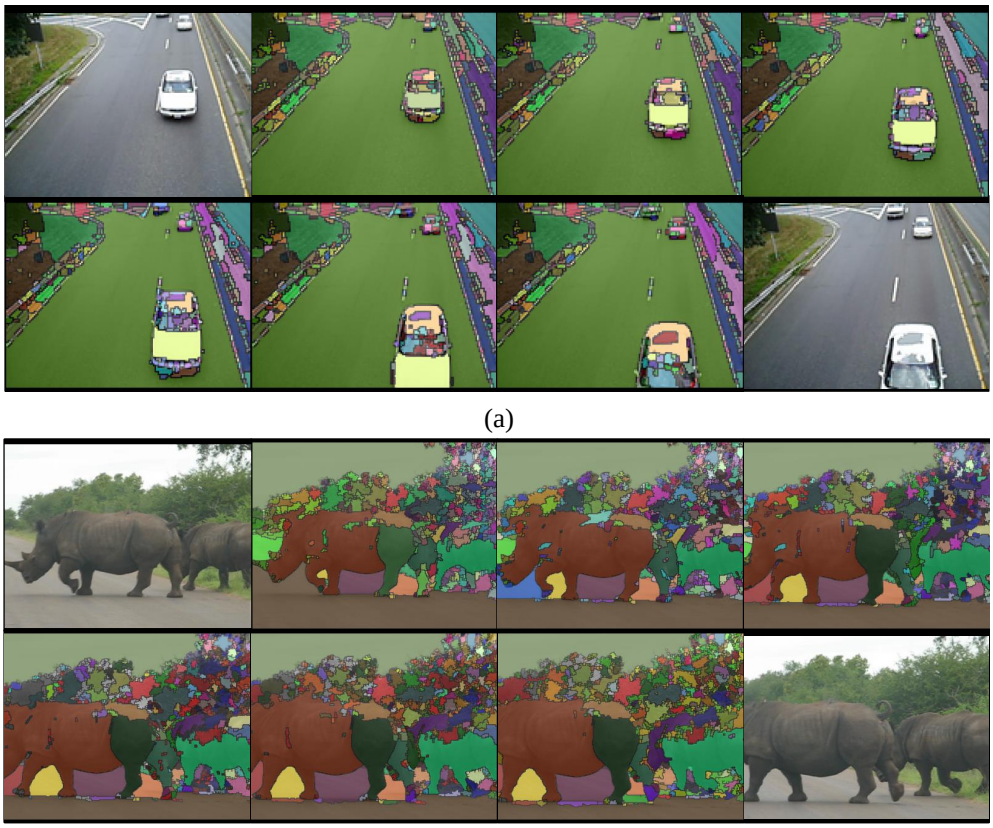


Figura 4.24: Casamento de regiões entre quadros, com correspondências atribuídas pelo algoritmo proposto. Os círculos vermelhos e as marcações em 'x'azul apontam as regiões referentes às sementes, com origem no primeiro quadro, que são perpetuados ao longo da sequência pelo casamento de regiões. Nos pares de confronto, 1-2, 2-3 e 3-4, as regiões correspondentes ganham o mesmo padrão de cores. As regiões sem correspondência no quadro subsequente analisado, como região escurecida referente ombro esquerdo do urso no quadro número 2 inferior, recebem uma nova rótulo e uma nova coloração/rótulo. O GT de todos os quadros está representado como uma contorno esbranquiçado em torno do urso.

Uma forma de se observar a propagação e casamento de regiões ao longo de quadros (Figura 4.25) é a representação dessas regiões distribuídas em forma de volumes ao longo do tempo, $x \times y \times$ tempo (Figura 4.26). Ao contrário da Figura 4.24 a regiões da sequência da Figura 4.25 não possuem regiões escurecidas, que é uma forma de ilustrar no confronto regiões sem cor-

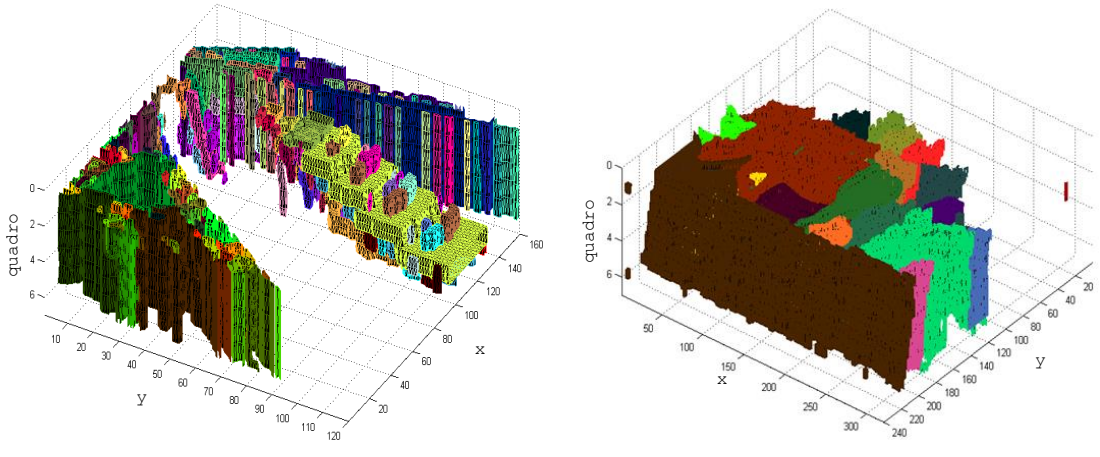
respondências para o quadro subsequente, todas as regiões exibem um padrão de cor. Com um mesmo padrão de cores para regiões correspondentes (Figura 4.25) é possível ilustrar o casamento de regiões em um volume espaço×tempo (Figura 4.26).



(a)

(b)

Figura 4.25: Ilustração do casamento de regiões em duas sequências de 6 quadros, regiões as quais recebem um mesmo padrão de cores para representar uma correspondência: (a) sequência *Traffic* e; (b) sequência *Rhino*.



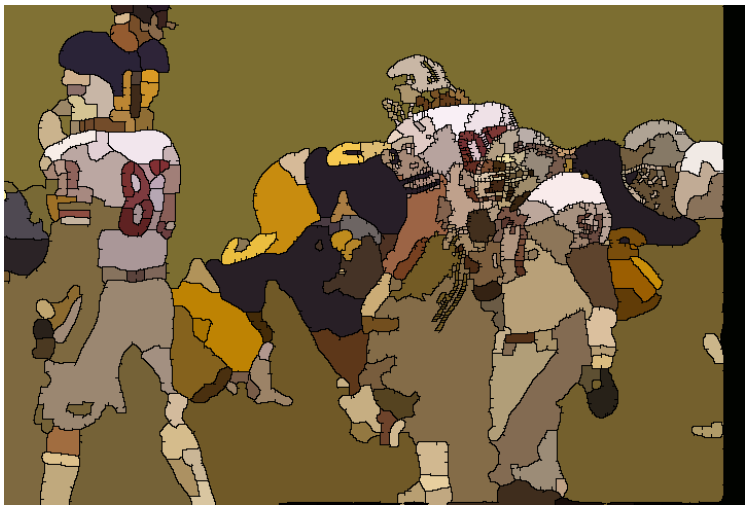
(a) (b)

Figura 4.26: Representação 3D das regiões que formam as sequências *Traffic* e *Rhino* da Figura 4.25(a) e (b), respectivamente. Os volumes exibidos nas imagens (a) e (b), *Traffic* e *Rhino*, representam regiões relevantes propagadas ao longo do tempo.

4.5 ESCALA MISTA ORIENTADA AO OBJETO

Aproveitando informações referentes ao posicionamento e tamanho do objeto fornecida pelo primeiro quadro e seu GT, é possível determinar regiões mais robustas e extensas para pontos mais afastados do objeto. Um processo inspirado na fisiologia do olho humano, que tem uma concentração de receptores na região da fóvea (Figura 4.27(a)), força uma distribuição de elementos de forma não uniforme, em que para objeto desejado há uma concentração maior de elementos/regiões por unidade de área, enquanto para as vizinhanças desse objeto, uma concentração menor de regiões é definida (Figura 4.27(b)) e regiões mais amplas são formadas.

(a)



(b)

Figura 4.27: Escala mista aplicada a uma imagem (*Football*) com foco no capacete do jogador: (a) a combinação de imagens borradas por Gaussianas de diferentes desvios padrão, produz semelhante ao que se tem quando se foca um objeto com o olhar, no caso, cabeça e capacete de jogador; (b) aplicando em (a) os métodos para criação de regiões desenvolvidos no Capítulo 4, obtém-se uma alta concertação no objeto de interesse e um maior número de elementos para representar áreas afastadas do objeto.

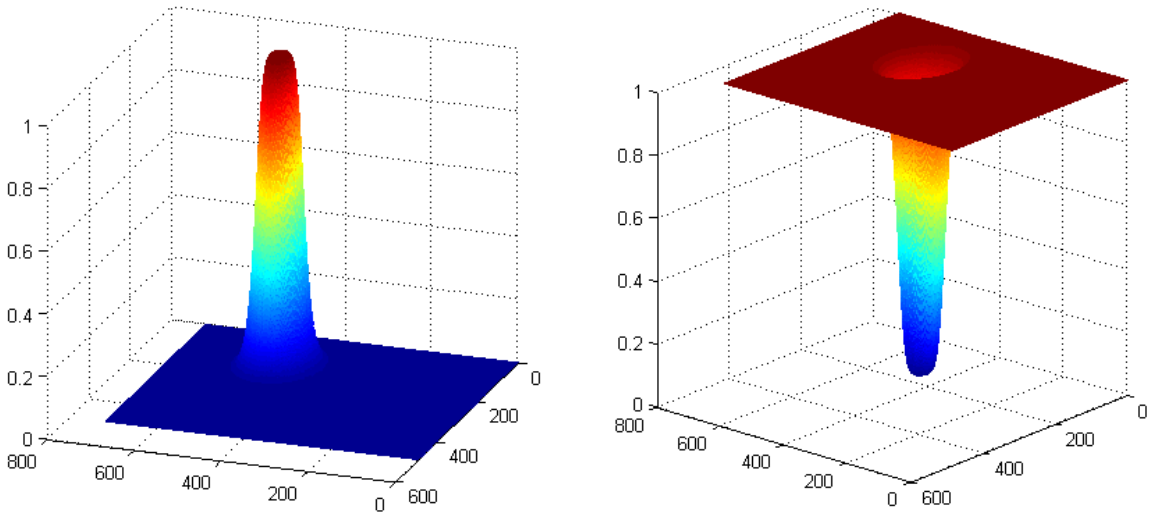
A técnica de escala mista consiste na adaptação de um passo dos métodos apresentados no Capítulo 4, modificando-se a equação (4.10) para a criação das regiões de forma que:

$$L(x,y,\sigma) = H(x,y)\big(I(x,y) * G(x,y,\sigma)\big) + \big(1 - H(x,y)\big)\big(I(x,y) * G(x,y,2\sigma)\big)\big), \quad (4.33)$$

em que H é uma máscara *Butterworth* tal que:

$$H(x,y) = \frac{1}{1 + \left(\frac{(x-x_s)^2+(y-y_s)^2}{2R}\right)^6}. \quad (4.34)$$

A equação (4.33) representa a filtragem de $I(x,y)$ por gaussianas de diferentes desvios padrões, $G(x,y,\sigma)$ e $G(x,y,2\sigma)$, que são combinadas por meio de uma janela *Butterworth*, $H(x,y)$ (Figura 4.28(a)), e seu complemento, $1 - H(x,y)$ (Figura 4.28(b)). A filtragem pela gaussiana de menor desvio padrão σ é multiplicada pela *Butterworth* centrada em (x_s, y_s), ressaltando o objeto, com detalhes menos desfocados que o fundo, para o qual o borramento é mais intenso e salientado pelo complemento da *Butterworth*.



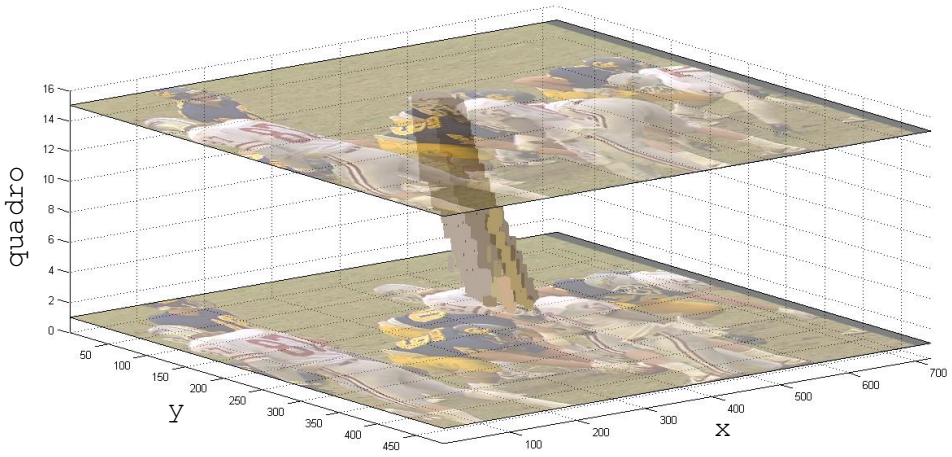
(a) $H(x,y)$ (b) $1-H(x,y)$

Figura 4.28: Janelas utilizadas para combinação de escalas: (a) *Butterwoth* relacionada a imagem da Figura 4.27(a), ponto de máximo e amplitude da janela estão vinculadas à posição do objeto e sua área em imagem, respectivamente; (b) representação do complemento de (a), ou seja, uma função constante 1 subtraída da *Butterwoth*.

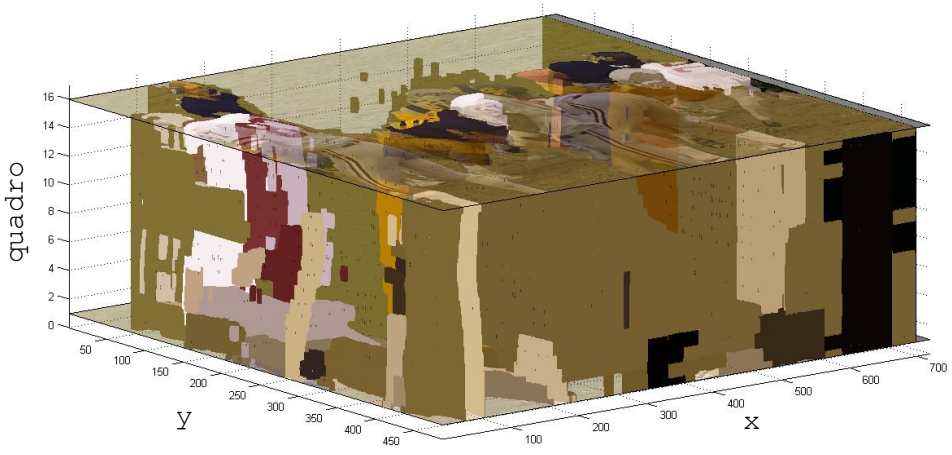
Na equação (4.33) a posição (x_s, y_s) é uma estimativa para a localização do centro do objeto. No casamento de regiões quadro a quadro adotada neste trabalho, apenas o primeiro quadro de uma sequência tem esta localização exata, determinada pelo GT. Nos demais quadros, estima-se (x_s, y_s) pelo centroide formado pelas sementes referentes ao objeto de um quadro. No caso de movimento, por exemplo, é como se o foco (região mais bem resolvida) estivesse sempre em atraso quanto à posição do objeto.

O valor 2 que multiplica o raio equivalente $\overline{R}$ na equação (4.34), aumenta a abrangência da área menos borrada, ou mais bem definida, de forma a compensar algum tipo de movimento ou mudança de dimensão do objeto. Fixou-se $\overline{R} = R_{gt}$ ao longo de toda sequência, onde R_{gt} representa o raio equivalente da área do objeto para o GT do primeiro quadro.

Foi observado em testes que, para um objeto em movimento, as sementes vão se extinguindo para ao longo de uma sequência de quadros (Figura 4.29 (a)), devido à falta de correspondências, consequente das divergências entre as imagens conforme o movimento. O mesmo acontece para as sementes referentes ao fundo (Figura 4.29 (b)), em maior volume devido à variação de escala.



(a)



(b)

Figura 4.29: A representação 3D do esquema de regiões em uma imagem de escala mista (Figura 4.27(a)) pode ser dividida nos conjuntos mais importantes para a segmentação proposta neste trabalho: (a) sementes referentes ao objeto (OB) propagadas ao longo de um grupo de quadros, verifica-se uma maior concentração de elementos para este conjunto do que para: (b) conjunto formado por sementes do fundo (BK).

4.5.1 Deslocamento normalizado do centroide

Não foram definidas métricas específicas para avaliar a coerência espaço-temporal do casamento de regiões ao longo dos quadros. Para avaliar como o movimento do objeto ao longo dos quadros pode influenciar a sua segmentação, foram construídos gráficos para o deslocamento normalizado do GT do objeto nas cenas estudadas. As componentes do deslocamento normalizado em termos absolutos do centroide (xc, yc) de objeto, segundo GT, entre dois quadros, q e $q+1$, é calculado como:

$$(dx_Q, dy_Q) = \frac{1}{RM_{GT}} |(xc_{q+1}, yc_{q+1}) - (xc_q, yc_q)| \tag{4.35}$$

em que RM_{GT} é a média dos raios equivalentes das regiões pertencentes ao GT:

$$RM_{GT} = \frac{1}{N_{GT}} \sum_{i \in GT} \sqrt{\frac{A_i}{\pi}} \tag{4.36}$$

O módulo do deslocamento normalizado do objeto de um quadro para o outro é a soma euclidiana das componentes do deslocamento.

$$d_Q = \sqrt{dx_Q^2 + dy_Q^2} \tag{4.37}$$

O objetivo desses deslocamentos normalizados é verificar o quanto em média as fronteiras das regiões estão afastando, de acordo com o movimento do objeto. Quando se normaliza o deslocamento pela média dos raios equivalentes da região do GT do objeto, RM_{GT}, esse deslocamento fica em termos desse raio médio. Essas curvas de deslocamento, calculados para cada sequência, são exibidos no Capítulo 6.

5 MAPAS DE PESOS E SEGMENTAÇÃO DE VÍDEOS VIA CORTES EM GRAFOS

5.1 INTRODUÇÃO

Neste capítulo serão apresentados os métodos utilizados para proceder a segmentação de objetos em vídeo via corte de grafos. Para avaliar as contribuições do descritor proposto, serão detalhados quatro modos de organização e ponderação de grafos, que serão aplicados em dois modos de segmentação, quadro a quadro e ao longo de um grupo de quados. A primeira forma de conectar e ponderar vértices de um grafo, representa fornecendo-se peso às ligações a partir da análise das distâncias entre posições e cores de forma direta (não ajustado – NA), a segunda utiliza o descritor proposto para realizar uma estimativa de movimento entre quadros (Ajustado – AJ), a terceira conecta os vértices correspondentes reforçando-se suas ligações com um peso infinto (Reforçado – RE); e a quarta agrupa as correspondências em vértices equivalentes (Equivalente – EQ).

5.2 ORGANIZAÇÃO DOS GRAFOS E DETERMINAÇÃO DOS MAPAS DE PESOS

Por intermédio da sobre-segmentação em regiões, obtida pelos métodos propostos no Capítulo 4, e a definição de regiões correspondentes também pelos métodos propostos, um vídeo pode ser interpretado como grafo, cujas relações entre elementos serão estudadas em favor da análise da contribuição do descritor local proposto para as regiões.

A partir desse ponto, as regiões são tratadas como vértices de um grafo que se estendem ao longo do tempo, em que o peso das relações entre esses vértices é determinado por características dessas regiões, como posição, cor, área e um vetor de movimento, estimado de acordo com o proposto no Capítulo 4.

Quatro tipos de grafos foram adotados para estudos, separados de acordo com organização e pesos de conexões entre as regiões entre quadros: (1) sem ajuste de movimento entre regiões (NA) (Figura 5.1(a)); (2) com ajuste no movimento entre regiões (AJ) (Figura 5.1(b)); (3) regiões correspondentes com pesos reforçados (RE) (Figura 5.1(c)); e (4) regiões correspondentes substituídas por elementos equivalentes (EQ) (Figura 5.1(d)).

A Figura 5.1 exibe as quatro variantes de organização de grafos, o grafo com elementos sem ajuste de movimento entre regiões, NA (Figura 5.1(a)), o grafo em que os elementos recebem um ajuste de movimento, AJ (Figura 5.1(b)), definido pela estimativa de movimento determinada no Capítulo 4. A ilustração exibe também o grafo no qual reforça-se ligações entre as correspondências, RE (Figura 5.1(c)), e aquele em se representa as correspondências por um único elemento equivalente, EQ (Figura 5.1(d)).

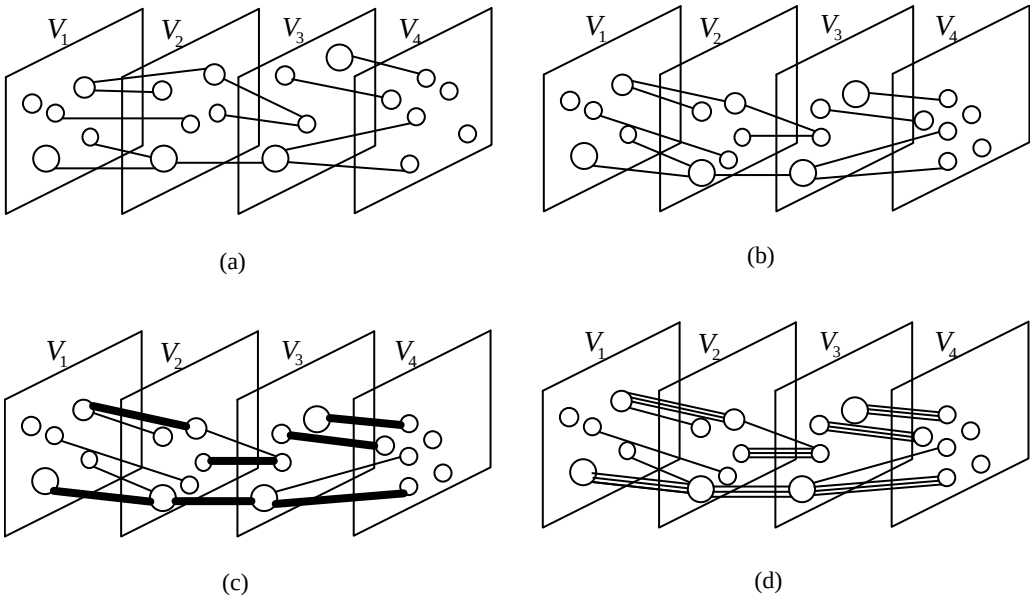


Figura 5.1: Ilustração para os quatro modos de grafos aplicados às sequências de vídeo estudadas: (a) o modo sem ajuste (NA) relaciona as regiões sem qualquer interferência na posição espacial das mesmas; (b) no modo ajustado (AJ), corrige-se a posição das regiões de um quadro em relação ao antecessor, por meio da estimativa de movimento proposta, tal correção é ilustrada pela reorganização da posição dos elementos em cada quadro em relação a disposição exibida em (a); (c) no grafo com pesos reforçados (RE), define-se uma ligação de peso infinito para as correspondências determinadas pelo algoritmo proposto, essa representação é feita pelas linhas mais espessas que ligam elementos de quados distintos; (d) para um grafo equivalente (EQ), as ligações triplas representam aqueles elementos considerados como um só, definidos pelas correspondências encontradas pelo algoritmo proposto.

Os casos RE e EQ são construídos com base no mapa de pesos do grafo AJ. Uma vez corrigido o movimento entre regiões e determinada os pesos de ligação dessas regiões, agora tratadas como vértices de um grafo, o grafo RE é construído ao se criar um ligações de pesos infinito entre aqueles elementos ditos correspondentes, pesos infinitos representados por linhas mais espessas na Figura 5.1(c).

No caso do grafo EQ, em vez dos pesos infinitos, trata-se essas regiões correspondentes como um único vértice, um vértice equivalente. Na Figura 5.1(d), por exemplo, as ligações triplas conectam regiões tratadas como um único vértice em um grafo equivalente, que é um grafo com menos elementos do que RE, mas que preserva suas propriedades para uma segmentação via corte de grafo.

A intenção da proposta dessas quatro formas de se organizar grafos é verificar a contribuição do descritor, uma vez que a ferramenta proposta é utilizada para se estimar o movimento das regiões entre quadros, podendo-se comparar uma segmentação aplicada a uma relação direta, sem ajuste (NA), a uma relação na qual as posições entre as regiões de um quadro para outro são ajustadas (AJ), antes da determinação da força de ligação entre elas.

As outras duas formas de avaliar as contribuições do descritor, envolvem como ele é capaz de relacionar as regiões entre quadros por meio das correspondências. Utiliza-se o mapa de pesos do grafo AJ, no qual diferencia-se a força de ligação entre aquelas regiões/vértices determinados como correspondentes (RE), ou simplesmente agrupa-se essas correspondências em vértices únicos, vértices equivalentes (EQ).

5.2.1 Grafos sem ajuste de movimento entre regiões

O princípio básico da segmentação de imagem/vídeo via grafos é utilizar informações referentes à cor e a posição das dos elementos que a compõe, pixels ou regiões, para a formação dos mapas de ponderação. Em vídeo, a contribuição da posição dos elementos no mapa de ponderação pode ser feita de maneira direta, medindo-se a distância euclidiana dos elementos intra quadros, sem se levar em consideração movimento que um objeto pode ter sofrido na passagem de um quadro para outro.

Em uma das quatro formas apresentadas neste trabalho para se relacionar as regiões de um vídeo sobre-segmentado, não se corrige o movimento entre essas regiões (NA). A força de ligação $w_{i,j}^{NA}$ entre o elemento i, pertencente ao conjunto de regiões de um quadro V_q, é calculada em relação ao elemento j, pertencentes quadro seguinte V_{q+1}. Utilizando-se no cálculo elementos de cor $\vec{\mathbf{L}}$ e posição $\vec{\mathbf{r}}$, forma-se uma matriz de pesos com componentes fornecidas pela equação:

$$w_{i,j}^{NA} = \begin{cases} e^{-\left(\frac{||\vec{\mathbf{L}}_i - \vec{\mathbf{L}}_j||}{2T}\right)^2} e^{-\left(\frac{||\vec{\mathbf{r}}_i - \vec{\mathbf{r}}_j||}{\overline{R}_i + \overline{R}_j}\right)^2}, & \text{se} \quad i \in V_q \text{ e } j \in V_{q+1} \\ 0, & \text{caso contrário} \end{cases}, \tag{5.1}$$

em que T é o limiar definido para a segmentação e determinação das regiões do quadro (Capítulo 4), $\overline{R}_i$ e $\overline{R}_j$ são os raios equivalentes das regiões, determinadas pela expressão $\overline{R} = \sqrt{A_i/\pi}$.

A equação 5.1 tem valor determinado pela multiplicação de duas Gaussianas, caso os vértices i e j pertençam a dois quadros subsequentes, e zero, caso se tratem de regiões/vértices de um mesmo quadro. Regiões de um mesmo quadro recebem pesos nulos entre si, partindo do princípio que o objeto já é bem definido na sobre-segmentação realizada, e regiões semelhantes devem ser relacionadas ao longo dos quadros subsequentes. Esses pesos nulos representam uma grande região esparsa na matriz de pesos (Figura 5.2).

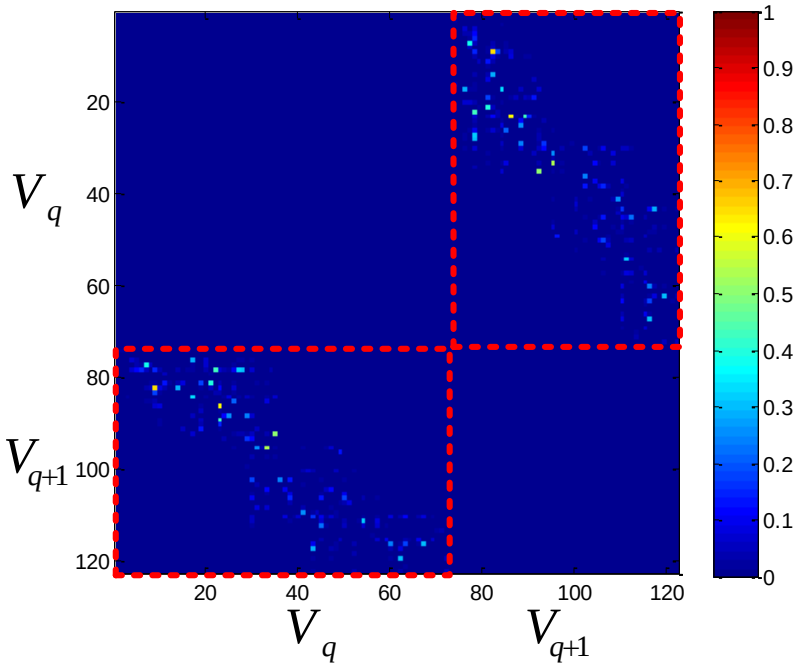


Figura 5.2: Mapa de pesos do grafo NA representado como um mapa de magnitudes. A relação se restringe aos conjuntos de elementos de dois quadros, V_q e V_{q+1}, entretanto a representação pode ser estendida para um mapa que contemple elementos de toda uma sequência. O mapa possui valores nulos para conexões de elementos no mesmo quadro, as regiões no mapa destacadas por retângulos com linhas pontilhadas representam as relações entre o quadro V_q e seu sucessor V_{q+1}.

5.2.2 Grafo com ajuste de movimento entre regiões

Para exemplificar o processo, pode-se aproveitar o mesmo exemplo de cálculo de vetor de movimento exibido na Figura 4.22(g), que contém os vetores que representam deslocamento estimado das regiões entre uma imagem e seu par estéreo rotacionado (Figura 5.3(a)). O efeito do ajuste/correção da posição dos elementos de uma imagem a partir do vetor de movimento, pode ser observado melhor em ilustração contendo esferas com raios iguais aos raios equivalentes $\overline{R}$ das regiões (Figura 5.3(b)), em mesma posição. Pode-se tratar a imagem original como um quadro de referência V_q, em que o vetor de movimento $\vec{\mathbf{f}}$ ajusta a posição dos elementos pertencentes a imagem rotacionada, um quadro subsequente V_{q+1} (Figura 5.3(c)).

Define-se pesos $w_{i,j}^{AJ}$ para a matriz de ponderação do caso AJ como:

$$w_{i,j}^{AJ} = \begin{cases} e^{-\left(\frac{||\vec{\mathbf{L}}_i - \vec{\mathbf{L}}_j||}{2T}\right)^2} e^{-\left(\frac{||\vec{\mathbf{r}}_i - (\vec{\mathbf{r}}_j + \vec{\mathbf{f}}_j)||}{\overline{R}_i + \overline{R}_j}\right)^2}, & \text{se} \quad i \in V_q \text{ e } j \in V_{q+1} \\ 0, & \text{caso contrário} \end{cases}, \tag{5.2}$$

em que as variáveis envolvidas são as mesmas que no caso NA da equação 5.1, com adição do vetor de movimento $\vec{\mathbf{f}}_j$ que corrige a posição $\vec{\mathbf{r}}_j$ do elemento $j \in V_{q+1}$ em relação a um quadro de referência V_q. Uma vez ajustadas as posições de acordo com o vetor de movimento estimado,

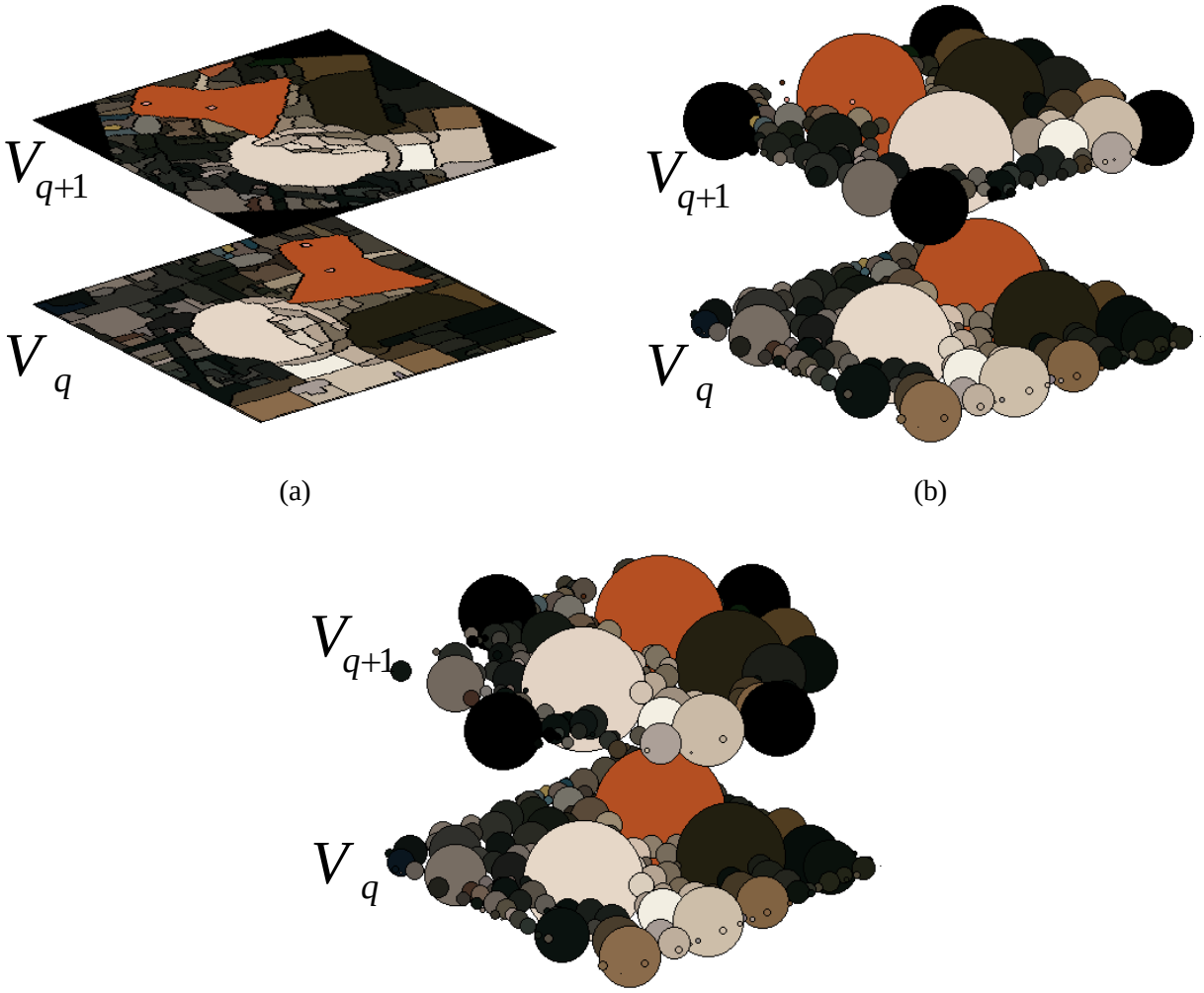


Figura 5.3: Ilustração da correção no movimento entre regiões proposta: (a) regiões são definidas para um quadro V_q e um quadro subsequente V_{q+1}, ilustrados por duas vista de uma cena, uma em seu aspecto origina e outra rotacionada; (b) representando as regiões de cada quadro como esferas de raio proporcional ao seu raio equivalente $\overline{R}$ é observar melhor o ajuste efetuado; (c) a correção na posição dos elementos de V_{q+1}, os posiciona em convergência com possíveis correspondências no quadro de referência V_q.

anula-se elementos que entre quais:

$$||\vec{\mathbf{r}}_i - (\vec{\mathbf{r}}_j + \vec{\mathbf{f}}_j^k)|| > 2(\overline{R}_i + \overline{R}_j). \tag{5.3}$$

Considera-se desconectados aqueles elementos que a distância das posições $\vec{\mathbf{r}}_i$ e $\vec{\mathbf{r}}_j$, mesmo quando ajustados pelo vetor de movimento $\vec{\mathbf{f}}_j$, estejam a uma distância maior que duas vezes a soma dos seus raios equivalentes $\overline{R_i}$ e $\overline{R_j}$. Essa consideração visa desconectar regiões não que tenham suas fronteiras próximas, mesmo com uma correção pelo vetor de movimento. O mesmo princípio não pode ser aplicado para o caso NA, pois sem informação quanto o movimento do objeto, pode-se desconectar regiões que ultrapassem qualquer limiar pré-definido.

Quando ajustado o movimento percebe-se um aumento de intensidade na força de ligação de alguns elementos (Figura 5.4) em relação ao caso não ajustado (Figura 5.2). Essa condição está relacionado a uma melhor convergência das posições das regiões aliado a sua proximidade em componentes de cor.

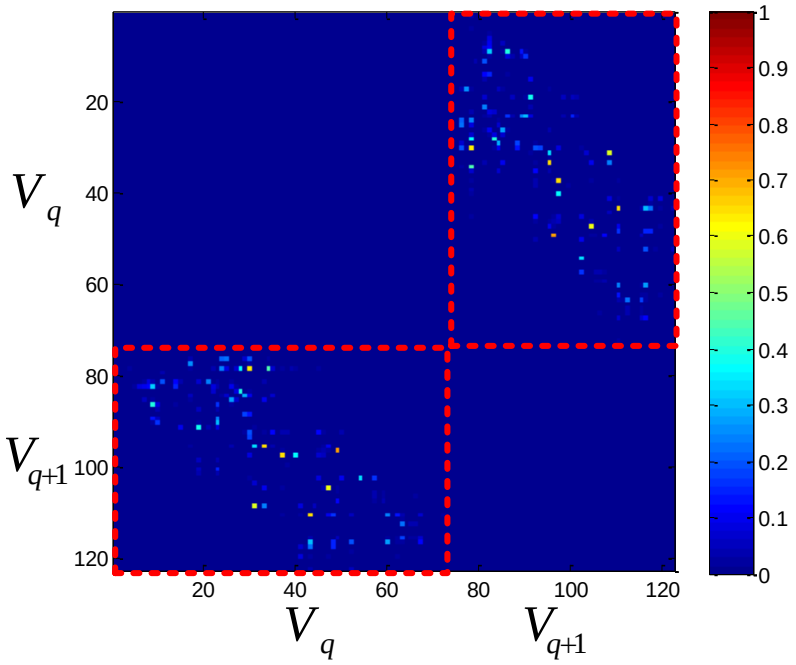


Figura 5.4: Mapa de pesos de um grafo AJ representado como um mapa de magnitudes. As relações são as mesmas para que as do mapa exibido na Figura 5.2, entretanto nota-se uma mudança na magnitude das componentes do mapa, oriunda da do ajuste na posição das regiões do quadro V_{q+1} em relação a V_q, que implica na mudança nos pesos em relação ao mapa NA.

5.2.3 Correspondências com pesos reforçados

A partir do último grupo de correspondências criadas para o cálculo do vetor de movimento (Figura 5.3(a)) pode-se determinar um grau maior de associação entre esse grupo de elementos. Repete-se o mesmo método de cálculo de pesos para o modelo AJ, substituindo-se os pesos dos pares de vencedores, ou agora, elementos equivalentes $[i, j] \in V_e$, por pesos $w_{i,j}^{RE}$ com valor muito grande, tratados infinitos:

$$w_{i,j}^{RE} = \begin{cases} \infty, & \text{se } [i,j] \in V_e, \\ w_{i,j}^{AJ}, & \text{caso contrário} \end{cases} \tag{5.4}$$

Este método de mapeamento de pesos cria uma força de ligação tal, que os elementos conectados por esses pesos infinitos passam a representar um único elemento. Entretanto, esse aumento no peso das ligações desses vértices não elimina as redundâncias criadas.

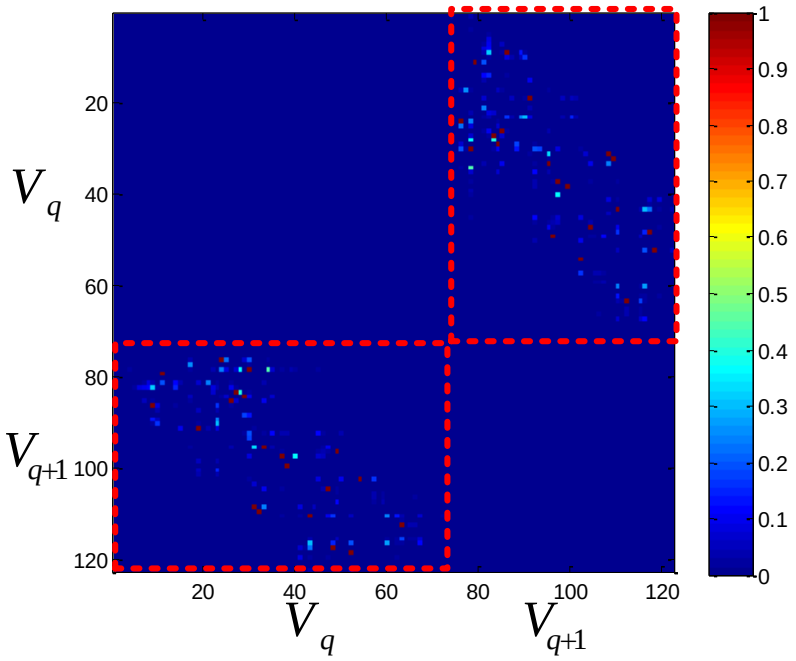


Figura 5.5: Mapa de pesos de um grafo RE representado como um mapa de magnitudes. Esse mapa tem em maior parte das componentes com o mesmo valor do mapa para o caso AJ (Figura 5.4). A diferença está nos pares de correspondência $[i_e, j_e]$, com $i_e \in V_q$ e $i_e \in V_q + 1$, que recebem um valor de ponderação muito alto, representando uma ligação infinita que pode ser observada no mapa pelos pontos em vermelho escuro.

5.2.4 Correspondências substituídas por elementos equivalentes

Em vez de se determinar pesos infinitos aos vértices referentes a regiões consideradas correspondentes entre quadros em sequência, como no modo RE, pode-se emergir os vértices equivalentes $[i_{e1}, i_{e2}, i_{e3}, ..., i_{en}] \in V_e$ ao longo de vários n quadros em um único vértice I_e tal que $[i_{e1}, i_{e2}, i_{e3}, ..., i_{en}] \equiv I_e$. Na Figura 5.6, que estende a Figura 5.3(a), temos em (a) e em (b) duas vistas do que retrata pares de elementos equivalentes emergidos, V_e, e aqueles vértices sem equivalência nos quadros V_q e V_{q+1}. Esse grafo equivalente pode ser visto como a combinação dos dois quadros (Figura 5.3(c)).

O mapa de pesos $\mathbf{W}^{EQ}$ tem componentes $w_{I,J}^{EQ}$ determinadas com base nas componentes do mapa de pesos ajustado:

$$w_{I,J}^{EQ} = \sum_{i \in I} \sum_{j \in J} w_{i,j}^{AJ}, \tag{5.5}$$

cada nova componente da matriz de pesos $\mathbf{W}^{EQ}$ será uma soma das relações de vizinhança dos seus elementos com referência ao mapa $\mathbf{W}^{AJ}$. A equação (5.5), para formação de um mapa de pesos equivalentes [45], tem como objetivo formar uma representação do grafo original que permita a realização de um corte/segmentação preservando as características originais mesmo com uma redução no número de elementos em análise.

Para um corte em grafo que envolva matrizes de tamanho $N \times N$ a redução no número de elementos é a alternativa mais direta para redução no curto computacional, sobressaltando o caso de segmentação de vídeos em que esse número N pode crescer linearmente com a adição de quadros. Constata-se uma redução no número de elementos para as matrizes para o caso EQ (Figura 5.6) em relação ao mapa de pesos de origem dessa redução AJ (Figura 5.4).

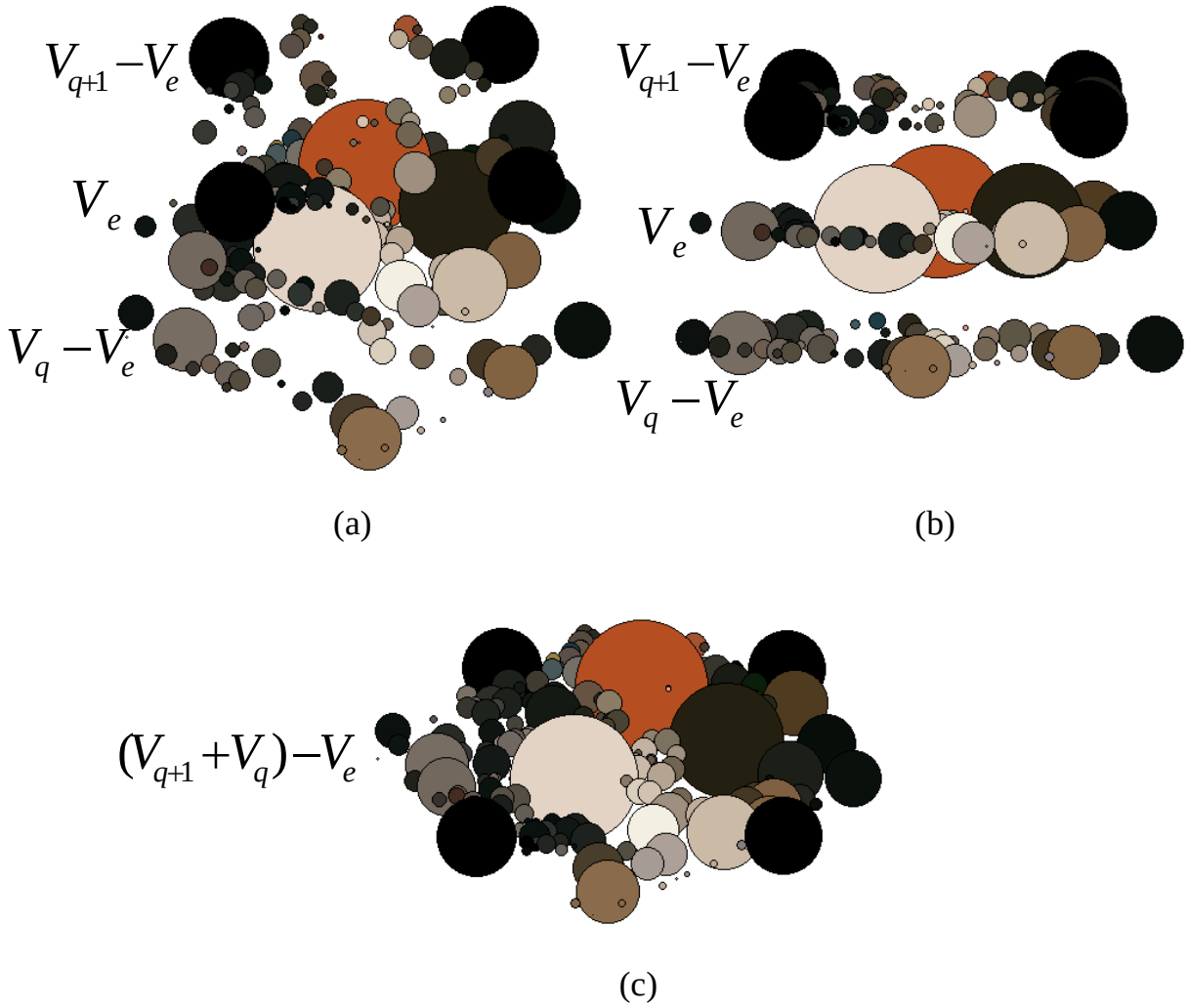


Figura 5.6: Ilustração para a equivalência de regiões referentes à Figura 5.3: (a) os quadros V_q e V_{q+1}, regiões/elementos na base e no topo, respectivamente, são exibidos sem os elementos equivalentes, representando pelo conjunto V_e ao centro; (b) outra vista para o apresentado em (a); (c) a correção de posição e substituição de elementos correspondentes por equivalências, permitem combinar dois quadros com características muito próximas de forma a serem tratados quase que como uma única imagem.

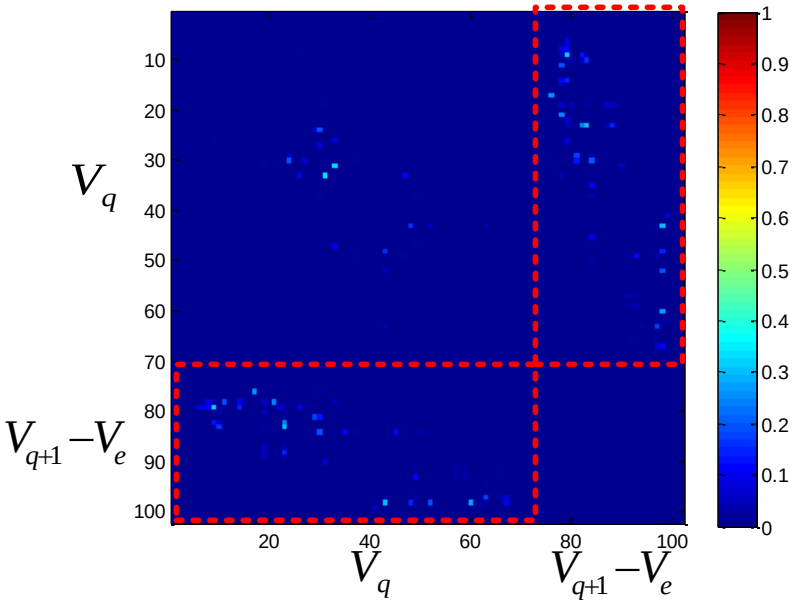


Figura 5.7: Mapa de pesos para o grafo EQ representado como um mapa de magnitudes. Esse mapa tem como base o mapa AJ, sendo que os elementos correspondentes são substituídos nós equivalentes. Isso elimina possíveis redundâncias ($V_{q+1} - V_e$) entre quadros, reduzindo o número de elementos submetidos à análise.

5.3 DETERMINAÇÃO DOS CORTES NOS GRAFOS

Oito formas de segmentação foram testadas neste trabalho, sendo separadas pelas 4 maneiras que seus elementos/regiões são agrupados e os 2 modos como o corte no grafo é aplicado. O corte de grafo pode ser aplicado quadro a quadro (Figura 5.8(a)) ou na totalidade de quadros (Figura 5.8(b)). O caso quadro a quadro envolve apenas um quadro e o subsequente, em uma segmentação feita em passos .

Apesar de envolver dois quadros, no modo de segmentação quadro a quadro apenas um quadro é segmentado na prática a cada passo, enquanto o outro fornece sementes (fundo em azul e objeto em vermelho, Figura 5.8). Para o primeiro quadro V_1 as sementes são determinadas por um GT, para os passos seguintes, as sementes são obtidas a partir das segmentações. Por exemplo, no passo 2 na (Figura 5.8(a)), o quadro segmentado V_2 segmentado no primeiro passo (representado por círculos coloridos), fornece sementes para a segmentação do terceiro quadro V_3, o processo se repete até todos os elementos da sequência passarem pelo corte.

A segunda forma de se segmentar um vídeo é aplicando um corte de grafos em uma sequência por completo (Figura 5.8(b)). Após o processamento quadro a quadro destinado ao casamento de regiões e do cálculo do vetor de movimento, aplica-se o corte no grafo contendo todos os elementos da sequência. O primeiro modo de corte, quadro a quadro, tem um esforço computacional menor do que o segundo modelo devido à quantidade de elementos agrupada.

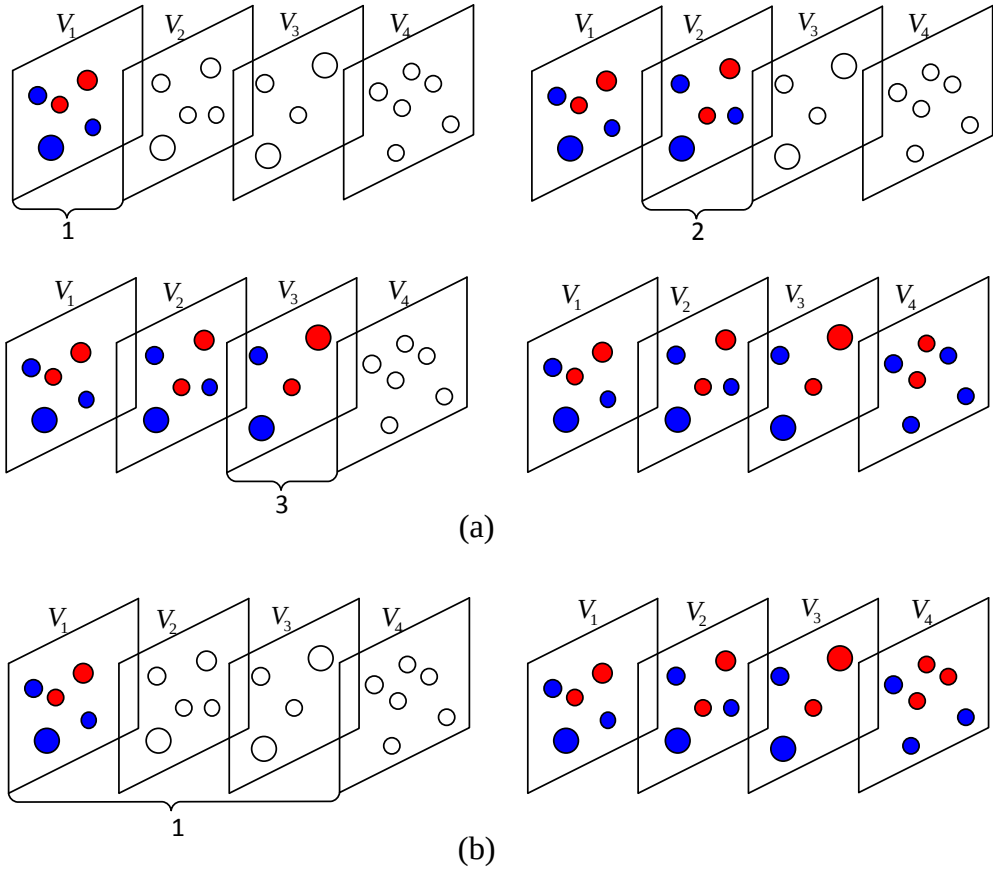


Figura 5.8: Ilustração para os dois padrões de segmentação adotados: (a) a segmentação quadro a quadro é procedida em um par de quadros, o quadro de referência segmentado oferece sementes para a segmentação de um quadro subsequente (elementos coloridos em vermelho para o objeto e em azul para o fundo). O primeiro quadro é segmentado a partir de um GT, os demais pelos métodos de corte de grafos proposto, a cada passo (esquerda para direita, de cima para baixo) um novo quadro é segmentado; (b) a segmentação em uma sequência de quadros possui um quadro como referência, o primeiro quadro segmentado com base no GT, que oferece sementes para a segmentação de toda sequência em um único passo.

Para se diferenciar os oito métodos em nomenclatura, adotou-se NAQ para a segmentação quadro a quadro sem ajuste de movimento, AJQ para a segmentação quadro a quadro com ajuste de movimento, REQ para a segmentação quadro a quadro com pesos reforçados e EQQ para a segmentação quadro a quadro com elementos equivalentes. O mesmo princípio de nomenclaturas pode ser definido para segmentações na totalidade de quadros de uma sequência testada, no mesmo padrão definido anteriormente temos: NAT, AJT, RET e EQT.

O princípio para a criação da matriz de pesos é semelhante para segmentação quadro a quadro e para aquela que contempla todos os quadros da sequência. No caso quadro a quadro uma imagem é segmentada a cada passo, mas duas imagens fornecem elementos para o grafo, um já segmentado fornecendo sementes e outro a ser segmentado. Quando se segmenta uma sequência por completo, utilizando-se elementos/regiões de todos os quadros, que se transformam em vértices do grafo. A construção do mapa de pesos reflete a força de ligação desses vértices.

5.3.1 Corte de grafo via *GrowCut*

O método utilizado para segmentar os grafos a partir de seus respectivos mapas de pesos criados neste Capítulo, foi o *GrowCut*, proposto na ref. [44] e apresentado no Capítulo 3. Os procedimentos adotados neste trabalho são semelhantes aos descritos na ref. [44], sofrendo algumas modificações para a aplicação no caso proposto.

Ao contrário do método original, quando aplica-se o *GrowCut* aos grafos utilizados neste trabalho, não se adota uma vizinhança de análise. Os próprios mapas de peso utilizados se valem de alguma informação quanto à abrangência da vizinhança de análise, como aquelas regiões que se te fronteiras afastadas, não sendo consideradas como desconectadas (equação (5.3)).

A cada iteração, o elemento i é atacado por seus $N-1$ vizinhos, sendo N o número de vértices do grafo a ser segmentado. O outro ponto de mudança em relação ao algoritmo original é a função de custos adotada. Para a ref. [44] tem-se a seguinte função:

$$g(i,j) = 1 - \frac{||\vec{\mathbf{I}_i} - \vec{\mathbf{I}_j}||^2}{max||\vec{\mathbf{I}}||^2}, \tag{5.6}$$

em que $\vec{\mathbf{I}_i}$ é a componente de cor RGB do elemento/pixel i e $\vec{\mathbf{I}_j}$ a componente RGB do elemento j. Subtraindo-se 1 pela distância euclidiana ao quadrado normalizada pela distância máxima ao quadrado entre as componentes de cor dos elementos, cria-se uma função monótona decrescente. A normalização limita o valor dessa função ao intervalo $[0, 1]$.

Na equação (5.6), sua característica decrescente aumenta o custo de se excluir dois elementos com cores próximas. No método proposto, essa função de custos deve envolver a magnitude dos pesos de ligação entre os elementos dos grafos, que é por si só fruto de uma função monotônicamente crescente e com valores dentro do intervalo $[0, 1]$. A função de custos para os elementos i e j nos grafos estudados é dada por:

$$g(i,j) = w(i,j), \tag{5.7}$$

em que $w(i,j)$ são as componentes do mapa de ponderação $\mathbf{W}$.

5.4 MÉTRICAS PARA ACURÁCIA E ERRO DE SOBRE-SEGMENTAÇÃO

Os métodos utilizados para medir acurácia da segmentação e erros de sobrestimação nos resultados oferecidos pelo algoritmo proposto, são baseadas na ref. [2], que utiliza o volume de supervoxels segmentados em relação a um *ground truth* para estimar o volume corretamente segmentado.

No caso estudado, a partir de uma segmentação SG (Figura 5.9(b)) a acurácia para um quadro é dada pelo número de pixels corretamente segmentados (Figura 5.9(c)) em relação a um *ground truth* (GT) (Figura 5.9(a)), dividido pelo número de elementos (pixels) deste GT. A sobrestimação (SE) segue o mesmo princípio, entretanto mede-se a quantidade de pixels segmentados que não pertencem à segmentação manual (Figura 5.9(d)).

Para a acurácia associada a um único quadro, divide-se o número de elementos contidos na intersecção entre a segmentação proposta SG e GT, pelo número de elementos de GT:

$$AC = \frac{\mathrm{n}(SG \cap GT)}{\mathrm{n}(GT)}, \tag{5.8}$$

em que $\mathrm{n}(\cdot)$ é a função que conta o número de pixels não nulos dentro da imagem binária. A proporção de pixels que ultrapassa a segmentação *ground truth* em um quadro é calculada como:

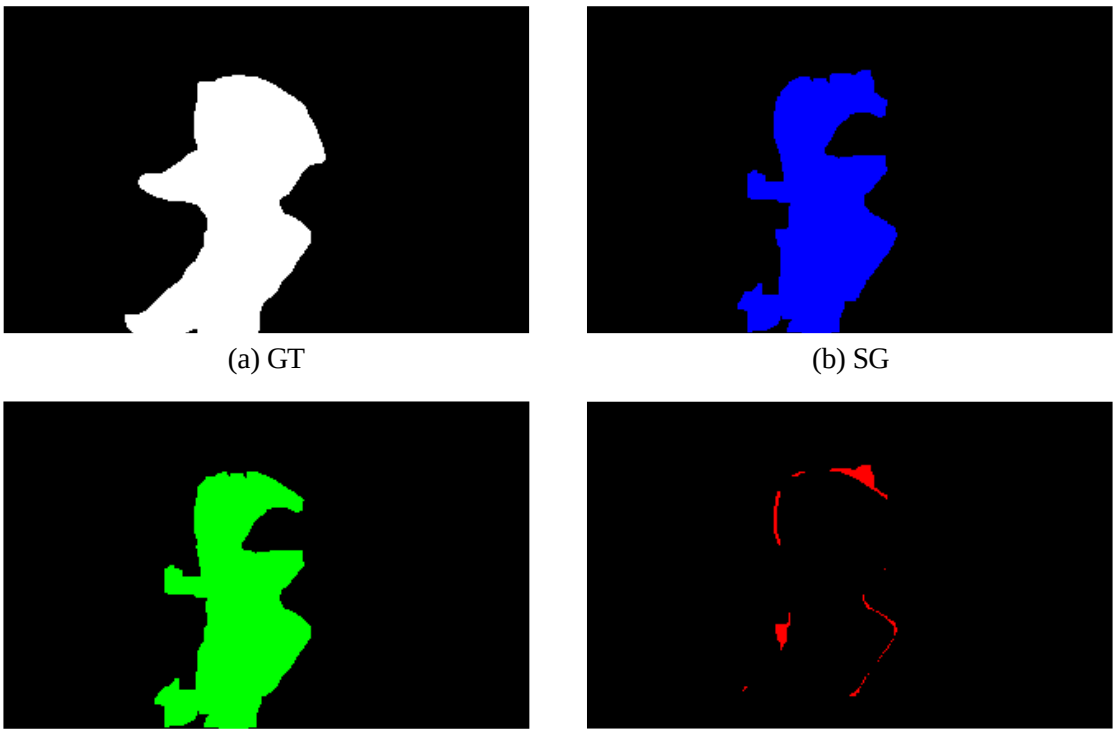
$$SE = \frac{\mathrm{n}(SG - GT))}{\mathrm{n}(GT)}. \tag{5.9}$$

Define-se também uma acurácia no volume 3D no espaço $x \times y \times$ tempo, para todos os elementos corretamente segmentados ao longo de um trecho de n_f quadros, como:

$$AC_{3D} = \frac{\sum_k^{nf} \mathrm{n}(SG_k \cap GT_k)}{\sum_k^{n_f} \mathrm{n}(GT_k)}. \tag{5.10}$$

a porcentagem de pixels que ultrapassa a segmentação *ground truth* de referência ao longo do trecho é calculada como:

$$SE_{3D} = \frac{\sum_k^{n_f} \mathrm{n}(SG_k - GT_k)}{\sum_k^{n_f} \mathrm{n}(GT_k)}. \tag{5.11}$$



(a) GT (b) SG

(c) SG ∩ GT (d) SG − GT

Figura 5.9: Ilustração representando a acurácia (AC) e sobrestimação (SE) do quadro 4 da sequência *Panda* exibida na Figura 4.24: (a) a segmentação manual do urso, ou chamado *ground truth* (GT) do objeto; (b) ao ser interseccionado pela segmentação (SG) proposta, no caso as regiões oriundas de sementes do primeiro quadro; (c) geram um conjunto de elementos/pixels, $SG \cap GT$, corretamente segmentados, que quando contabilizados e divididos pela soma de elementos do GT fornecem a AC; (d) quando se subtrai os elementos do conjunto GT do conjunto SG, obtém-se a quantidade de pixels que ultrapassam a segmentação de ideal (GT), ao se dividir a soma desses elementos que ultrapassam uma segmentação de referência, pela soma dos elementos dessa segmentação de referência, obtém-se o erro de sobrestimação SE.

6 RESULTADOS

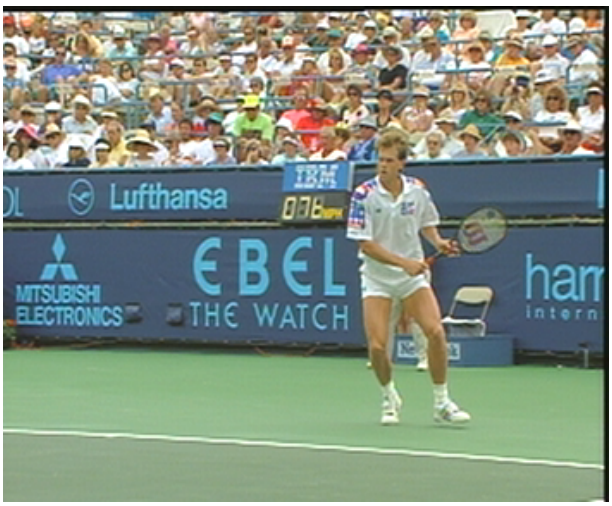
6.1 INTRODUÇÃO

Neste capítulo serão apresentados e discutidos os resultados obtidos, quanto ao rastreamento e segmentação de vídeos a partir dos métodos propostos. Primeiramente, é discutida a contribuição do algoritmo proposto para o casamento de regiões e para o rastreamento de objeto e fundo ao longo dos quadros. Em um segundo passo, discute-se a influência desse casamento de regiões na redução da quantidade de elementos para os grafos. Por fim, avalia-se as segmentações quadro a quadro ou ao longo da sequência, por meio do corte de grafos *GrowCut* para os tipos de mapas propostos NA, AJ, RE e EQ. Esses mapas foram pensados de forma a se avaliar a contribuição do descritor proposto no ajuste na posição das regiões (AJ), no reforço de pesos entre regiões correspondentes (RE) e na definição de equivalências para essas regiões correspondentes (EQ), todos em relação a forma mais simples e difundida de análise, comparando-se as posições e cores das região (NA).

6.2 SEQUÊNCIAS TESTADAS

As sequências de 9 quadros testadas foram: *Stefan*, *Angelfish*, *Trainer*, *Mobile* e *Panda*, apresentadas visualmente na Figura 6.1. A resolução natural das sequências *Stefan* e *Mobile* é 352×288, que podem ser obtidas, bem como seus *ground truth* para os objetos de interesse (Figura 6.2), na ref. [46]. O restante das sequências e suas segmentações manuais são encontradas na ref. [47], sendo que a resolução natural dos quadros dessas sequências é 320×200. Os quadros iniciais f_i de cada sequência foram definidos de forma que, ao longo dos 8 quadros subsequentes, houvesse um movimento constante do objeto em análise e poucas deformações.

Para cada sequência, o limiar T foi aplicado antes da segmentação *watershed* para a formação das regiões (Capítulo 4), escolhido de forma a criar regiões estáveis ao longo dos quadros. A Figura 6.3 apresenta gráficos para os deslocamentos normalizados dos centroides dos objetos de interesse nas sequências testadas. Esses deslocamentos são obtidos com base no GT da sequência e nas dimensões das regiões que o compõe, esses gráficos ajudam a avaliar os resultados.



(a) *Stefan*; $n = 1$; $T = 8$; $f_i = 1$

(b) *Angelfish*; $n = 2$; $T = 12$; $f_i = 22$

(c) *Trainer*; $n = 3$; $T = 10$; $f_i = 1$

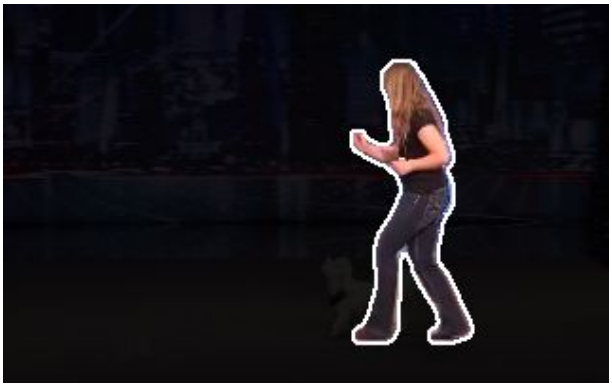
(d) *Mobile*; $n = 4$; $T = 10$; $f_i = 10$

(e) *Panda*; $n = 5$; $T = 5$; $f_i = 12$

Figura 6.1: Quadros iniciais das sequências de 9 quadros utilizadas para teste. f_i define a posição desse quadros iniciais na sequências originais encontradas em [46] e [47] . Abaixo das figuras define-se os parâmetros utilizados para a criação de regiões, a oitava n e o limiar T. As sequências são: (a) *Stefan*; (b) *Angelfish*; (c) *Trainer*; (d) *Mobile* e; (e) *Panda*.

(a) *Stefan*

(b) *Angelfish*



(c) *Trainer*

(d) *Mobile*

(e) *Panda*

Figura 6.2: Segmentação *ground truth* aplicada ao objeto de interesse para o primeiro quadro de cada uma das sequências testadas, esse GT do primeiro quadro fornece as sementes que são base para os processos de segmentação implementados: (a) na sequência *Stefan*, objeto de interesse é o tenista; (b) na *Angelfish* é o peixe de cores azul e amarela; (c) na *Trainer* é o indivíduo que orienta cachorro; (d) na *Mobile* envolve o conjunto calendário, bola e trem de brinquedo; (e) na *Panda*

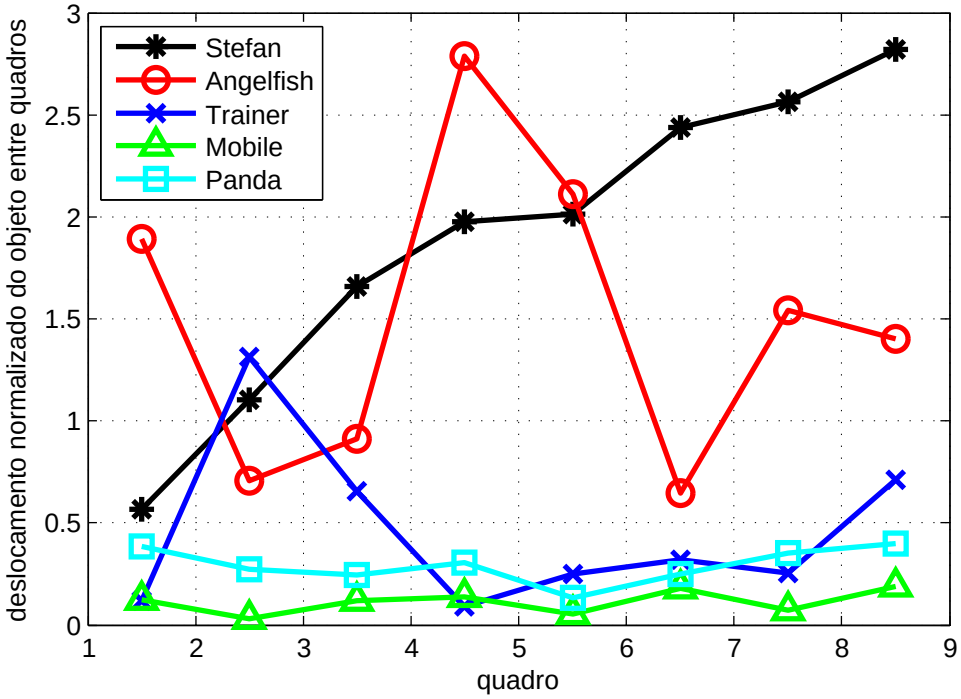


Figura 6.3: Gráfico dos deslocamentos normalizados entre quadros do centroide dos GT dos objetos para todas as sequências.

6.2.1 Espaço de escalas

Utilizou-se cinco escalas distintas nas cinco sequências estudadas, escalas as quais determinam a segmentação de regiões e, por consequência, os descritores dessas regiões (Capítulo 4). As cinco sequências, *Stefan*, *Angelfish*, *Trainer*, *Mobile* e *Panda*, estão distribuídas em escalas distintas de forma a cobrirem 5 oitavas no processo, indicadas pelo índice n na Figura 6.1. O desvio padrão que define a posição da escala na oitava (equação 4.33), foi fixado o mesmo para todas as sequências, $\sigma = 2$.

O intuito de representar as sequências em escalas ou oitavas distintas é verificar o nível de representatividade que uma escala pode fornecer para os objetos, oferecendo resultados satisfatórios e ao mesmo tempo reduzindo o esforço computacional relacionado a processamento em um nível de escala baixo. Também tem-se interesse em analisar a contribuição de diferentes níveis de escala para futuros trabalhos envolvendo segmentação hierárquica.

6.3 RASTREAMENTO DE REGIÕES E DE OBJETOS

Não foram determinadas métricas específicas para se estimar espacialmente a precisão do rastreamento fornecida pelo algoritmo proposto, ou seja, uma forma de se avaliar os deslocamentos das regiões e sua coerência espacial. Avaliou-se neste trabalho a contribuição do casamento/rastreamento de regiões para a segmentação do objeto ao longo de uma sequência, isto é, erros de casamento de regiões dentro do objeto, ou para o fundo, não foram levadas em consideração. Por exemplo, para o tenista na sequência *Stefan*, o erro quanto ao casamento de uma perna esquerda em um quadro com uma perna direta de outro quadro é ignorado, pois ambos elementos pertencem ao objeto de desejo.

A análise do rastreamento fornecido pelo algoritmo proposto focou na contribuição que esse tem na segmentação das sequências. Verifica-se a acurácia e a sobrestimação para uma segmentação levando-se em consideração apenas o casamento de regiões, sem o corte de grafo. Essa análise é feita de duas maneiras: (1) avalia-se o nível de conexão entre dois quadros subsequentes ao se utilizar o GT para definir as regiões corretamente casadas no outro, para objeto (OBQ) e fundo (BKQ); (2) rastreando as sementes fornecidas pelo primeiro quadro para toda a sequência, avalia-se o nível de mudanças do objeto dentro da sequência a partir do primeiro quadro, para objeto (OBT) e fundo (BKQ).

Pode-se exemplificar as duas formas de análise do rastreamento com a sequência *Angelfish* na Figura 6.4. O padrão de cores que se repete de um quadro para o outro, representa o casamento de regiões ao longo da sequência, sendo que uma nova cor é atribuída a cada nova área sem correspondência. Regiões escuras representam elementos sem correspondência dentro do par de confronto. A cada par de confrontos, 1-2, 2-3, 3-4 e assim por diante, a manutenção de um mesmo padrão de cor para uma região dentro do objeto ou fora dele, no caso o peixe, determina se há um erro ou não de rastreamento quadro a quadro. As regiões coloridas em quadros subsequentes (inferiores) que convergem com regiões dentro do GT (contorno esbranquiçado) na imagem de referência (superiores) representam o acerto de rastreamento quadro a quadro.

Os círculos vermelhos na Figura 6.4 representam as sementes do objeto determinadas pelo o GT do primeiro quadro, que se propagam ao longo da sequência pelo confronto de regiões. Os marcadores em ‘x’na cor azul representam a mesma propagação de sementes para o fundo, com base no GT da primeira imagem da sequência. Essas sementes propagadas a partir do primeiro quadro, são base para segmentações ao longo de toda a sequência, aproveitando as correspondências nos mapas de peso, RET e EQT.

Os gráficos exibidos na Figura 6.5 registram as taxas de acerto AC (equação (5.8)) e erro de sobrestimação SE (barras verticais) por quadro para as sequências estudadas. As curvas OBQ re-

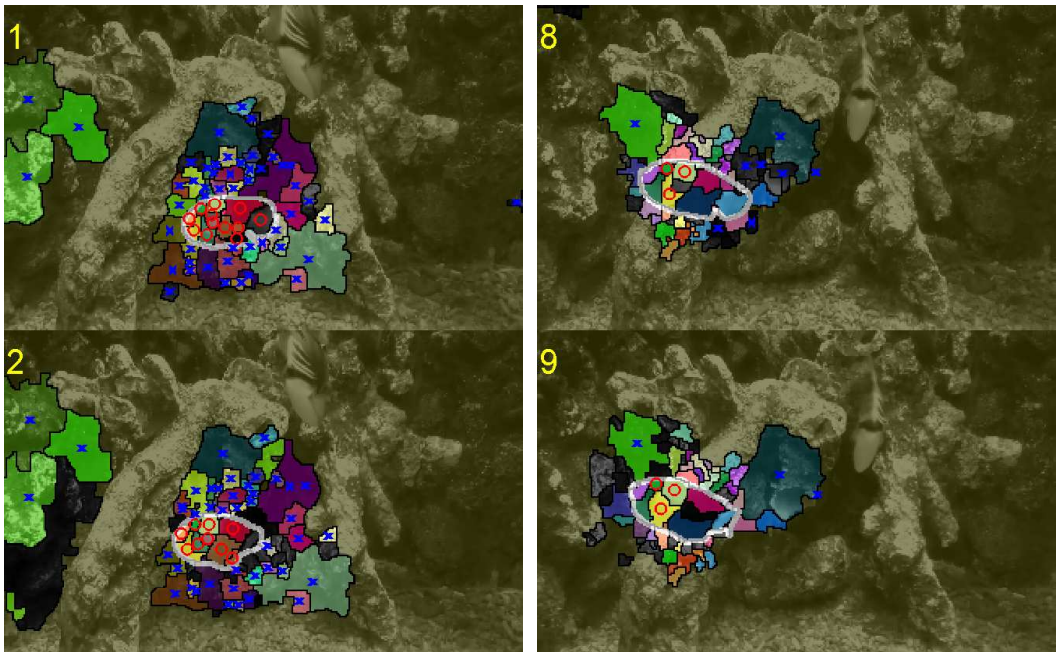
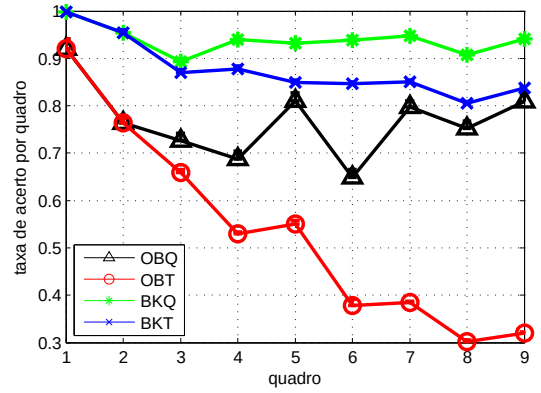


Figura 6.4: Rastreamento de sementes e convergência de regiões para a sequência *Angelfish*. São exibidos pares de confronto, par 1-2 e o 8-9. A manutenção do padrão de cores dentro do par indicam um casamento de regiões, as regiões escuras são regiões sem correspondência dentro do par de confronto. As circunferências vermelhas e as marcações em 'x'azuis representam sementes, para objeto e fundo, respectivamente, que tem base no GT do primeiro quadro e são propagadas ao longo do sequência por meio do algoritmo de rastreamento proposto.

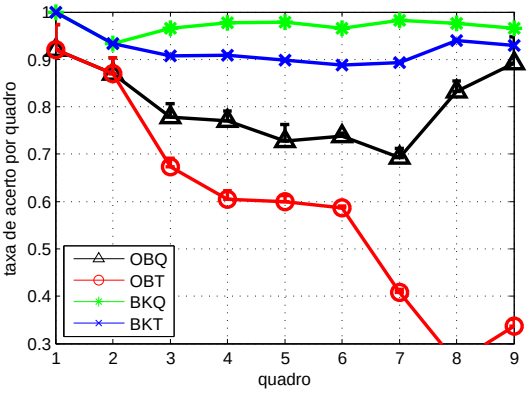
gistram um nível de semelhança do objeto para quadro o seu antecessor, da mesma maneira temos as curvas BKQ que medem a porcentagem do fundo pode ser encontrada no quadro antecessor por meio do confronto de regiões. As curvas OBT e BKT relacionam a área do objeto e fundo de um quadro, com as regiões que o objeto e do fundo no primeiro quadro. Essa relação é fruto do casamento de regiões do primeiro propagadas até o quadro de desejo.

Os resultados para o rastreamento aparentam indicar uma correlação do movimento entre quadros (Figura 6.3) com capacidade do algoritmo proposto em achar correspondências para o objeto ao longo de uma sequência (Figura 6.5). Tal comportamento era esperado, pois movimentos ou distorções nas regiões e nas suas vizinhanças provocam uma queda no desempenho do descritor. Além do movimento, a escala pode ser um fator que influencia nessa queda de desempenho, pois quanto menores os agrupamentos dentro da imagem, menor a sua representatividade e singularidade, da mesma forma que o algoritmo original se propõe a eliminar pontos irrelevantes, deixando apenas os chamados pontos-chave.

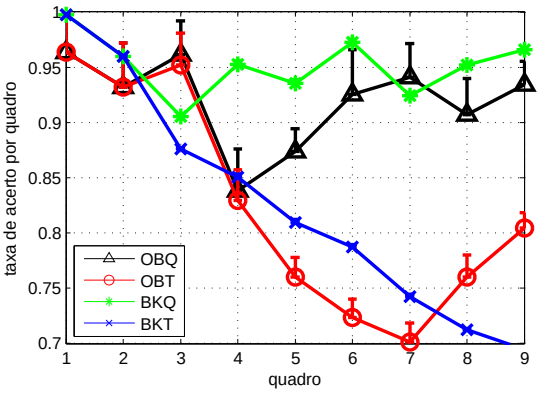
Ao se acumular uma taxa de acerto (equação (5.10)) ou de sobrestimação (equação (5.11)) do primeiro quadro ao quadro final da sequência, pode-se analisar o desempenho do algoritmo de rastreamento dentro do volume $x \times y \times$ tempo. A Tabela 6.1 exibe essa representação de erro no volume.



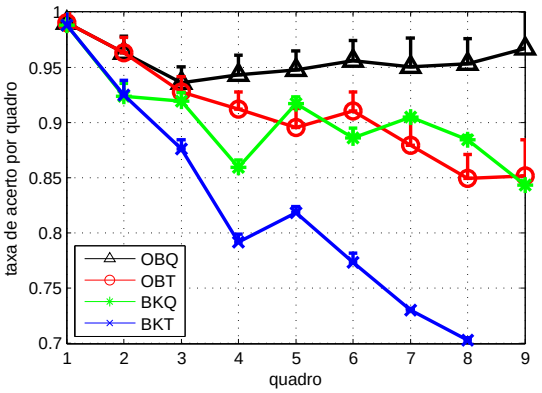
(a) *Stefan*



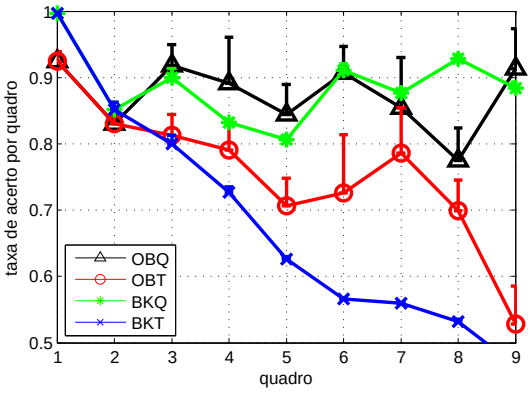
(b) *Angelfish*



(c) *Trainer*



(d) *Mobile*



(e) *Panda*

Figura 6.5: Gráficos para acurácia (AC) e erro de sobrestimação (SE, barras verticais) para o rastreamento de regiões ao longo dos quadros. OBQ e BKQ são curvas que representam a porção de regiões que são encontradas em um quadro originadas de um quadro antecessor pelo processo de casamento de regiões. OBT e BKT representam as porções de área referentes ao objeto e ao fundo no primeiro quadro que são encontradas nos quadros seguintes pelo processo de rastreamento proposto.

Tabela 6.1: Tabela para acurácia (AC_{3D}) e erro sobrestimação (SE_{3D}) no volume composto pelo objeto rastreado ao longo 9 dos quadros das sequências. OBQ retrata um nível de casamento entre quadros, a porção de área dos quadros que tem origem em um quadro antecessor. OBT relaciona a porção do volume formado por regiões do objeto na primeiro propagadas pelos demais quadros pelo rastreamento proposto.

Sequência	**Percentual**	OBQ	OBT
Stefan	AC_{3D}	76,79	54,15
	SE_{3D}	1,62	0,8
Angelfish	AC_{3D}	80,33	56,53
	SE_{3D}	3,07	1,73
Trainer	AC_{3D}	91,97	82,14
	SE_{3D}	3,22	2,4
Mobile	AC_{3D}	95,63	90,91
	SE_{3D}	1,96	1,85
Panda	AC_{3D}	87,37	75,97
	SE_{3D}	4,43	4,4

Para o rastreamento de regiões, as sequências *Stefan* e *Angelfish* têm os piores desempenhos para as duas situações, no casamento quadro a quadro e na propagação de sementes do objeto no primeiro quadro. O movimento de ambos em relação a dimensões das regiões que os compõe é mais relevante que para as outras sequências. Para *Stefan* (Figura 6.3) o deslocamento entre quadros, d_Q, se mantém acima de uma unidade, enquanto a sequência mais estável e de melhor desempenho, *Mobile* (d), não ultrapassa esse valor em nenhuma situação.

A hipótese de que a escala é um fator de influência no casamento e rastreamento de regiões é reforçada quando se compara o casamento de regiões nas sequências *Stefan* e *Trainer* (Figuras 6.6 e 6.7, respectivamente). Apesar de apresentarem o mesmo padrão de movimento para os indivíduos em cenas (Figuras 6.6 e 6.7, mapa de movimento), um movimento lateralizado da câmera que produz um deslocamento de objeto e fundo, o rastreamento para a sequência *Trainer* ($n = 3$, terceira oitava) apresenta melhor desempenho na análise quadro a quadro, acurácia de 91,97% contra 76,79%. Ao longo dos 9 quadros o volume das regiões rastreadas do objeto a partir do primeiro quadro (OBT) representa 82,14% do indivíduo em *Trainer*, enquanto para mesma situação tem-se 54,15% para a sequência *Stefan*.

Em termos de sobrestimação, a sequência *Trainer* tem um desempenho mais baixo que a *Stefan*, ultrapassando as fronteiras da segmentação GT de maneira mais acentuada que o rastreamento na sequência *Stefan*. Atribui-se esse comportamento à escala/oitava, quanto maior as regiões formadas no processo de agrupamentos, maior a área relacionada a um erro de rastreamento ou sobrestimação.

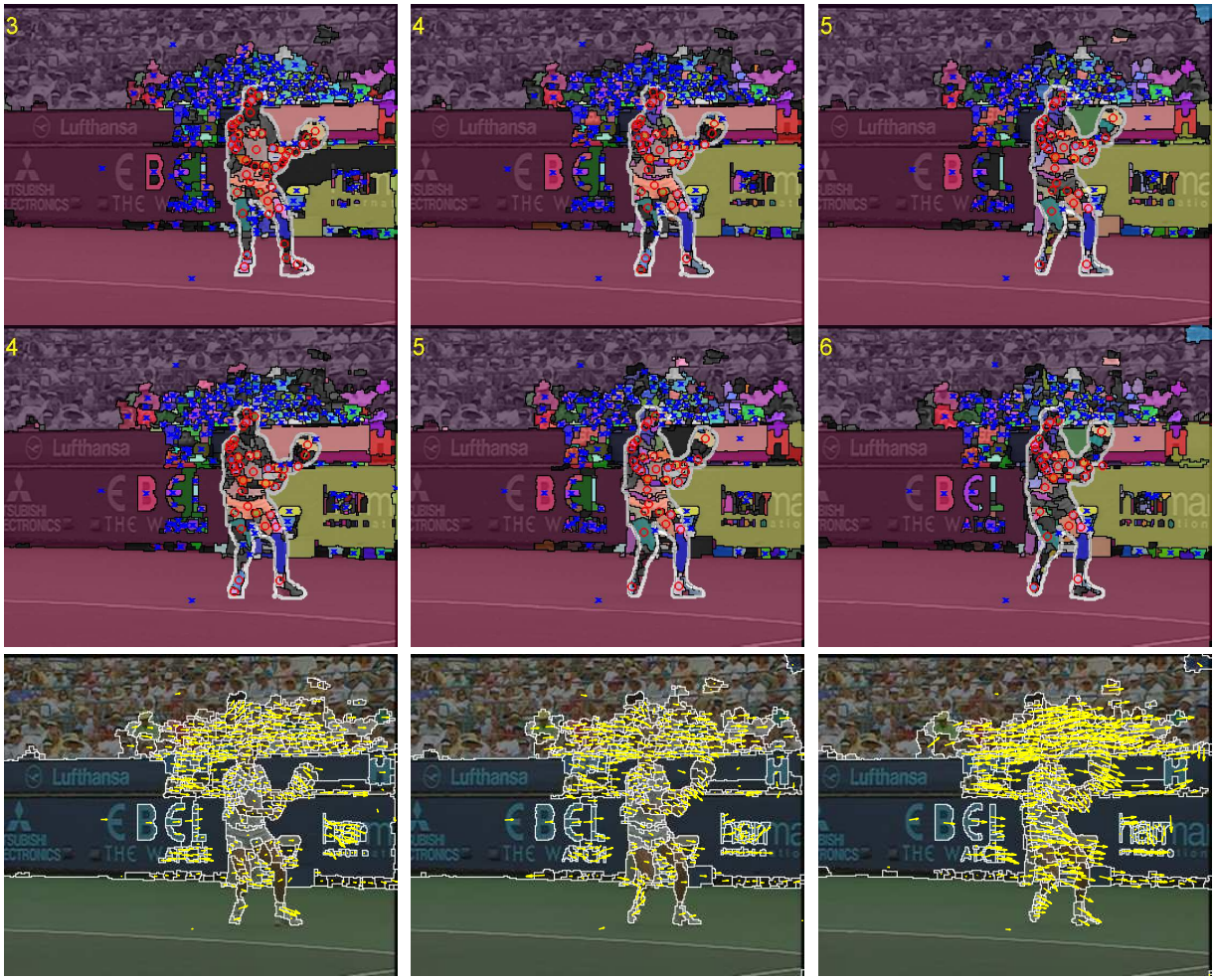


Figura 6.6: Pares de confronto e casamento de regiões dos quadros 3-4, 4-5 e 5-6 da sequência *Stefan* (pares de imagem superiores). Regiões correspondentes tem um mesmo rótulo e recebem uma mesma cor, regiões escuras representam regiões sem correspondência para o par de confronto. Circunferências vermelhas e os marcadores 'x' em azul são as sementes referentes ao objeto e ao fundo, respectivamente, propagadas ao longo dos quadros pelo processo de rastreamento proposto. Abaixo de cada par de confronto encontra-se o respectivo mapa de movimento, calculado em conjunto com o casamento de regiões.

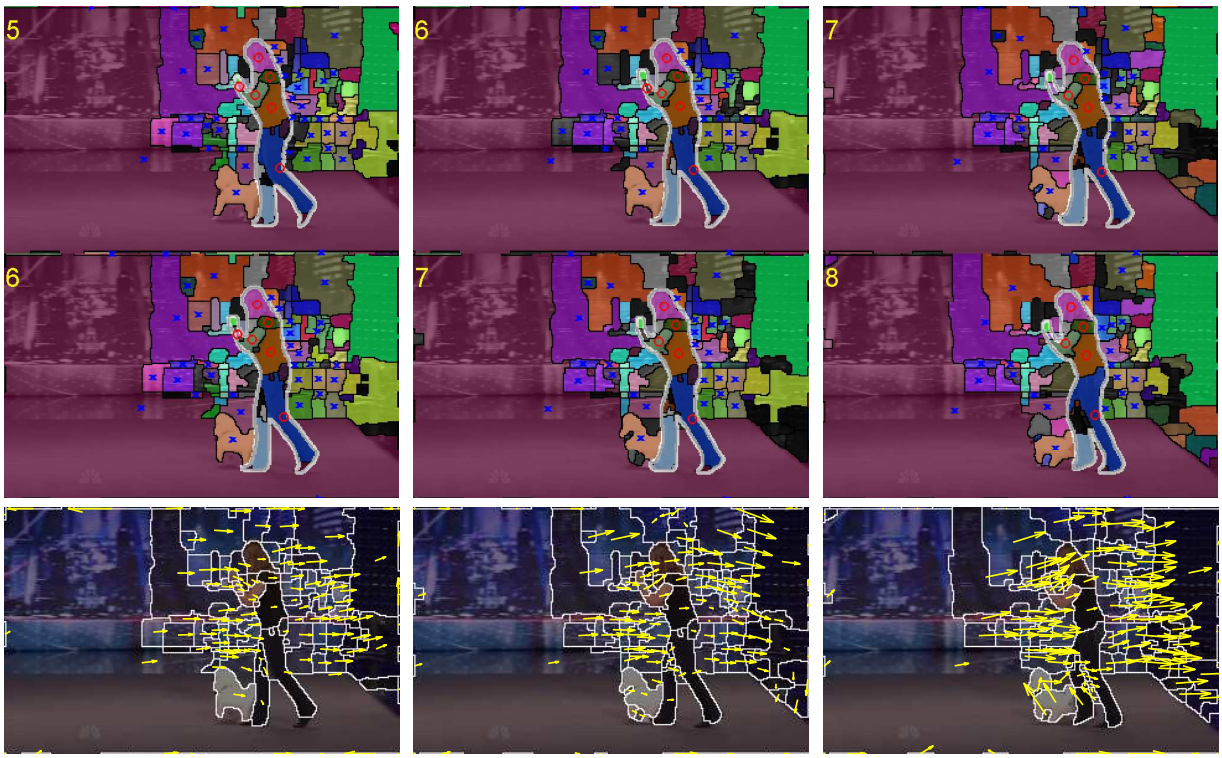


Figura 6.7: Pares de confronto e casamento de regiões dos quadros 5-6, 6-7 e 7-8 da sequência *Trainer* (pares de imagem superiores). Regiões correspondentes tem um mesmo rótulo e recebem uma mesma cor, regiões escuras representam regiões sem correspondência para o par de confronto. Circunferências vermelhas e os marcadores 'x' em azul são as sementes referentes ao objeto e ao fundo, respectivamente, propagadas ao longo dos quadros pelo processo de rastreamento proposto. Abaixo de cada par de confronto encontra-se o respectivo mapa de movimento, calculado em conjunto com o casamento de regiões.

A Tabela 6.2 apresenta valores de acurácia (AC_{3D}) e erro de sobrestimação (SE_{3D}) para a segmentação dos objetos ao longo do volume formado pelos 9 quadros nas 5 sequências testadas. Na Tabela, BKQ* e BKT* são os percentuais de acurácia e sobrestimação BKQ e BKT com base no volume do GT do objeto.

Tabela 6.2: Resultados para acurácia (AC_{3D}) e erro sobrestimação (SE_{3D}) no volume composto pelo fundo rastreado ao longo 9 dos quadros das sequências. BTQ retrata um nível de casamento entre quadros, a porção de área dos quadros que tem origem em um quadro antecessor. BKT relaciona a porção do volume formado por regiões do fundo na primeiro propagadas pelos demais quadros pelo rastreamento proposto. BTQ* e BKT* têm como referência o objeto, ou seja, a porção do fundo rastreada em relação ao tamanho do objeto de interesse.

Sequência	**Percentual**	BKQ	BKT	BKQ*	BKT*
Stefan	AC_{3D}	93,9	87,65	1.349	1.260
	SE_{3D}	0,61	0,52	8,76	7,46
Angelfish	AC_{3D}	97,16	92,21	4.049	3.843
	SE_{3D}	0,1	0,1	3,69	1,97
Trainer	AC_{3D}	95,17	82,58	1.504	1.305
	SE_{3D}	0,22	0,18	3,44	2,87
Mobile	AC_{3D}	90,27	80,37	76,34	67,96
	SE_{3D}	0,68	0,68	0,58	0,57
Panda	AC_{3D}	88,74	67,93	454,5	347,9
	SE_{3D}	0,9	0,73	4,61	3,74

A análise do casamento de regiões relacionadas ao fundo (BK) retorna aspectos importantes para a segmentação do objeto. O movimento do objeto dentro de uma cena, revela regiões oclusas que podem ser casadas com regiões que pertencem ao objeto, configurando um erro de sobrestimação. Os erros de sobrestimação para o casamento de elementos BK, limitam a acurácia para a segmentação do objeto.

Essa limitação pode ser observada com maior relevância para a sequência *Stefan*, uma vez que a Tabela 6.2 exibe para essa sequência um erro de sobrestimação de cerca de 8%, para os dois casos BKQ* e BKT*. Isto é, ao se segmentar uma imagem a partir do corte em grafos em um mapas de pesos do tipo RE ou EQ, que têm ponderações ou equivalências baseados no casamento de regiões, a acurácia não ultrapassará os 92%, pois 8% do objeto foi incorretamente relacionado a uma região pertencente ao fundo. Essa sobrestimação também pode estar relacionada a uma má definição dos contornos das regiões em relação ao GT.

Ao contrário do rastreamento de regiões para o objeto, a taxa de acerto para o casamento de regiões relacionadas ao fundo foi menor em agrupamentos oriundos de escalas maiores. As sequências *Mobile* ($n = 4$, quarta oitava, Figura 6.8) e *Panda* ($n = 5$, quinta oitava, Figura 6.9) têm as maiores quedas em taxa de acerto para o casamento de regiões BK.

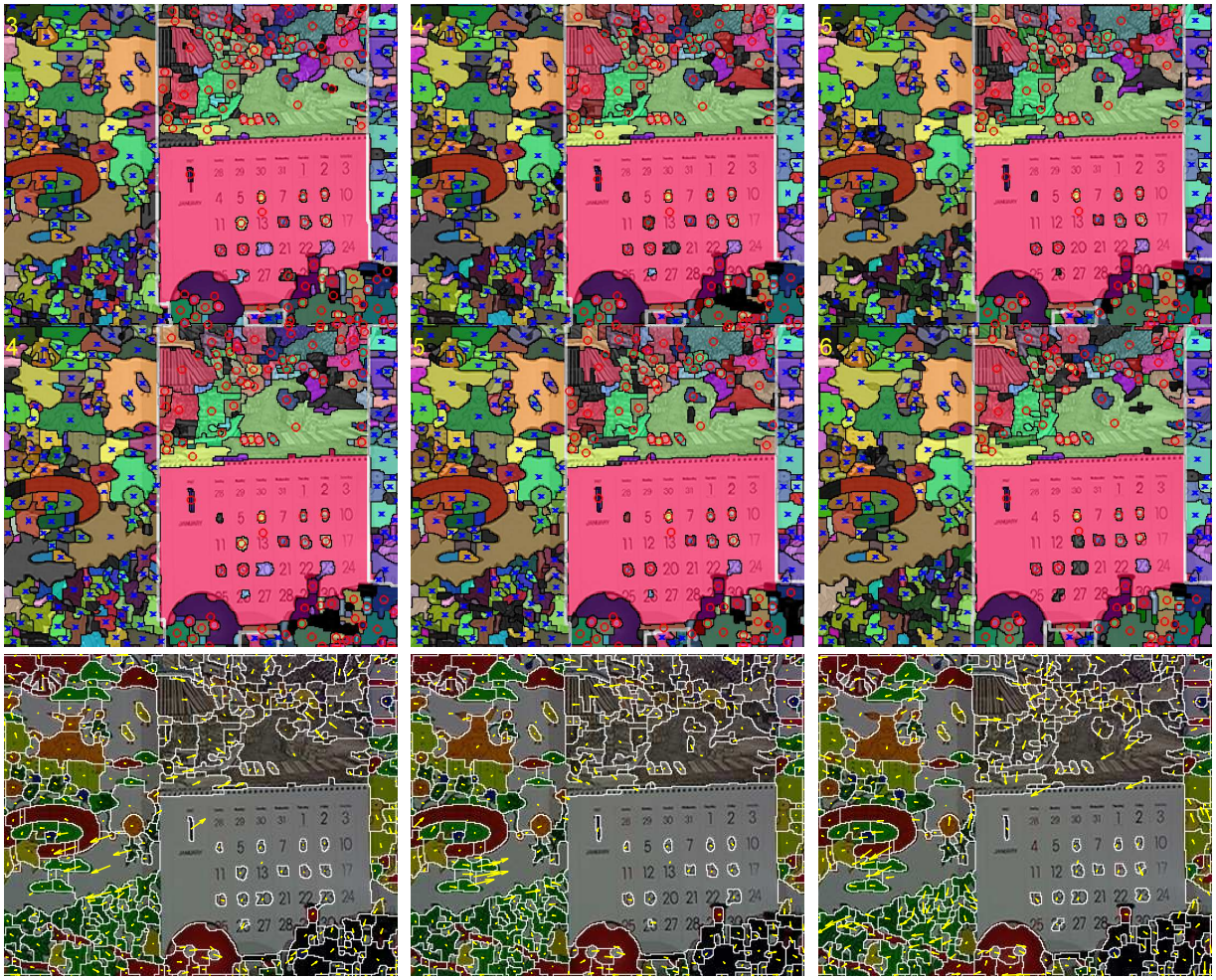


Figura 6.8: Pares de confronto e casamento de regiões dos quadros 3-4, 4-5 e 5-6 da sequência *Mobile* (pares de imagem superiores). Regiões correspondentes tem um mesmo rótulo e recebem uma mesma cor, regiões escuras representam regiões sem correspondência para o par de confronto. Circunferências vermelhas e os marcadores 'x' em azul são as sementes referentes ao objeto e ao fundo, respectivamente, propagadas ao longo dos quadros pelo processo de rastreamento proposto. Abaixo de cada par de confronto encontra-se o respectivo mapa de movimento, calculado em conjunto com o casamento de regiões.

Figura 6.9: Pares de confronto e casamento de regiões dos quadros 5-6, 6-7 e 7-8 da sequência *Panda* (pares de imagem superiores). Regiões correspondentes tem um mesmo rótulo e recebem uma mesma cor, regiões escuras representam regiões sem correspondência para o par de confronto. Circunferências vermelhas e os marcadores 'x' em azul são as sementes referentes ao objeto e ao fundo, respectivamente, propagadas ao longo dos quadros pelo processo de rastreamento proposto. Abaixo de cada par de confronto encontra-se o respectivo mapa de movimento, calculado em conjunto com o casamento de regiões.

6.3.1 Redução no número de elementos em grafos via equivalências

Uma contribuição importante para o casamento de regiões é a possibilidade da redução de elementos para a construção de um grafo, procedimento aplicado neste trabalho para a criação dos mapas tipo EQ. O número total de regiões/elementos N dentro dos 9 quadros para cada sequência estudada, pode ser distribuído em uma média por quadros $\overline{N}_1$ ou uma média por pares de quadros $\overline{N}$.

O número de elementos EE dentro de um grafo do tipo EQ retrata a quantidade de elementos com a qual é possível se representar um grafo primordial, no caso, os grafos do tipo AJ. Para uma análise em termos normalizados, a média de elementos por quadro $\overline{N}$ serve de base para medir-se um nível de compressão fornecida por um grafo equivalente, criado a partir do casamento de regiões. Compara-se a quantidade média de elementos a cada dois quadros, $\overline{N}_2$, com o número de elementos para sua versão comprimida, equivalente, $\overline{NE}_2$. Mesma comparação pode ser feita entre o total de regiões/elementos na sequência, N, com o número de elementos da versão equivalente NE. Esses dados, para as 5 sequências, estão expressos na Tabela 6.3.

Tabela 6.3: Relação de elementos em valores absolutos e normalizados. A normalização se dá pelo número médio de elementos por quadro $\overline{N}_1$, calculado com base no número total de elementos/regiões dentro da sequência N dividido uniformemente pelos seus 9 quadros. $\overline{N}_2$ é o número médio de elementos a cada dois quadros e $\overline{NE}_2$ o número de elementos equivalentes entre dois quadros, determinado pelos casamentos de regiões. O mesmo princípio de equivalência pode ser adotado para toda a sequência com o número de elementos NE.

Sequência	**Valor**	$\overline{N}_1$	$\overline{N}_2$	$\overline{NE}_2$	N	NE
Stefan	absoluto	434,6	869,1	572,9	3.911	1.577
	normalizado	1	2	1,32	9	3,63
Angelfish	absoluto	67,89	135,8	90,13	611	238
	normalizado	1	2	1,33	9	3,51
Trainer	absoluto	115	230	145,1	1.035	350
	normalizado	1	2	1,26	9	3,04
Mobile	absoluto	355,7	711,3	416,3	3.201	845
	normalizado	1	2	1,17	9	2,38
Panda	absoluto	128,7	257,3	157,9	1.158	352
	normalizado	1	2	1,23	9	2,74

O número médio de elementos por quadro $\overline{N}_1$ é uma boa referência, por se tratar da quantidade média de regiões envolvida no processo de cálculo de descritores. As sequência que envolvem o maior esforço computacional nesse processo são a *Stefan* e *Mobile*, com uma média de 434,6 e 355,7 elementos por quadro. No corte quadro a quadro, que envolve grafos que relacionam dois quadros, a sequência *Mobile* sofre a maior compressão de elementos. Ao invés de mapas de peso,

com dimensão média de 711×711, pode-se representar pares de quadros na sequência *Mobile* com mapas reduzidos para uma dimensão 416×416.

No caso de grafos construídos com elementos de toda a sequência, a compressão é significativa, ao invés de um crescimento no número de componentes de $9^2\times$, 81 vezes, na matriz de pesos, em relação a média por quadros, tem-se mapas de peso equivalentes com um crescimento de máximo $3{,}65^2\times$, cerca de 13 vezes, em relação a um mapa construído com a $\overline{N}_1$ elementos. Os descritores propostos são calculados utilizando matrizes com dimensão $\overline{N}_1 \times \overline{N}_1$.

6.4 SEGMENTAÇÃO DE OBJETOS

Os resultados da segmentação estão dispostos de acordo com a forma que o corte de grafos é aplicado, quadro a quadro ou na sequência por completo. As duas modalidades de corte são divididas em quatro formas de se ponderar as ligações do grafo, NAQ, AJQ, REQ e EQQ para os casos quadro a quadro e NAT, AJT, RET e EQT para a segmentação em um grupo de quadros.

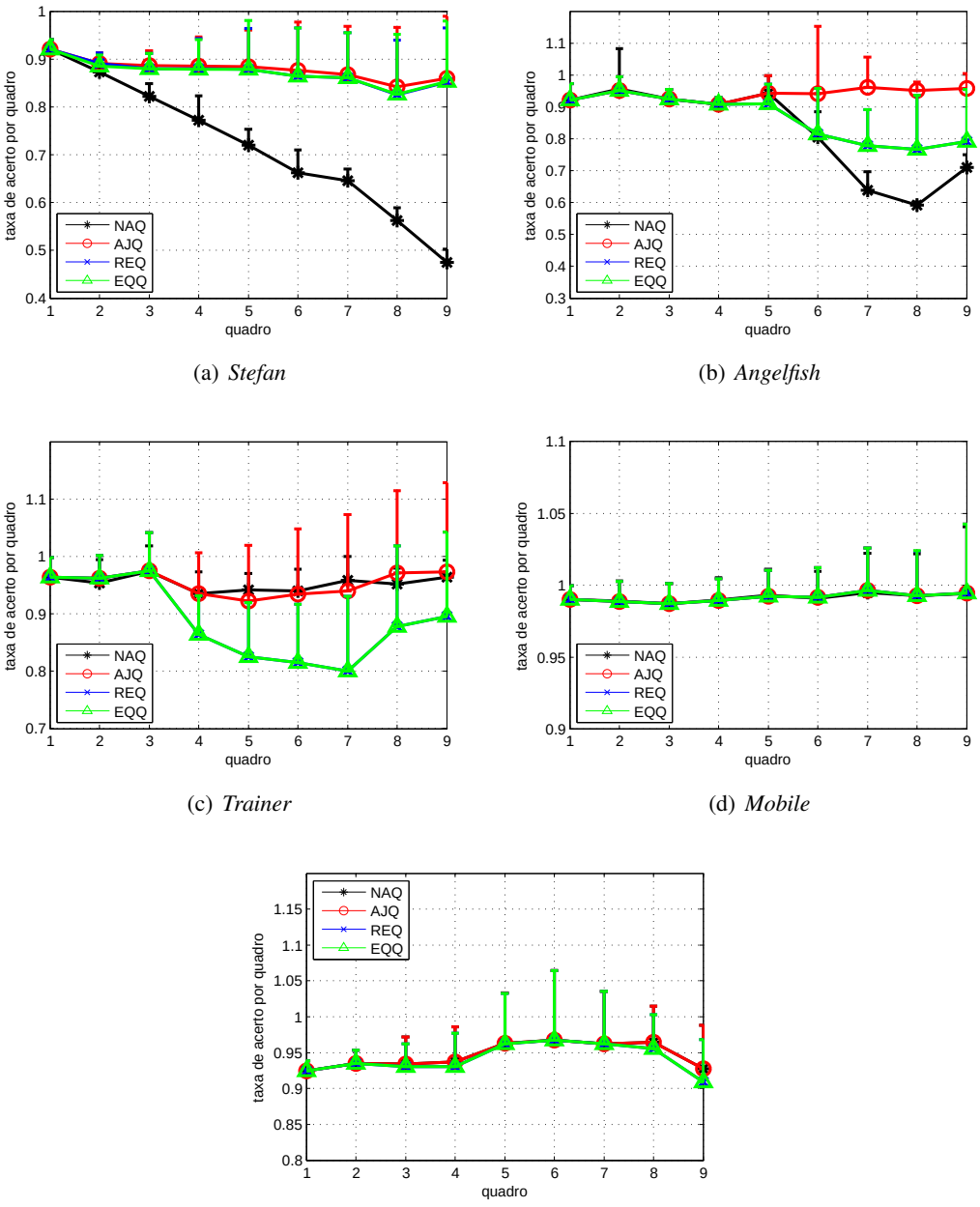
Os gráficos da Figura 6.10 exibem a taxa de acerto por quadro (AC) e erro de sobrestimação (SE) representado pelas barras verticais, para segmentações efetuadas em cortes quadro a quadro nas 5 sequências estudadas. Nota-se que, quando não aplicado um ajuste nas posições dos elementos de um quadro para o outro, caso NAQ, as duas sequências segmentadas a partir de agrupamentos em escalas menores, *Stefan* e *Angelfish*, apresentam um baixo desempenho em relação aos outros casos que utilizam um vetor de movimento para correção de deslocamentos.

Os gráficos da Figura 6.11 registram a taxa de acerto por quadro para o corte de grafo aplicado a todos 9 os quadros das sequências estudadas. Os modos de organização e ponderação das ligações dos grafos são separados em NAT, AJT, RET e EQT. Para as sequências *Stefan* e *Angelfish* há uma queda de desempenho em relação aos mesmo padrões de ponderação para um corte efetuado quadro a quadro.

Comparando o corte quadro a quadro (Figura 6.10) com o efetuado ao longo de toda a sequência (Figura 6.11), ao se observar o resultado para as sequências com regiões agrupadas em na quarta e na quinta oitava, *Mobile* (d) e *Panda* (e), respectivamente, nota-se uma certa manutenção no desempenho para ambos os casos, tendo a sequência *Panda* um aumento na sobrestimação (barras verticais) do caso quadro a quadro em relação aquele aplicado em toda a sequência.

Como esperado, os cortes que operam com o casamento de regiões, sejam por uma ligação reforçada entre essas (REQ ou RET) ou a emersão de um grupo de elementos em um único nó equivalente (EQQ ou EQT) têm comportamentos semelhantes para as sequências. As curvas

para esses casos se super posicionam em quase todas as taxas de erro por quadro (Figura 6.10 e 6.11), destacando-se a sequência *Trainer* na qual o corte no mapa no padrão EQT tem melhor desempenho do que o RET (Figura 6.11(c)) e o melhor desempenho entre todas a formas de mapa e corte.

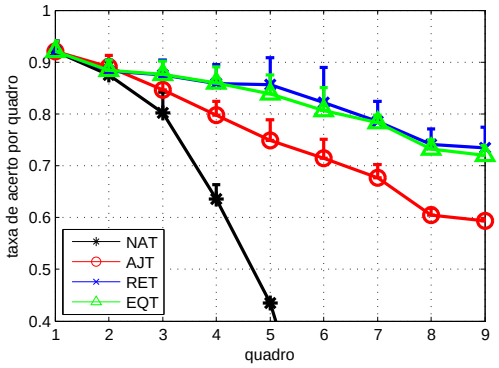


(a) *Stefan* (b) *Angelfish*

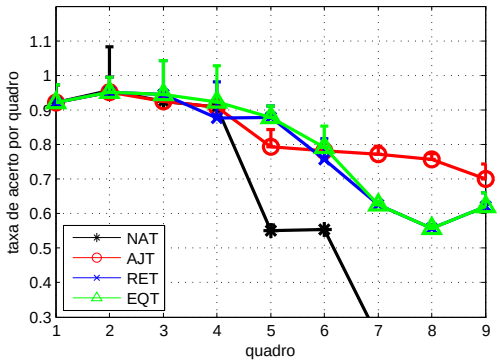
(c) *Trainer* (d) *Mobile*

(e) *Panda*

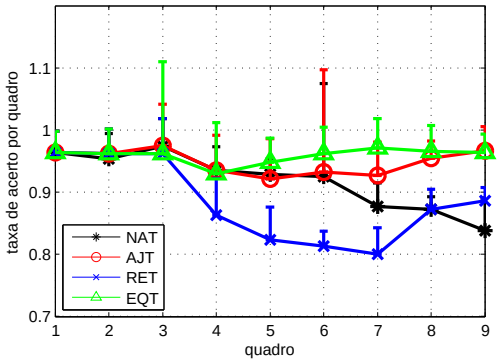
Figura 6.10: Taxa de acerto (AC) e erro sobrestimação (SE, barras verticais) por quadro para as segmentações aplicadas aos grafos no modo quadro a quadro nas sequências estudadas. NAQ representa acurácias para o corte em em um grafo no qual não há correção de movimento de um quadro para outro para atribuição e pesos de ligação. AJQ representa o grafo cujas relações de vizinhança recebem a correção do vetor de movimento proposto. REQ é tem o mesmo mapa que AJQ com reforços de ligação nos nós/regiões correspondentes. EQQ são as curvas para a segmentação em um grafo equivalente, onde emerge-se nós correspondentes em um mesmo elemento.



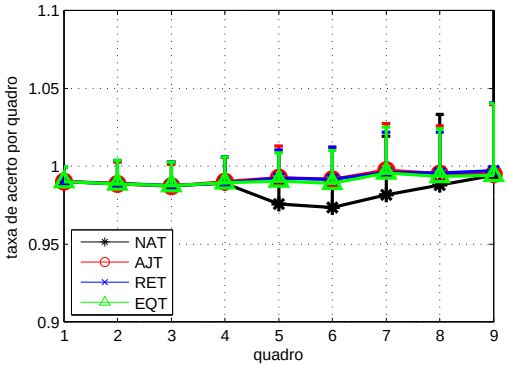
(a) *Stefan*



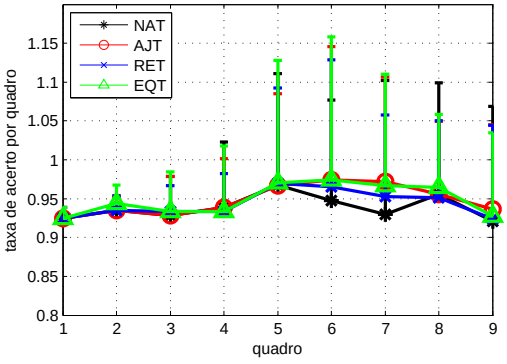
(b) *Angelfish*



(c) *Trainer*



(d) *Mobile*



(e) *Panda*

Figura 6.11: Taxa de acerto (AC) e erro sobrestimação (SE, barras verticais) por quadro para as segmentações aplicadas aos grafos compostos por elementos dos 9 quadros das sequências estudadas. NAT representa acurácias para o corte em em um grafo no qual não há correção de movimento de um quadro para outro para atribuição e pesos de ligação. AJT representa o grafo cujas relações de vizinhança recebem a correção do vetor de movimento proposto. RET é tem o mesmo mapa que AJT com reforços de ligação nos nós/regiões correspondentes. EQT são as curvas para a segmentação em um grafo equivalente, onde emerge-se nós correspondentes em um mesmo elemento.

A Tabela 6.4 registra os valores para AC_{3D} e SE_{3D}, taxas de acerto e de erro de sobrestimação, para o volume formado pelo conjunto de 9 quadros.

Tabela 6.4: Resultados para acurácia (AC_{3D}) e erro sobrestimação (SE_{3D}) no volume referente ao objeto segmentado após aplicação do corte de grafos. São exibidos resultados para 4 tipos de mapa de pesos no corte quadro a quadro, NAQ, AJQ, REQ e EQQ, e para os 4 tipos de mapa no corte de grafo aplicado em todo grupo de quadros, NAT, AJT, RET e EQT.

		Mapa de pesos							
Sequência	**Percentual**	NAQ	AJQ	REQ	EQQ	NAT	AJT	RET	EQT
Stefan	AC_{3D}	72,56	86,1	85,86	85,24	43,33	77,33	82,47	80,89
	SE_{3D}	4,15	7,39	7,92	7,74	1,49	2,99	5,01	3,94
Angelfish	AC_{3D}	80,87	94,07	85,61	85,61	54,03	82,72	77,78	78,66
	SE_{3D}	5,1	6,44	9,33	9,33	2,59	3,18	4,79	4,82
Trainer	AC_{3D}	95,33	95,23	88,39	88,39	91,73	94,78	88,07	95,9
	SE_{3D}	3,55	9,6	9,23	9,23	4,96	6,13	3,97	5,38
Mobile	AC_{3D}	99,14	99,15	99,15	99,15	98,54	99,21	99,21	99,1
	SE_{3D}	2,13	2,22	2,22	2,23	3,16	2,21	2,12	2,24
Panda	AC_{3D}	94,59	94,59	94,17	94,17	93,85	94,74	94,3	94,84
	SE_{3D}	5,14	5,14	5,01	5,01	9,82	8,42	7,87	9,38

Uma discussão aliada a uma inspeção visual nos resultados pode ajudar a entender melhor os dados da Tabela 6.4. No corte de grafos quadro a quadro em baixas escalas, *Stefan* (Figura 6.12) e *Angelfish* (Figura 6.13), o ajuste nas posições da regiões entre quadros, elevou a acurácia de 72,56% (NAQ) para 86,1% (AJQ) em *Stefan* (Figura 6.12) e de 80,87% (NAQ) para 94,07% (AJQ) em *Angelfish* (Figura 6.12). Nos grafos com pesos reforçados (REQ) e equivalente EQQ há um desempenho superior do caso sem ajuste, NAQ, entretanto abaixo do grafo com de ajuste pelo vetor de movimento, AJQ.

Ainda para as duas sequências *Stefan* e *Angelfish*, ao se realizar um corte de grafos ao pelas regiões formadas pelos 9 quadros, o desempenho sem ajuste, NAT, fica abaixo dos 50% para *Stefan*, retratando uma acurácia nula para os quadros finais (Figura 6.12 (a)). O corte no grafo equivalente da sequência *Stefan* apresenta taxa de acerto abaixo de AJQ, 80, 89% contra 86,1%, o maior para a sequência, entretanto, há uma redução na sobrestimação (Figuras 6.12(b) e 6.14(b)) e na quantidade de elementos entre os dois tipos de grafos, de 3911 nós para 1577 (Tabela 6.3), representa uma troca de operações com matriz de pesos de 15.295.921 componentes por uma de 2.486.029. Essa troca entre esforço computacional, taxa de acerto e sobrestimação pode ser uma discussão para trabalhos futuros.

(a)

(b)

Figura 6.12: Comparação de segmentação, do 5º ao 7º quadro da sequência *Stefan* relativos às segmentações NAQ (a) e AJQ (b). Observa-se uma maior acurácia para o mapa com posições ajustadas, AJQ, bem como uma maior sobrestimação.

(a)

(b)

Figura 6.13: Comparação de segmentação, do 7º ao 9º quadro da sequência *Angelfish* relativos às segmentações NAQ (a) e AJQ (b). Observa-se uma maior acurácia para o mapa com posições ajustadas, AJQ.

Para uma oitava intermediária, como é o caso da sequência *Trainer*, destaca-se o comportamento para os grafos equivalentes, EQQ (Figura6.15(a)) e EQT (Figura6.15(b)). No corte quadro a quadro, NAQ tem o melhor desempenho, com taxa de acerto de 95,33% dentro dos 9 quadros

(a)

(b)

Figura 6.14: Comparação de segmentação, do 7º ao 9º quadro da sequência *Stefan* relativos às segmentações NAT (a) e EQT (b). Observa-se uma baixa acurácia para a aplicação do corte de grafos formado por toda a sequência com um mapa de pesos obtido sem uma correção de movimento, NAT. Um mapa equivalente, EQT, que leva em consideração o movimento entre quadros, promove uma melhor segmentação com razoável acurácia e baixa sobrestimação.

(Tabela 6.4), superando o caso ajustado AJQ, com uma melhor acurácia e com uma menor sobrestimação. REQ juntamente a EQQ têm os piores desempenhos dentro da segmentação quadro a quadro, entretanto quando aplicado ao longo de toda a sequência, o corte em um grafo equivalente apresenta o melhor acurácia entre todos os casos, tendo uma sobrestimação cerca de 2% maior que NAQ.

Explorando novamente a troca entre a redução no número de elementos do grafo utilizado para corte e o desempenho da segmentação, tem-se para a sequência *Trainer* as maires acurácias para todos os casos em NAQ e EQT, 95,33% e 95,9%, respectivamente, e uma redução de 1035 elementos para 350 (Tabela 6.3). Ao invés de operações de corte efetuadas em uma matriz de 1.071.225 componentes, um grafo equivalente pode ser segmentado com uma matriz de 122.500 componentes, cerca de 9$\times$ menor.

Ressalta-se novamente a necessidade de estudos futuros relacionados ao *trade off* entre, esforço computacional, acurácia e sobrestimação. Apesar da compressão fornecida pelos grafos equivalentes, deve-se atentar ao fato de que a construção dos vetor de movimento, em conjunto com o casamento de regiões, é feita a partir de operações de matrizes com aproximadamente $\overline{N}_1^2$ elementos, no caso da sequência *Trainer*, matrizes em média com 13.225 componentes para se

rastrear o objeto e se calcular o vetor de movimento. Para escalas menores, ou grandes deslocamentos entre regiões (Figura 6.3), a correção de movimento pelo vetor de movimento se mostrou importante para o aumento na acurácia.

(a)

(b)

Figura 6.15: Comparação de segmentação, do 7º ao 9º quadro da sequência *Trainer* relativos às segmentações EQQ (a) e EQT (b). Observa-se uma maior acurácia para o mapa equivalente construído por regiões de toda a sequência, EQT. No caso quadro a quadro, EQT, além de uma acurácia mais baixa, observa-se uma sobrestimação relevante.

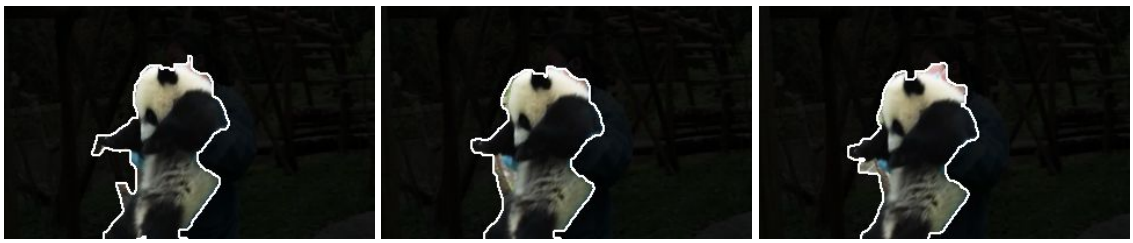
As sequências mais estáveis, como um movimento relativo entre regiões menos acentuadas (Figura 6.3), são as de maior escala, *Mobile* (4ª oitava) e *Panda* (5ª oitava). Nessas duas sequências há uma relação inversa a apresentada para a *Mobile* e *Panda*, a segmentação quadro a quadro promove uma menor sobrestimação do que o corte aplicado em toda a extensão de quadros. *Mobile* e *Panda* mantém valores de acurácia próximas aos 99% e 94%, respectivamente, para todos os mapas e tipos de corte.

Um bom desempenho dos mapas sem correção de movimento, NAQ e NAT, nas sequências *Mobile* e *Stefan*, indicam uma não necessidade de rastreamento e cálculo de vetor de movimento para situações de pouco movimento e distorção entre quadros. Entretanto, trabalhos futuros podem averiguar melhor a relação entre esforço computacional e desempenho, pode-se determinar uma quantidade de quadros ideal para uma corte ao longo de uma sequência longa e espera-se que diante de pouco movimento o número de iterações para o cálculo de vetor de movimento e rastreamento de regiões seja baixo.

(a)

(b)

Figura 6.16: Comparação de segmentação, do 7º ao 9º quadro da sequência *Mobile* relativos às segmentações EQQ (a) e EQT (b). Os resultados para sequência *Mobile* são bastante próximos, nos dois casos destacados, trabalha-se com grafos equivalentes, que comprimem os mapas de peso para segmentação, preservando o resultado final.



(a)

(b)

Figura 6.17: Comparação de segmentação, do 5º ao 7º quadro da sequência *Panda* relativos às segmentações EQQ (a) e EQT (b). Observa-se um maior nível de sobrestimação para o corte aplicado em toda a sequência, EQT.

7 CONCLUSÃO

O presente trabalho apresentou uma proposta de algoritmo que generaliza conceitos da transformação SIFT para grafos de regiões, visando aplicação em representações de uma imagem com relações de vizinhança menos triviais que a proporcionada por pixels. O descritor local proposto tem como objetivo o casamento de regiões entre quadros de vídeo para a segmentação de objetos ao longo de cenas, simplificando a análise de um volume de vídeo (espaço$\times$tempo) ao se eliminar redundâncias espacias e temporais.

No processo de desenvolvimento do algoritmo de criação de descritores locais para as regiões, foi desenvolvido um método de agrupamento que alia velocidade e automaticidade da técnica de *watershed* com a velocidade e precisão do algoritmo SLIC. Esse método de agrupamento de regiões se mostrou eficaz no propósito de conservar características de um objeto e cenas no decorrer de quadros de um vídeo, características importantes para a estabilidade dos descritores e a sua eficácia no casamento de regiões e importante para a segmentação de objetos em cenas.

O descritor proposto conseguiu conservar as propostas do SIFT, realizando o casamento de regiões mediante transformações geométricas nas cenas e nos objetos, como mudanças na escala das imagens, sua orientação e transformações no seu nível de intensidade. Não foram realizados testes específicos para avaliação direta da eficácia do descritor, ficando inicialmente restrita a uma avaliação perceptual.

As contribuições do descritor proposto foram analisadas de maneira indireta, com a comparação das segmentações em grafos que utilizam ou não o descritor proposto para um ajuste de posição dos elementos antes da determinação da força de ligação entre esses de um quadro para um quadro subsequente.

Outra forma de avaliação se baseou na análise da melhora na acurácia das segmentações em trechos de vídeo, ao se inserir informações nos grafos a respeito daqueles elementos que são correspondentes ao longo do trecho, quando não atribuídas forças de ligação mais altas, cada elemento e suas correspondências foram representados por um único nó dentro de um grafo equivalente destinado a segmentação.

Os resultados se mostraram favoráveis a utilização do algoritmo proposto na segmentação de vídeos via grafo, principalmente em situações de grande movimento relativo entre regiões de um objeto. Apesar dos testes terem sido aplicados em trechos curtos de vídeo, 9 quadros, em um número baixo de sequências, alguns resultados significativos foram encontrados, como

a necessidade da correção de movimento em uma segmentação realizada simultaneamente em todos os quadros da sequência. Essa correção é promovida por um fluxo óptico obtido por meio do algoritmo proposto.

7.1 CONSIDERAÇÕES FINAIS

Foi apresentada uma proposta de algoritmo que traz conceitos da transformação SIFT para o domínio dos grafos de região, criando descritores para as regiões que aumentam a discriminação entre elas, objetivando o rastreamento de objetos ao longo de cenas. Esse rastreamento se mostrou eficiente no aprimoramento da segmentação de objetos ao longo de cenas, ao ser utilizado na correção do movimento entre as regiões. O estudo do casamento de regiões para a criação de grafos equivalentes, com uma quantidade reduzida de elementos, apresentou resultados promissores para a utilização do algoritmo proposto em trabalhos futuros, visando a redução do esforço computacional na segmentação em volumes de vídeo.

7.2 TRABALHOS FUTUROS

Para trabalhos futuros espera-se aperfeiçoar o descritor e sua forma de aplicação em vídeos, avaliando a eficiência para sua utilização comparado a outros trabalhos. A princípio, deve-se melhorar a técnica de agrupamentos propostas. A união das técnicas *watershed* e SLIC para a criação de regiões/superpixels se mostrou resultados promissores, entretanto, se faz necessário a definição mais precisa dos parâmetros para essa adaptação. O método de agrupamentos por escalas pode ser utilizado para segmentações em hierarquia, aproveitando as características da imagem em diferentes estágios.

Os ganhos da adição de informações quanto os canais de cores no descritor proposto, um diferencial em relação ao SIFT, pode ser detalhadamente explorados. O descritor pode ser submetido a testes mais objetivos quanto o confronto de características entre cenas, com outras propostas de transformações entre imagens. Assim como os pontos-chave algoritmo SIFT, podem ser definidas regiões-chave, um grupo restrito de regiões para quais serão calculados os descritores que fornecem um melhor casamente de regiões.

No campo da segmentação, novas formas de organizar os grafos podem ser definidos, e outras técnicas de cortes em grafos aplicadas. A análise principal recai na definição de um equilíbrio entre esforço computacional, acurácia e erros sobre-segmentação na análise de volumes de vídeo.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] VAZQUEZ-REINA, A. et al. Multiple hypothesis video segmentation from superpixel flows. In: *Computer Vision–ECCV 2010*. [S.l.]: Springer, 2010. p. 268–281.

[2] XU, C.; CORSO, J. J. Evaluation of super-voxel methods for early video processing. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.], 2012. p. 1202–1209.

[3] NAGAHASHI, T.; FUJIYOSHI, H.; KANADE, T. Video segmentation using iterated graph cuts based on spatio-temporal volumes. In: *Computer Vision–ACCV 2009*. [S.l.]: Springer, 2010. p. 655–666.

[4] YANG, F.; LU, H.; YANG, M.-H. Robust superpixel tracking. *Image Processing, IEEE Transactions on*, IEEE, v. 23, n. 4, p. 1639–1651, 2014.

[5] GRUNDMANN, M. et al. Efficient hierarchical graph-based video segmentation. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. [S.l.], 2010. p. 2141–2148.

[6] HICKSON, S. et al. Efficient hierarchical graph-based segmentation of rgbd videos. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. [S.l.], 2014. p. 344–351.

[7] CHANG, J.; WEI, D.; FISHER, J. A video representation using temporal superpixels. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. [S.l.], 2013. p. 2051–2058.

[8] WANG, W.; NEVATIA, R. Robust object tracking using constellation model with superpixel. In: *Computer Vision–ACCV 2012*. [S.l.]: Springer, 2013. p. 191–204.

[9] BRENDEL, W.; TODOROVIC, S. Video object segmentation by tracking regions. In: IEEE. *Computer Vision, 2009 IEEE 12th International Conference on*. [S.l.], 2009. p. 833–840.

[10] LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Springer, v. 60, n. 2, p. 91–110, 2004.

[11] TANG, F.; TAO, H. Object tracking with dynamic feature graph. In: IEEE. *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*. [S.l.], 2005. p. 25–32.

[12] BATTIATO, S. et al. SIFT features tracking for video stabilization. In: IEEE. *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*. [S.l.], 2007. p. 825–830.

[13] SONG, Y.-Z. et al. Robust visual tracking using structural region hierarchy and graph matching. *Neurocomputing*, Elsevier, v. 89, p. 12–20, 2012.

[14] ZHAO, Y. et al. Experts-shift: Learning active spatial classification experts for keyframe-based video segmentation. In: IEEE. *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*. [S.l.], 2011. p. 622–627.

[15] ACHANTA, R. et al. SLIC superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 34, n. 11, p. 2274–2282, 2012.

[16] BEAR, M. F.; CONNORS, B. W.; PARADISO, M. A. *Neuroscience*. [S.l.: s.n.].

[17] GUYTON, A. C.; HALL, J. E. *Textbook of medical physiology*. 11. ed. [S.l.]: Elsevier Saunders, 2006. 613–637 p. Hardcover. ISBN 0721602401.

[18] GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. [S.l.]: Prentice Hall Upper Saddle River, NJ, 2002.

[19] SOUZA, G. d. S. et al. A visão através dos contrastes. *estudos avançados*, SciELO Brasil, v. 27, n. 77, p. 45–60, 2013.

[20] WIESEL, T. N.; HUBEL, D. H. et al. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J Neurophysiol*, v. 26, n. 6, p. 1003–1017, 1963.

[21] EDELMAN, S. Receptive fields for vision: From hyperacuity to object recognition. 1995.

[22] LINDEBERG, T. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, Taylor & Francis, v. 21, n. 1-2, p. 225–270, 1994.

[23] BONDY, J. A.; MURTY, U. S. R. *Graph theory with applications*. [S.l.]: Macmillan London, 1976.

[24] STAWIASKI, J. *Mathematical morphology and graphs: Application to interactive medical image segmentation*. Tese (Doutorado) — Ph. D. dissertation, Paris School Mines, Paris, France, 2008.

[25] HARARY, F. *Graph theory*. [S.l.]: Addison-Wesley, Reading, MA, 1969.

[26] DUNNE, P. Looking for consistency in the construction and use of Feynman diagrams. *Physics Education*, IOP Publishing, v. 36, n. 5, p. 366, 2001.

[27] HOLDSWORTH, J. *Feynman diagram*. 2008. `https://en.wikipedia.org/wiki/Feynman_diagram`. [Online; Acesso em 30/12/2015].

[28] ROBINSON, I.; WEBBER, J.; EIFREM, E. *Graph databases*. [S.l.]: "O'Reilly Media, Inc.", 2013.

[29] MISLOVE, A. et al. Measurement and analysis of online social networks. In: ACM. *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. [S.l.], 2007. p. 29–42.

[30] KIM, M.; LESKOVEC, J. Modeling social networks with node attributes using the multiplicative attribute graph model. *arXiv preprint arXiv:1106.5053*, 2011.

[31] BACKSTROM, L. et al. Four degrees of separation. In: ACM. *Proceedings of the 4th Annual ACM Web Science Conference*. [S.l.], 2012. p. 33–42.

[32] CAO, L.; KRUMM, J. From gps traces to a routable road map. In: ACM. *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*. [S.l.], 2009. p. 3–12.

[33] GONZALEZ, H. et al. Adaptive fastest path computation on a road network: a traffic mining approach. In: VLDB ENDOWMENT. *Proceedings of the 33rd international conference on Very large data bases*. [S.l.], 2007. p. 794–805.

[34] SONKA, M.; HLAVAC, V.; BOYLE, R. *Image processing, analysis, and machine vision*. [S.l.]: Cengage Learning, 2014.

[35] LINDEN, R. Técnicas de agrupamento. *Revista de Sistemas de Informação da FSMA*, v. 1, n. 4, p. 18–36, 2009.

[36] NAJMAN, L.; COUPRIE, M. Watershed algorithms and contrast preservation. In: SPRINGER. *Discrete geometry for computer imagery*. [S.l.], 2003. p. 62–71.

[37] MEYER, F. Topographic distance and watershed lines. *Signal processing*, Elsevier, v. 38, n. 1, p. 113–125, 1994.

[38] BOYKOV, Y. Y.; JOLLY, M.-P. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In: IEEE. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. [S.l.], 2001. v. 1, p. 105–112.

[39] SHI, J.; MALIK, J. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 22, n. 8, p. 888–905, 2000.

[40] FORD, A.; ROBERTS, A. Colour space conversions. *Westminster University, London*, v. 1998, p. 1–31, 1998.

[41] ROERDINK, J. B.; MEIJSTER, A. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta informaticae*, IOS Press, v. 41, n. 1, p. 187–228, 2000.

[42] WOLBERG, G. Sampling, reconstruction, and antialiasing. Citeseer, 2004.

[43] SHEWCHUK, J. R. Delaunay refinement algorithms for triangular mesh generation. *Computational geometry*, Elsevier, v. 22, n. 1, p. 21–74, 2002.

[44] VEZHNEVETS, V.; KONOUCHINE, V. GrowCut - Interactive multi-label N-D image segmentation by cellular automata. In: CITESEER. *proc. of Graphicon*. [S.l.], 2005. p. 150–156.

[45] GALASSO, F. et al. Spectral graph reduction for efficient image and streaming video segmentation. In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.].

[46] BAJIć, I. V. *Segmented foreground objects*. `http://www.sfu.ca/~ibajic/#data`. [Online; Acesso em 12/01/2016].

[47] VIDEO Object Co-Segmentation. `https://www.ece.nus.edu.sg/stfpage/eleclf/video_coseg.html`. [Online; Acesso em 12/01/2016].